The Hong Kong Polytechnic University
Department of Applied Mathematics

Optimality Conditions via Exact Penalty Functions

Kaiwen MENG

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

May 2011

CERTIFICATE OF ORIGINALITY

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgment is made in the text.

Kaiwen MENG

To my family

Abstract

The purpose of this thesis is, to study optimality conditions for constrained optimization problems in finite dimension spaces from the viewpoint of exact penalty functions. The tools that we use are mainly from the modern variational analysis popularized by Rockafellar and Wets' classical book. The problem models that we focus on are nonlinear programming and mathematical programs with complementarity constraints. We aim at developing a unified framework and providing a detailed exposition of optimality conditions from exactness of penalty functions. In this connection, we intend to answer questions as to when penalty functions are exact and how optimality conditions of the original constrained problems can be inherited from those of exact penalty functions.

We study sufficient conditions for penalty terms to possess local error bounds, which guarantee exactness of penalty functions. We give characterizations for a stronger version of the local error bound property in terms of strong slopes, subderivative and regular subgradients for points outside the referenced set. In particular, we give full characterizations of the local error bound property for the elementary max function of a finite collection of smooth functions. With the aid of these characterizations, we show that the quasinormality constraint qualification implies the existence of a local error bound. We also study sufficient and necessary conditions for the existence of local error bounds by virtue of various limits defined on the boundary of the referenced set.

We study first- and second-order necessary and sufficient conditions for penalty functions to be exact. These conditions are expressed by subderivatives, second-order subderivatives, and parabolic subderivatives, which are the notions that have been utilized to formulate tight optimality conditions for optimization problems. In our investigation, the kernels of these derivatives, representing directions at which derivatives vanish, play an key role. In particular, we show an interesting auxiliary result which asserts that, the polar cone of the subderivative kernel of an extended real-valued function at a local minimum is the same as the positive hull of its regular subgradients at the same point.

We show how Karush-Kuhn-Tucker conditions and second-order necessary conditions in nonlinear programming, and strong and Mordukhovich stationarities in mathematical programs with complementarity constraints, can be derived from exactness of penalty functions under some additional conditions on constraint functions. In presenting these additional conditions, it turns out that the kernels of (parabolic) subderivatives of penalty terms are very crucial. By virtue of these kernels and a variational description of regular subgradients, we show necessity and sufficiency of these additional conditions. We also present conditions in terms of the original data by applying (generalized) Taylor expansions to calculate these kernels.

Acknowledgments

Mom, Dad, Brothers (Kaixue and Kaike), Sisters (Kaixiu, Kaiqin and Kaimei), Nephews (Zhengwei, Dongdong, Mengzhu and Zhiqiu), and Nieces (Zhenghua, Mengjiao, Mengjun and Yuanyuan), this thesis is dedicated to all of you. I would never have made it without your love, faith, and support. This thesis is also dedicated to my wonderful and amazing girlfriend, Tian Feifei, for her love, encouragement, support and tolerance.

I would like to express my gratitude to all the people who contributed in making my three and a half years stay at Hong Kong Polytechnic University a fun time, full of invaluable experiences.

First of all, I wish to express deep gratitude to my supervisor Professor Xiaoqi Yang. Thanks to him for providing me a challenging research topic and helping me to find the right research direction to follow, for teaching me how to analyze everything with patience, and for careful reading of my manuscripts and useful comments and suggestions on them over the last three and a half years. I am greatly indebted to him, not only professionally, but also personally, for providing me with the opportunity to study at Hong Kong Polytechnic University and for always being there when I needed his help.

A number of people have given me great help and advice. Thanks to Professor Xiaojun Chen (Hong Kong Polytechnic University) for paying attention to my research work and for always encouraging me. Thanks to Dr. Yaping Fang (Sichuan University) for the stimulating discussion on how to give characterizations for semi-closed polyhedra during a joint work. Thanks to Professor Xuexiang Huang (Chongqing University) for helping me to find a counterexample to illustrate that in the context of mathematical programs with complementarity constraints (for short, MPCC), the MPCC Guignard constraint qualification is not the weakest one for Mordukhovich stationarity. Thanks to Dr. Houduo Qi (University of Southampton) for offering me a short visit to him and for a helpful discussion on features of good research work. Thanks to Professor Daniel Ralph (University of Cambridge) for guiding me during my three months stay at University of Cambridge and for drawing my attention to the theory of 2-regularity which could be an interesting topic for my future research. Thanks to Professor R. Tyrrell Rockafellar (University of Washington) for posing a good research question to my supervisor and then to me at a conference held in Beijing, which motivates me to try hard to work it out and learn a lot from chasing it. In addition, I would like to acknowledge a number of editors and anonymous referees for giving me valuable comments and suggestions which have contributed to a better presentation of certain results.

I also wish to thank many good friends and colleagues: An Congpei, Chen Chunrong, Chen Zhangyou, Chen Zhe, Fang Donghui, Feng Zhiguo, Huang Haiqiao, Hu Yaohua, Li Xuesong, Tian Boshi, Xiao Yibin, Zhang Fuwen and Zhou Hongjun for their friendship and help. They made my time at Hong Kong Polytechnic University more enjoyable. Thanks to Sharon, a truly great Hong Konger and friend, who can speak excellent Putonghua and made me smile every time I felt sad.

Thanks are also due to Research Committee of Hong Kong Polytechnic University and Research Grants Council of Hong Kong (with grant number: PolyU 5303/05E) for providing me financial support over the last three and a half years. I sincerely thank supporting staff at Department of Applied Mathematics: Mr. Fu To (Benny), Ms. Hau Ling-ling (Cynthia), Miss Yiu Yee-wah (Eva), Mrs. Fanny Cheung, Ms. Man Wai-lin (Ida), Ms. Au Wah-yin (Jojo), Mr. Cheng Tak-shing (Joseph), Miss Law Siu-lin (Margaret), Mr. Lo Charn-tong (Peter), Ms. Li Mang Wan (Shermie) for their kindness and support. I shall always treasure the stimulating research environment and the excellent research facilities of the Department of Applied Mathematics at Hong Kong Polytechnic University.

This thesis is based on the following papers written by the author during the period of stay at the Department of Applied Mathematics, The Hong Kong Polytechnic University as a graduate student:

1. K. W. Meng and X. Q. Yang, Optimality conditions via exact penalty function. SIAM Journal on Optimization, 20(6):3208-3231, 2010.

2. K. W. Meng and X. Q. Yang, Characterizations of local error bounds, (in preparation).

3. K. W. Meng and X. Q. Yang, Second-order optimality conditions via exact penalty functions, (in preparation).

The author of the thesis participated in another project during the last three and a half years, which results in a paper as follows:

4. Y. P. Fang, K. W. Meng and X. Q. Yang, Piecewise linear multi-criteria programs: the continuous case and its discontinuous generalization, Operations Research, (accepted).

However, the content of this paper is by no means related to the topic of the thesis, and therefore, we do not include it in the thesis.

Contents

Chapter 1

Preview and Introduction

1.1 Review on Nonlinear Programming Problems

Consider the nonlinear programming problem

$$\begin{array}{lll} \text{(NLP)} & \min & f(x) \\ & \text{s.t.} & g_i(x) \leq 0, \quad i \in I, \\ & & h_j(x) = 0, \quad j \in J, \end{array}$$

where $I = \{1, 2, \cdots, m\}$, $J = \{m+1, m+2, \cdots, m+q\}$, and the functions $f, g_i, h_j : R^n \to R$ are assumed to be continuously differentiable. Associated with (NLP), the Lagrange function $L : R^n \times R^{m+q} \to R$ is defined by

$$L(x, \lambda) := f(x) + \sum_{i \in I} \lambda_i g_i(x) + \sum_{j \in J} \lambda_j h_j(x),$$

and the generalized Lagrange function $\tilde{L} : R^n \times R \times R^{m+q} \to R$ is defined by

$$\tilde{L}(x, \lambda_0, \lambda) := \lambda_0 f(x) + \sum_{i \in I} \lambda_i g_i(x) + \sum_{j \in J} \lambda_j h_j(x).$$

In what follows, let C be the feasible set of (NLP) and let $\bar{x} \in C$ be fixed. Note that any point in C is called a feasible point of (NLP). If $f(\bar{x}) \leq f(x)$ for all feasible $x \neq \bar{x}$ in some neighborhood of $\bar{x}$, then $\bar{x}$ is called a local minimum. If $f(\bar{x}) < f(x)$ for all feasible $x \neq \bar{x}$ in some neighborhood of $\bar{x}$, then $\bar{x}$ is called a strict local minimum.

And if there exist some positive integer κ and some positive number τ such that

$$f(\bar{x}) + \tau\|x - \bar{x}\|^{\kappa} \leq f(x)$$

for all feasible x in some neighborhood of $\bar{x}$, then $\bar{x}$ is called a strict local minimum of order κ. It is clear that a strict local minimum of any order κ is a strict local minimum.

First-order necessary and sufficient conditions and constraint qualifications

In what follows, let the index set of active inequality constraints at $\bar{x}$ be defined by

$$I(\bar{x}) := \{i \in I \mid g_i(\bar{x}) = 0\},$$

let the (Bouligand) tangent cone to C at $\bar{x}$ be defined by

$$T_C(\bar{x}) := \{w \in R^n \mid \exists t_k \to 0+, \exists w_k \to w, \text{ s.t. } \bar{x} + t_k w_k \in C \ \forall k\},$$

and let the (first-order) linearized tangent cone to C at $\bar{x}$ be defined by

$$L_C(\bar{x}) = \left\{ w \in R^n \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^T w \leq 0 & \forall i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T w = 0 & \forall j \in J \end{array} \right\}.$$

When $\bar{x}$ is a local minimum of (NLP), the basic primal necessary condition for (NLP) can be expressed as follows:

$$\nabla f(\bar{x})^T w \geq 0 \quad \forall w \in T_C(\bar{x}).$$

Conversely, if the first-order sufficient condition holds as follows:

$$\nabla f(\bar{x})^T w > 0 \quad \forall w \in T_C(\bar{x}) \backslash \{0\},$$

then $\bar{x}$ is a strict local minimum of (NLP). Since $T_C(\bar{x}) \subset L_C(\bar{x})$ holds automatically, a stronger version of the first-order sufficient condition can be expressed as follows:

$$\nabla f(\bar{x})^T w > 0 \quad \forall w \in L_C(\bar{x}) \backslash \{0\}. \tag{1.1.1}$$

The Fritz John condition (also known as the FJ condition), named after Fritz John [90], holds at $\bar{x}$ if there exists a vector $(\lambda_0, \lambda) \in R \times R^{m+q}$ such that $\lambda_0 \geq 0$, $(\lambda_0, \lambda) \neq (0, 0)$,

$$\nabla_x \tilde{L}(\bar{x}, \lambda_0, \lambda) = 0 \quad \lambda_i \geq 0, \ \lambda_i g_i(\bar{x}) = 0 \quad \forall i \in I.$$

We call such a vector (λ_0, λ) an FJ multiplier, and denote by $\mathrm{FJ}(\bar{x})$ the set of all FJ multipliers of (NLP) at $\bar{x}$. It is clear that $\mathrm{FJ}(\bar{x}) \cup \{(0,0)\}$ is a polyhedral cone, and it is well-known that $\mathrm{FJ}(\bar{x}) \neq \emptyset$ whenever $\bar{x}$ is a local minimum of (NLP). The primal form of the FJ condition at $\bar{x}$ can be expressed as the inconsistency of the following system:

$$\nabla f(\bar{x})^T w < 0, \quad \nabla g_i(\bar{x})^T w < 0 \; \forall i \in I(\bar{x}), \quad \nabla h_j(\bar{x})^T w = 0 \; \forall j \in J.$$

The Karush-Kuhn-Tucker condition (also known as the KKT condition) holds at $\bar{x}$ if there exists a vector $\lambda \in R^{m+q}$ such that

$$\nabla_x L(\bar{x}, \lambda) = 0, \quad \lambda_i \geq 0, \; \lambda_i g_i(\bar{x}) = 0 \quad \forall i \in I.$$

We call such a vector λ a KKT multiplier, and denote by $\mathrm{KKT}(\bar{x})$ the set of all KKT multipliers of (NLP) at $\bar{x}$. It is clear to see that $\mathrm{KKT}(\bar{x})$ is a polyhedral set and that $\lambda \in \mathrm{KKT}(\bar{x})$ if and only if $(1, \lambda) \in \mathrm{FJ}(\bar{x})$. The KKT conditions were originally named after Harold W. Kuhn and Albert W. Tucker, who first published the conditions in [96]. Later scholars discovered that the necessary conditions for this problem had been stated by William Karush in his master's thesis [93]. The primal form of the KKT condition at $\bar{x}$ can be expressed as the inconsistency of the following system:

$$\nabla f(\bar{x})^T w < 0, \quad \nabla g_i(\bar{x})^T w \leq 0 \; \forall i \in I(\bar{x}), \quad \nabla h_j(\bar{x})^T w = 0 \; \forall j \in J.$$

It should be noticed that KKT conditions may not hold at local minima of (NLP) unless some regularity conditions are satisfied. By regularity conditions, we mean various conditions imposed on the problem data, some of which may depend on the constraint functions only, while some of which may depend on the objective function as well. When regularity conditions are independent of the objective functions, they are more often known as the constraint qualifications (CQs) in the literature. For various CQs appeared in the literature, we refer to the survey papers by Peterson [127] and Bazaraa et al. [15], and the text books [17, 16].

In the following, we list a number of CQs that are frequently used in the literature or has been studied recently, and discuss their relationships.

The linear independence constraint qualification [50] (LICQ) holds at $\bar{x}$ if the vectors $\{\nabla g_i(\bar{x}), i \in I(\bar{x})\} \cup \{\nabla h_j(\bar{x}), j \in J\}$ are linearly independent. If $\bar{x}$ is a local minimum of (NLP), the LICQ at $\bar{x}$ implies that $\mathrm{KKT}(\bar{x})$ is a singleton.

The Mangasarian-Fromovitz constraint qualification [108] (MFCQ) holds at $\bar{x}$ if the gradients of the equality constraints are linearly independent at $\bar{x}$, and there exists $w \in R^n$ such that $\nabla g_i(\bar{x})^T w < 0$ for all $i \in I(\bar{x})$ and $\nabla h_j(\bar{x})^T w = 0$ for all $j \in J$. Applying a theorem of alternatives [107], the equivalent dual form of MFCQ at $\bar{x}$ asserts that the vector pair $(\{\nabla g_i(\bar{x}), i \in I(\bar{x})\}, \{\nabla h_j(\bar{x}), j \in J\})$ is positive-linearly independent. Here, the vector pair $(\{a_1, \ldots, a_k\},\ \{a_{k+1}, \ldots, a_l\})$ is said to be positive-linearly independent if

$$\sum_{i=1}^{l} \lambda_i a_i = 0, \quad \lambda_i \geq 0 \quad \forall i = 1, \ldots, k \Longrightarrow \lambda_1 = \ldots = \lambda_l = 0,$$

otherwise, it is positive-linearly dependent. If $\bar{x}$ is a local minimum of (NLP), then the MFCQ at $\bar{x}$ amounts to the boundedness of $\mathrm{KKT}(\bar{x})$ [62], and also amounts to the emptiness of the set $(\mathrm{FJ}(\bar{x}) \cap \{0\} \times R^{m+q})$ according to the dual form of MFCQ.

The constant rank constraint qualification (CRCQ) holds at $\bar{x}$ if the rank for each subset of the gradients of the active inequality constraints and the equality constraints at a neighborhood of $\bar{x}$ is constant, or in other words for each $I' \subset I(\bar{x})$ and $J' \subset J$, if the vectors $\{\nabla g_i(\bar{x}), i \in I'\} \cup \{\nabla h_j(\bar{x}), j \in J'\}$ are linearly dependent, then the vectors $\{\nabla g_i(x), i \in I'\} \cup \{\nabla h_j(x), j \in J'\}$ are linearly dependent for all x in some neighborhood of $\bar{x}$. The CRCQ was introduced by Janin [87], and its weaker version, called the relaxed CRCQ, has been recently studied in [111]. If the constraints of (NLP) are all defined by affine functions, the CRCQ is obviously satisfied at every feasible point. Moreover, if the CRCQ holds at $\bar{x}$ and some equality constraint $h_j(x) = 0$ is replaced by two inequality constraints: $h_j(x) \leq 0$ and $-h_j(x) \leq 0$, the CRCQ still holds at $\bar{x}$ with the new description of the feasible set. Note that the MFCQ does not enjoy this property.

The constant positive linear dependence constraint qualification (CPLD) holds at $\bar{x}$ if for each $I' \subset I(\bar{x})$ and $J' \subset J$, the positive-linear dependence of the vector pair $(\{\nabla g_i(\bar{x}), i \in I'\}, \{\nabla h_j(\bar{x}), j \in J'\})$ implies the linear dependence of the vectors $\{\nabla g_i(x), i \in I'\} \cup \{\nabla h_j(x), j \in J'\}$ for all x in some neighborhood of $\bar{x}$. The CPLD was introduced for use in the analysis of SQP methods by Qi and Wei [131] who conjectured that CPLD could be a constraint qualification. This conjecture is proved to be true in [4].

The quasi-normality constraint qualification (QNCQ) holds at $\bar{x}$ if there exist no nonzero vector $\lambda \in R_+^m \times R^q$ and no sequence $x_k \to \bar{x}$ such that

$$\sum_{i \in I} \lambda_i \nabla g_i(\bar{x}) + \sum_{j \in J} \lambda_j \nabla h_j(\bar{x}) = 0,$$

and for all k, $\lambda_i g_i(x_k) > 0$ for all i with $\lambda_i > 0$, and $\lambda_j h_j(x_k) > 0$ for all j with $\lambda_j \neq 0$. The QNCQ was introduced by Hestenes [75], and a slightly stronger CQ, called the pseudo-normality constraint qualification (PNCQ) has been proposed and investigated by Bertsekas and Ozdaglar [22] for constrained optimization problems with not only equality and inequality constraints but an abstract set constraint. If the functions g_i are all concave and the functions h_j are all linear, then the QNCQ holds at each feasible point.

The Abadie constraint qualification [1] (ACQ) holds at $\bar{x}$ if $T_C(\bar{x}) = L_C(\bar{x})$. Note that the ACQ at $\bar{x}$ was referred to as $\bar{x}$ being a regular point in [75], and as $\bar{x}$ being a quasi-regularity point in [22].

The Guignard constraint qualification [65] (GCQ) holds at $\bar{x}$ if $T_C(\bar{x})^* = L_C(\bar{x})^*$, where for a given subset A of R^n, $A^* := \{v \in R^n \mid v^T x \leq 0 \quad \forall x \in A\}$ stands for the polar cone of A.

If $\bar{x}$ is a local minimum of (NLP), then the KKT condition holds at $\bar{x}$ provided that one of the CQs described previously is satisfied. The relationships among these CQs are as follows:

$$\text{LICQ} \Longrightarrow (\text{MFCQ or CRCQ}) \Longrightarrow \text{CPLD} \Longrightarrow \text{QNCQ} \Longrightarrow \text{ACQ} \Longrightarrow \text{GCQ}.$$

Most of the implications are straightforward. In particular, Andreani [4] showed the implications MFCQ $\Longrightarrow$ CPLD, CRCQ $\Longrightarrow$ CPLD and CPLD $\Longrightarrow$ QNCQ, and demonstrated by several examples that the reverse implications do not hold, see also Qi and Wei [131]. Moreover, Hestenes [75] showed the implication QNCQ $\Longrightarrow$ ACQ, while Janin [87] showed the implication CRCQ $\Longrightarrow$ ACQ. Note that the MFCQ is neither weaker nor stronger than the CRCQ, see [87]. However, if the CRCQ holds, there exists an alternative representation of the feasible set for which the MFCQ holds, see a recent paper by Lu [103]. Among all CQs, the GCQ is the weakest one in the sense that GCQ holds at $\bar{x}$ if and only if the KKT condition holds at $\bar{x}$ whenever a continuously differentiable objective function f has a local minimum at $\bar{x}$ relative to C, see Gould and Tolle [64] for the original version of this result, and Theorem 6.11 of Rockafellar and Wets [141] for new features of this result.

It was established by Robinson [133] that the MFCQ holds at $\bar{x}$ if and only if the set-valued mapping $M : R^{m+q} \rightrightarrows R^n$ defined by

$$M(y) = \left\{ x \in R^n \;\middle|\; \begin{array}{ll} g_i(x) \leq y_i, & i \in I \\ h_j(x) = y_j, & j \in J \end{array} \right\}, \tag{1.1.2}$$

has the Aubin property [141] (also known as the pseudo-Lipschitz continuity [7]) at $\bar{y} = 0$ for $\bar{x}$ or equivalently the inverse mapping M^{-1} is metric regularity [85, 47] at $\bar{x}$ for $\bar{y}$. This stability property highlights the special role of MFCQ among all the other CQs. See Chapter 9 of Rockafellar and Wets [141] for extensive discussions on the notion of metric regularity, the Aubin property, and the calmness variant of the Aubin property.

Let $d_C(x)$ be the distance of the point x from C and let

$$S(x) := \sum_{i \in I} \max\{g_i(x), 0\} + \sum_{j \in J} |h_j(x)| \quad \forall x \in R^n. \tag{1.1.3}$$

We say that S is a local error bound for C at $\bar{x}$ if there exist some $\tau > 0$ and a neighborhood V of $\bar{x}$ such that

$$\tau d_C(x) \leq S(x) \quad \forall x \in V. \tag{1.1.4}$$

We say that S is a global error bound for C if the inequality (1.1.4) holds with $V = R^n$. Since all the norms in a finite dimensional space are equivalent, the right-hand side of the inequality (1.1.4) can be replaced by functions induced from other norms, such as the function

$$\max\{0, \, g_i(x), i \in I, \, |h_j(x)|, j \in J\}, \tag{1.1.5}$$

which is induced from the ℓ_∞ norm and was considered in [112]. According to Henrion and Outrata [73], and Dontchev and Rockafellar [47], S is a local error bound for C at $\bar{x}$, if and only if the set-valued mapping M given by (1.1.2) is calm at $\bar{y} = 0$ for $\bar{x}$ or equivalently the inverse mapping M^{-1} is metric subregularity at $\bar{x}$ for $\bar{y}$. It should be noticed that the metric subregularity, unlike the metric regularity, may not imply the stability, as pointed out by Dontchev and Rockafellar [47].

It is well-known that the ACQ holds at $\bar{x}$ if (1.1.4) is satisfied, and that the converse is in general not true, see in particular Henrion and Outrata ([73], Proposition 1 and Example 1). Since the Aubin property implies calmness, (1.1.4) holds under the MFCQ. Janin [87] essentially showed that (1.1.4) holds if the CRCQ holds at $\bar{x}$. Very recently, Minchenko and Tarakanov [112] showed that if the QNCQ holds at $\bar{x}$ and the gradients of g_i and h_j are locally Lipschitz continuous, then the function defined by (1.1.5) is a local error bound for C at $\bar{x}$ or equivalently (1.1.4) holds. However, the assumption on Lipschitz continuity of the gradients is not necessary, as will be seen from a characterization of local error bounds presented in Chapter 2. As such, the existence of local error bounds can be treated as a constraint qualification that takes a position between the QNCQ and the ACQ. To sum up, we have

$$\text{QNCQ} \Longrightarrow (1.1.4) \Longrightarrow \text{ACQ}.$$

Originated from the practical implementation and numerical considerations of iterative methods for solving optimization problems, the study of error bounds has received increasing attention in many interesting areas such as sensitivity and stability analysis, subdifferential calculus, exact penalty functions, and optimality conditions, see [36, 25, 84, 82, 91, 95, 132, 86] and especially the excellent survey papers by Lewis and Pang [98], and Pang [124] for more details. It should be noticed that the notion of error bound is closely related with the notions of weak sharp minima [49, 30], calmness [141] and subregularity [47]. These notions are equivalent with each other in the sense that each of them can be used to interpret the others. Sufficient conditions ensuring local or global error bounds have been studied in [100, 148, 41, 42, 101, 153, 29, 118, 69, 170, 171, 172, 114] under the convexity assumption, in [104, 106, 43, 105] under the analyticity assumption, and in [115, 116, 79, 156, 155, 158, 157, 10, 11, 119] for general lower semi-continuous functions. With other particular structures being imposed, sufficient conditions ensuring local or global error bounds can also be found in [151, 27, 28, 117, 174, 44, 173, 134, 71, 73, 70, 72].

Second-order necessary and sufficient conditions and constraint qualifications

When second-order optimality conditions are discussed in this thesis, all the functions in defining (NLP) are assumed to twice continuously differentiable. The critical cone of (NLP) at $\bar{x}$ is defined by

$$\mathcal{V}(\bar{x}) := \left\{ w \in R^n \;\middle|\; \begin{array}{ll} \nabla f(\bar{x})^T w \leq 0 & \\ \nabla g_i(\bar{x})^T w \leq 0 & \forall i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T w = 0 & \forall j \in J \end{array} \right\}.$$

If $\mathrm{KKT}(\bar{x}) \neq \emptyset$, then the inequality $\nabla f(\bar{x})^T w \leq 0$ in the definition of the critical cone $\mathcal{V}(\bar{x})$ can be replaced by the equality $\nabla f(\bar{x})^T w = 0$, and for any $\lambda \in \mathrm{KKT}(\bar{x})$, $\mathcal{V}(\bar{x})$ can be reformulated as

$$\mathcal{V}(\bar{x}) = \left\{ w \in R^n \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^T w \leq 0 & \forall i \in \{i \in I(\bar{x}) \mid \lambda_i = 0\} \\ \nabla g_i(\bar{x})^T w = 0 & \forall i \in \{i \in I(\bar{x}) \mid \lambda_i > 0\} \\ \nabla h_j(\bar{x})^T w = 0 & \forall j \in J \end{array} \right\},$$

see [66, 20, 23]. The reduced critical cone of (NLP) at $\bar{x}$ is defined by

$$\mathcal{V}^r(\bar{x}) := \left\{ w \in R^n \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^T w = 0 & \forall i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T w = 0 & \forall j \in J \end{array} \right\}.$$

If $\mathrm{KKT}(\bar{x}) \neq \emptyset$, then $\mathcal{V}^r(\bar{x}) \subset \mathcal{V}(\bar{x})$, and if in addition, the strict complementarity condition holds at $\bar{x}$ with respect to some $\bar{\lambda} \in \mathrm{KKT}(\bar{x})$ (i.e., $\bar{\lambda}_i > 0$ for all $i \in I(\bar{x})$), then $\mathcal{V}^r(\bar{x}) = \mathcal{V}(\bar{x})$, see [3, 23].

There can be found in the literature four different types of second-order necessary conditions for a local minimum $\bar{x}$ of (NLP). They are as follows.

There exists at least one $\bar{\lambda} \in \mathrm{KKT}(\bar{x})$ such that

$$\text{(SON1)} \quad w^T \nabla^2_{xx} L(\bar{x}, \bar{\lambda}) w \geq 0 \quad \forall w \in \mathcal{V}(\bar{x}).$$

There exists at least one $\bar{\lambda} \in \mathrm{KKT}(\bar{x})$ such that

$$\text{(SON1)}^r \quad w^T \nabla^2_{xx} L(\bar{x}, \bar{\lambda}) w \geq 0 \quad \forall w \in \mathcal{V}^r(\bar{x}).$$

For each $w \in \mathcal{V}(\bar{x})$, there exists some $\lambda \in \mathrm{KKT}(\bar{x})$ such that

$$\text{(SON2)} \quad w^T \nabla^2_{xx} L(\bar{x}, \lambda) w \geq 0. \tag{1.1.6}$$

For each $w \in \mathcal{V}(\bar{x})$, there exists some $(\lambda_0, \lambda) \in \mathrm{FJ}(\bar{x})$ such that

$$\text{(SON3)} \quad w^T \nabla^2_{xx} \tilde{L}(\bar{x}, \lambda_0, \lambda) w \geq 0.$$

It is clear that (SON1) $\Longrightarrow$ (SON1)^r and (SON1) $\Longrightarrow$ (SON2) $\Longrightarrow$ (SON3). The (SON2) and (SON3) are distinguished by the use of the entire set of multipliers rather than a single multiplier vector as is the case for the (SON1) and the (SON1)^r. Since the (SON3) relies on FJ multipliers, it holds without any constraint qualification, see Proposition 5.48 of [23], Ben-Tal [19], and Ben-Tal and Zowe [20].

Under the LICQ at $\bar{x}$, which implies that KKT$(\bar{x})$ is a singleton, the (SON1)^r was first obtained by McCormick [109], see also [50]. McCormick [109] actually showed the (SON1)^r under a weaker condition: each $w \in L_C(\bar{x})\backslash\{0\}$ is tangent to a once-differentiable arc, emanating from $\bar{x}$ and contained in the feasible set, and each nonzero vector $w \in \mathcal{V}^r(\bar{x})$ is the tangent of an arc $\alpha(\theta)$, twice differentiable, along which $g_i(\alpha(\theta)) \equiv 0$ for all $i \in I(\bar{x})$ and $h_j(\alpha(\theta)) \equiv 0$ for all $j \in J$, where $\theta \in [0, \varepsilon]$, $\varepsilon > 0$. Under the LICQ, the (SON1), a slightly stronger condition than the (SON1)^r, has also been established in the text books [21, 60, 120, 16], and furthermore, even a stronger condition than the (SON1) can be obtained which asserts that the unique KKT multiplier $\bar{\lambda}$ satisfies

$$w^T \nabla^2_{xx} L(\bar{x}, \bar{\lambda}) w \geq 0 \quad \forall w \in L_C(\bar{x}),$$

see Theorem 3.3 of [19] and Theorem 4.4.3 in the text book by Bazaraa et al. [16].

Recently, Andreani et al.(2010) showed in Theorem 3.1 of [3] that the CRCQ implies the (SON1). Note that KKT$(\bar{x})$ may not be a singleton under the CRCQ. They actually showed that for any $\bar{\lambda} \in$ KKT$(\bar{x})$, the (SON1) holds. Moreover, they showed that the (SON1)^r holds for any $\bar{\lambda} \in$ KKT$(\bar{x})$, under any CQ ensuring KKT$(\bar{x}) \neq \emptyset$, and the weak constant-rank (WCR) condition at $\bar{x}$ (i.e., the rank of the vectors $\{\nabla g_i(x), i \in I(\bar{x})\} \cup \{\nabla h_j(x), j \in J\}$ does not change in some neighborhood of $\bar{x}$). The WCR condition, originated with Penot [126], implies the ACQ for (NLP) with only equality constraints, see [126]. However, when inequality constraints are involved, the WCR condition may not guarantee that KKT$(\bar{x}) \neq \emptyset$, see [4] for an example.

The counterexamples given in [6, 5, 13] demonstrate that the MFCQ alone cannot guarantee the (SON1) or the (SON1)^r, i.e., there may not exist a fixed $\bar{\lambda} \in$ KKT$(\bar{x})$ such that $\nabla^2_{xx} L(\bar{x}, \bar{\lambda})$ is positive semi-definite on $\mathcal{V}(\bar{x})$ or $\mathcal{V}^r(\bar{x})$, unless additional conditions are imposed, see Baccari [12] and Baccari and Trad [13]. Therefore, any CQ weaker than the MFCQ does not imply the (SON1) or the (SON1)^r.

Of particular note in derivation of the (SON1) is the work by Rockafellar [137] who employed a perturbation method that does not require any CQ by estimating the generalized subgradients of the optimal value function associated with a parameterized nonlinear programming problem.

The (SON2), developed initially by Ioffe [83], has been extensively studied, see [19, 20, 94, 138, 139, 26, 24, 18, 31]. In what follows, we shall recall some constraint qualifications using the information of second-order derivatives, which can guarantee the (SON2). Ben-Tal's Constraint Qualification (BTCQ) [19, 20] holds at $\bar{x}$ if the vectors $\{\nabla h_j(\bar{x}), j \in J\}$ are linearly independent, and for each $w \in \mathcal{V}(\bar{x})$, there exists some $z \in R^n$ such that

$$\begin{aligned} &\nabla g_i(\bar{x})^T z + w^T \nabla^2 g_i(\bar{x}) w < 0 \quad \forall i \in I(\bar{x}, w), \\ &\nabla h_j(\bar{x})^T z + w^T \nabla^2 h_j(\bar{x}) w = 0 \quad \forall j \in J, \end{aligned}$$

where $I(\bar{x}, w) := \{i \in I(\bar{x}) \mid \nabla g_i(\bar{x})^T w = 0\}$ is the active index set of the inequality constraints at $\bar{x}$ in the direction w.

The second-order linearized tangent set to C at $\bar{x}$ in the direction $w \in L_C(\bar{x})$ [94, 23] is defined by

$$L_C^2(\bar{x} \mid w) = \left\{ z \in R^n \;\middle|\; \begin{array}{ll} \langle \nabla g_i(\bar{x}), z\rangle + \langle w, \nabla^2 g_i(\bar{x}) w\rangle \le 0 & \forall i \in I(\bar{x}, w) \\ \langle \nabla h_j(\bar{x}), z\rangle + \langle w, \nabla^2 h_j(\bar{x}) w\rangle = 0 & \forall j \in J. \end{array} \right\}.$$

If $w \notin L_C(\bar{x})$, the set $L_C^2(\bar{x} \mid w)$ is interpreted as an empty set. The second-order tangent set $T_C^2(\bar{x} \mid w)$ [141] to C at $\bar{x}$ for a vector $w \in T_C(\bar{x})$ consists of vectors z such that there are sequences $t_k \to 0+$ and $z_k \to z$ such that $\bar{x} + t_k w + \frac{1}{2} t_k^2 z_k \in C$ for all k. If $w \notin T_C(\bar{x})$, $T_C^2(\bar{x} \mid w)$ is interpreted as an empty set. The second-order Abadie constraint qualification (SACQ) [94] holds at $\bar{x}$ if

$$T_C^2(\bar{x} \mid w) = L_C^2(\bar{x} \mid w) \quad \forall w \in \mathcal{V}(\bar{x}),$$

and the second-order Guignard constraint (SGCQ) [94] holds at $\bar{x}$ if

$$\operatorname{clconv}[T_C^2(\bar{x} \mid w)] = L_C^2(\bar{x} \mid w) \quad \forall w \in \mathcal{V}(\bar{x}),$$

where the set clconv$A$ stand for the closed convex hull of A. It was shown in [94] that

$$\text{MFCQ} \Longrightarrow \text{BTCQ} \Longrightarrow \text{SACQ} \Longrightarrow \text{SGCQ}.$$

If the local error bound property (1.1.4) holds at $\bar{x}$, it is easy to check by definition that the SACQ holds at $\bar{x}$, see also He and Sun [68]. Therefore, any CQ, which implies the local error bound property, can be used to derive the (SON2).

The feasible point $\bar{x}$ is a strict local minimum of order 2 for (NLP), if one of the following second-order sufficient conditions is satisfied:

There exists some $\bar{\lambda} \in \mathrm{KKT}(\bar{x})$ such that

$$\text{(SOS1)} \quad w^T \nabla^2_{xx} L(\bar{x}, \bar{\lambda}) w > 0 \quad \forall w \in \mathcal{V}(\bar{x}) \backslash \{0\}.$$

For each $w \in \mathcal{V}(\bar{x}) \backslash \{0\}$, there exists some $\lambda \in \mathrm{KKT}(\bar{x})$ such that

$$\text{(SOS2)} \quad w^T \nabla^2_{xx} L(\bar{x}, \lambda) w > 0.$$

For each $w \in \mathcal{V}(\bar{x}) \backslash \{0\}$, there exists some $(\lambda_0, \lambda) \in \mathrm{FJ}(\bar{x})$ such that

$$\text{(SOS3)} \quad w^T \nabla^2_{xx} \tilde{L}(\bar{x}, \lambda_0, \lambda) w > 0.$$

It is clear that (SOS1) $\Longrightarrow$ (SOS2) $\Longrightarrow$ (SOS3). Note that the (SOSi) and the (SONi) with $i = 1, 2, 3$ have no gap in the sense that the only change between them is between a strict and non-strict inequality. The (SOS1) was first considered by McCormick [109] who actually showed that $\bar{x}$ is a strict local minimum under the (SOS1), see also the text books [50, 21, 60, 16, 120]. The (SOS1) was slightly extended to an FJ-type condition by Han and Mangasarian [66] who showed that if there exists some $(\bar{\lambda}_0, \bar{\lambda}) \in \mathrm{FJ}(\bar{x})$ such that

$$w^T \nabla^2_{xx} \tilde{L}(\bar{x}, \bar{\lambda}_0, \bar{\lambda}) w > 0 \quad \forall w \in \mathcal{V}(\bar{x}) \backslash \{0\},$$

then $\bar{x}$ is a strict local minimum of (NLP). They demonstrated by an example that the (SOS1) may not applied because $\mathrm{KKT}(\bar{x}) = \emptyset$, but the generalized result can be applied. The (SOS2) was studied by Ioffe [83], Ben-Israel et al. [18], Rockafellar [138, 139], Burke [26, 24], Burke and Poliquin [31]. The (SOS3) was given in Proposition 5.48 of [23], see also [19, 20].

Optimality conditions and exact penalty functions

As pointed out by Rockafellar and Wets [141, p.35], penalty functions have some earlier history in association with numerical methods, see e.g., [40, 2, 143, 32], but in optimization they were popularized by Fiacco and McCormick's book [50]. Over the last sixty years, penalty functions have been extensively investigated both theoretically and practically. In the vast literature on penalty functions, much attention has been paid to the notion of exact penalization. This notion can be well explained in the context of nonlinear programming from the rich connections between optimality conditions of (NLP) and exact penalty functions that are attached with (NLP).

There can be found in the literature many various penalty functions associated with (NLP), most of which fall into a category of functions with the form

$$P(x,\mu) := f(x) + \mu Q(\|v(x)\|), \tag{1.1.7}$$

where μ, the penalty parameter, is a nonnegative real number,

$$v(x) := (\max\{0, g_1(x)\}, \ldots, \max\{0, g_m(x)\}, h_{m+1}(x), \ldots, h_{m+q}(x))^T \quad \forall x \in R^n,$$

$\|\cdot\|$ is a vector norm in R^{m+q}, and $Q : R_+ \to R_+$ is a nonnegative function with the property that $Q(t) = 0$ if and only if $t = 0$. By setting $Q(t) = t^2$ and using the ℓ_2 norm, we obtain from $P(x,\mu)$ the classical quadratic penalty function

$$f(x) + \mu\left(\sum_{i\in I}(g_i(x)_+)^2 + \sum_{j\in J}|h_j(x)|^2\right),$$

which dates back to an idea of Courant (1943) [40, 50] and has been fully developed in the book by Fiacco and McCormick [50]. By setting $Q(t) = t$ and using the ℓ_1 norm, we obtain from $P(x,\mu)$ the well-known l_1 penalty function

$$f(x) + \mu\left(\sum_{i\in I} g_i(x)_+ + \sum_{j\in J}|h_j(x)|\right), \tag{1.1.8}$$

which was first introduced by Eremin [48] and at essentially the same time by Zangwill [169], as pointed out by Burke [26]. Han and Mangasarian [66] employed the penalty function $P(x,\mu)$ with an additional requirement on Q as follows:

$$0 < Q'(0+) := \lim_{t\to 0+}\frac{Q(t) - Q(0)}{t} < +\infty. \tag{1.1.9}$$

The penalty function $P(x,\mu)$ is said to be *exact* at a local minimum $\bar{x}$ of (NLP), if $\bar{x}$ is an unconstrained local minimum of $P(x,\mu)$ for all sufficiently large but finite values of μ. For short, this property is referred to as *exact penalization*. The question as to under what circumstances the penalty function $P(x, \mu)$ is exact at local minima of (NLP), is the main concern of many research work. Note that exactness of $P(x, \mu)$ using a specific norm implies exactness for all other norms as well. Moreover, Han and Mangasarian ([66], Theorem 4.2) showed that the exactness of $P(x, \mu)$ does not depend on the specific form of Q as long as the property (1.1.9) is satisfied. Therefore, exactness of the classical l_1 penalty function amounts to exactness of many other penalty functions. The central roles that the l_1 penalty functions play in the theory of constrained optimization have been comprehensively investigated by Burke [26] and many references therein.

Among various regularity conditions for exact penalization, the notion of calmness, originally formulated by Rockafellar and first appearing in the paper by Clarke [35], can be utilized to give some full characterizations of exact penalization. In general terms, calmness can be described as a basic regularity condition under which we can study the sensitive properties of certain variational systems. Since the appearance, the notion of calmness has been extensively used in the literature, see [36, 136, 25, 154, 45, 152]. To be precise, we consider the perturbed nonlinear programming problems

$$\begin{array}{lll} \mathrm{NLP}(y) & \min & f(x) \\ & \text{s.t.} & x \in M(y), \end{array}$$

where M is a set-valued mapping defined by (1.1.2). Let $\bar{y} = 0 \in R^{m+q}$. It is clear that $\bar{x} \in M(\bar{y}) = C$ and that $\mathrm{NLP}(\bar{y})$ is exactly the same with (NLP). According to Burke [25], the problem $\mathrm{NLP}(\bar{y})$ is said to be calm at $\bar{x}$ if there exist a number $\bar{\mu} \geq 0$ and a neighborhood U of $\bar{x}$ such that,

$$f(x) + \bar{\mu}\|y - \bar{y}\| \geq f(\bar{x}) \quad \forall x \in M(y) \cap U,$$

where $\|z\|$ stands for the norm of z in R^{m+q} and we can specify the norm to be the ℓ_1 norm without loss of generality. Note that this definition of calmness varies from Definition 6.4.1 of Clarke [36] in that the variable y is not restricted to a neighborhood of $\bar{y}$ in order for the above inequality to hold. It was shown by [25] that the restriction on the perturbation y is redundant when the functions g_i and h_j are continuous. Calmness can also be defined independent of the existence of a local minimum of $\mathrm{NLP}(\bar{y})$. The problem $\mathrm{NLP}(y)$ is said to be calm at $\bar{y}$ if

$$\liminf_{y \to \bar{y}} \frac{V(y) - V(\bar{y})}{\|y - \bar{y}\|} > -\infty,$$

where $V : R^{m+q} \to R \cup \{\pm\infty\}$ is the value function defined by

$$V(y) = \inf\{f(x) \mid x \in M(y)\}.$$

If $M(y)$ is an empty set, then $V(y)$ is assigned the value $+\infty$. It is easy to see that if the problem NLP(y) is said to be calm at $\bar{y}$, then for any global minimum x of the problem NLP(y), the problem NLP($\bar{y}$) is calm at x, see Proposition 6.4.2 [36] and Proposition 2.2 of [25]. Clarke [36] showed that calmness of NLP($\bar{y}$) implies exactness of penalty functions, while the reverse implication was first established by Burke [25, 26]. Therefore, the notion of calmness is in a sense equivalent to the notion of exact penalization.

Howe [76] showed that if there exists no nonzero $w \in \mathcal{V}(\bar{x})$ or equivalently the first-order sufficient condition (1.1.1) for (NLP) holds at $\bar{x}$, then $\bar{x}$ is a strict local minimum of the l_1 penalty function for all μ sufficiently large. Rosenberg [142] extended Howe's result to the Lipschitzian case and provided a sharp lower bound for all exact penalty parameters. Similar result can be found in [26, 14].

Note that the calmness property of the problem NLP($\bar{y}$) and Howe's result both rely on the objective function f. Other regularity conditions for exact penalization that are independent of the objective function f, can be found in the literature. Han and Mangasarian [66] showed that, if $\bar{x}$ is a strict local minimum of (NLP) and the MFCQ holds at $\bar{x}$, then $\bar{x}$ is a local minimum of the l_1 penalty function for all μ sufficiently large. Pietrzykowski [128] obtained the same result by assuming the LICQ, a stronger condition than the MFCQ. Lasserre [97] also employed the LICQ but provided a slightly different result from the corresponding result of Han and Mangasarian, and Pietrzykowski, since Lasserre did not assume that the local minimum $\bar{x}$ is a strict one. It follows from Clarke's elementary exact penalty penalization theorem ([36], Proposition 2.4.3) that, the l_1 penalty function is exact at $\bar{x}$ provided that the local error property (1.1.4) holds at $\bar{x}$, see also Proposition 3.111 of [23] and Corollary 2.6 of [147]. This fundamental result indicates that the QNCQ or any stronger CQ is sufficient for the exactness of the l_1 penalty function, since the QNCQ implies the local error property (1.1.4).

Exactness of the l_1 penalty function is also closely related to second-order sufficient and necessary conditions of (NLP). Han and Mangasarian [66] essentially showed that, if the (SOS1) holds at a feasible point $\bar{x}$ with respect to some $\bar{y} \in \mathrm{KKT}(\bar{x})$, then $\bar{x}$ is a strict local minimum of the l_1 penalty function for all $\mu > \|\bar{y}\|_\infty$. This result subsumes and sharpens the result by Charalambous [33] who considered the problem (NLP) with inequality constraints only. Similar result can be found in Lasserre [97]. Parallel to the results of Charalambous, Han and Mangasarian, and Lasserre, a more elegant result obtained by Burke ([26], Theorem 4.7) asserts that for all $\mu > \|\bar{y}\|_\infty$, $\bar{x}$ is a strict local minimum of order 2 for the l_1 penalty function.

It is well-known that if the l_1 penalty function is exact at $\bar{x}$, then both the KKT condition and the (SON2) hold at $\bar{x}$, see in particular Han and Mangasarian ([66], Theorem 4.8) and Rockafellar ([139], Corollary 4.5). This indicates that the l_1 penalty function is qualified for detecting both the KKT condition and the (SON2) in the sense that any condition ensuring the exactness of the l_1 penalty function at some local minimum of (NLP), guarantees the KKT condition and the (SON2) at this local minimum.

Thanks to the efforts of many researchers, the equivalence of first- and second-order optimality conditions for (NLP) and the l_1 exact penalty function is now well understood. But there is another type of exact penalty functions whose optimality conditions have no direct connections with those of (NLP), such as the so-called lower order l_p $(0 < p < 1)$ penalty function

$$f(x) + \mu \left(\sum_{i \in I} g_i(x)_+ + \sum_{j \in J} |h_j(x)| \right)^p \tag{1.1.10}$$

which is obtained from $P(x, \mu)$ by setting $Q(t) = t^p$ and using the ℓ_1 norm. Note that when $p = 1$, the above function reduces to the l_1 penalty function. This type of penalty functions was first introduced in Luo et al. [105] for the study of mathematical programs with equilibrium constraints, and has been studied extensively in [80, 145, 160, 161, 110, 162]. Note that the l_p $(0 < p < 1)$ penalty function is, in general, non-Lipschitz because it is defined via the function $t^p : R_+ \to R_+$ which is non-Lipschitz at $t = 0$ relative to R_+.

Similarly as the case for the l_1 penalty function, it is easy to show that the l_p penalty function is exact at $\bar{x}$ if and only if there exist a number $\bar{\mu} \geq 0$ and a neighborhood U of $\bar{x}$ such that,

$$f(x) + \bar{\mu}\|y - \bar{y}\|^p \geq f(\bar{x}) \quad \forall x \in M(y) \cap U.$$

The latter property was referred to as the generalized calmness-type conditions by Rubinov and Yang in their book [144], where the global version of the generalized calmness-type conditions by virtue of value functions is also discussed. The general exact penalty result for subanalytic systems ([105], Theorem 2.1.2) asserts that if all functions in defining (NLP) are continuous subanalytic, then there exists some $p \in (0, 1]$ such that the l_p penalty function is exact at $\bar{x}$. If the l_p penalty function is exact at $\bar{x}$, then for any $0 < p' < p$, the $l_{p'}$ penalty function is also exact at $\bar{x}$.

Although the l_p penalty functions have a greater chance to be exact than the l_1 penalty function, their exactness, however, does not in general imply the KKT conditions. Consider the simple problem of minimizing $-x$ subject to $x^2 \leq 0$, for which the KKT condition does not hold at the local minimum $x = 0$. The l_1 penalty function for this problem is not exact at $x = 0$, but the l_p (with $p = 0.5$) penalty function is exact at $x = 0$. Therefore, not every l_p exact penalty function is qualified for detecting KKT conditions. Yang and Meng [161] showed that if a type of conditions in terms of (generalized) second-order derivatives of the constraints is satisfied, then KKT conditions can be derived from the l_p exact penalty functions. Their technique is conducted by first applying (generalized) Taylor expansions to estimate the Dini upper-directional derivatives of the l_p exact penalty functions, and then by using the Farkas' Lemma.

We end this section by emphasizing that the terminology 'exact penalization' appeared in the literature may be different from the one that we previously reviewed. The exact penalty functions found in [46, 59, 67, 129, 39] are distinguished by their differentiability, while exact penalty functions having the form $P(x, \mu)$ are commonly believed to be non-differentiable. The exact penalty functions appeared in the context of augmented Lagrangian theory are also different because the penalty terms associated with augmented Lagrangian functions could take negative values. The augmented Lagrangian method was proposed by Hestenes [74] and Powell [130] for the equality constrained problem, and extended to the inequality constrained problem by Rockafellar [135, 140]. The augmenting functions, referred to as additional penalty terms, considered in these work are quadratic. Further developments have been done by Rockafellar and Wets [141] in connection with convex augmenting functions, by Huang and Yang [80] in connection with level-boundedness augmenting functions, by Zhou and Yang [175, 176] in connection with valley-at-0 augmenting functions. For an overview of modified Lagrangians and their usage in numerical optimization, see Bertsekas [21] and Nocedal and Wright [120].

1.2 Review on Mathematical Programs with Complementarity Constraints

Consider the mathematical program with complementarity constraints

$$\begin{array}{lll} \text{(MPCC)} & \min & f(x) \\ & \text{s.t.} & g_i(x) \leq 0, & i \in I, \\ & & h_j(x) = 0, & j \in J, \\ & & G_k(x) \geq 0, H_k(x) \geq 0, G_k(x)H_k(x) = 0, & k \in K, \end{array}$$

where f, I, J, g_i, h_j are given as in (NLP), $K = \{m+q+1, \cdots, m+q+l\}$, and G_k, H_k are all assumed to be continuously differentiable functions from R^n to R. Let $\bar{x}$ be a fixed feasible point of (MPCC). The index sets $I(\bar{x}), \alpha, \beta, \gamma$ depending on $\bar{x}$ are defined as follows:

$$\begin{aligned} I(\bar{x}) &= \{k \in I \,|\, g_i(\bar{x}) = 0\}, \\ \alpha &= \{k \in K \,|\, 0 = G_k(\bar{x}) < H_k(\bar{x})\}, \\ \beta &= \{k \in K \,|\, G_k(\bar{x}) = H_k(\bar{x}) = 0\}, \\ \gamma &= \{k \in K \,|\, G_k(\bar{x}) > H_k(\bar{x}) = 0\}. \end{aligned}$$

For a given set A, the set of all partitions of A is given by

$$\mathcal{P}(A) = \{(A_1, A_2) \mid A_1 \cup A_2 = A,\ A_1 \cap A_2 = \emptyset\}.$$

It is clear to see that (MPCC) is essentially an (NLP) problem. Moreover, (MPCC) is closely related to many other (NLP) problems with different various forms. First, we can associate with (MPCC) two (NLPs) [146], the *tightened* NLP

$$\begin{array}{lll} \text{(TNLP)} & \min & f(x) \\ & \text{s.t.} & g_i(x) \leq 0, & i \in I, \\ & & h_j(x) = 0, & j \in J, \\ & & G_k(x) = 0, H_k(x) \geq 0 & k \in \alpha, \\ & & G_k(x) \geq 0, H_k(x) = 0, & k \in \gamma, \\ & & G_k(x) = 0, H_k(x) = 0, & k \in \beta, \end{array}$$

and the *relaxed* NLP

$$
\begin{array}{lll}
\text{(RNLP)} & \min & f(x) \\
& \text{s.t.} & g_i(x) \leq 0, \qquad\qquad\quad i \in I, \\
& & h_j(x) = 0, \qquad\qquad\quad j \in J, \\
& & G_k(x) = 0, H_k(x) \geq 0 \quad k \in \alpha, \\
& & G_k(x) \geq 0, H_k(x) = 0, \quad k \in \gamma, \\
& & G_k(x) \geq 0, H_k(x) \geq 0, \quad k \in \beta.
\end{array}
$$

Then, for each partition $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$, we can associate with (MPCC) an (NLP) as follows [54, 165]:

$$
\begin{array}{lll}
\text{NLP}(\beta_1, \beta_2) & \min & f(x) \\
& \text{s.t.} & g_i(x) \leq 0, \qquad\qquad\quad i \in I, \\
& & h_j(x) = 0, \qquad\qquad\quad j \in J, \\
& & G_k(x) = 0, H_k(x) \geq 0 \quad k \in \alpha, \\
& & G_k(x) \geq 0, H_k(x) = 0, \quad k \in \gamma, \\
& & G_k(x) \geq 0, H_k(x) = 0, \quad k \in \beta_1, \\
& & G_k(x) = 0, H_k(x) \geq 0, \quad k \in \beta_2.
\end{array}
$$

Next, by the so-called NCP-function [51] $\phi : R^2 \to R$ with the property that $\phi(a, b) = 0$ if and only if $a \geq 0, b \geq 0, ab = 0$, we can reformulate (MPCC) as an (NLP) as follows [146, 160]:

$$
\begin{array}{lll}
\text{(MPCC)}_\phi & \min & f(x) \\
& \text{s.t.} & g_i(x) \leq 0, \qquad\qquad\quad i \in I, \\
& & h_j(x) = 0, \qquad\qquad\quad j \in J, \\
& & \phi(G_k(x), H_k(x)) = 0 \quad k \in K.
\end{array}
$$

Note that the composition function $\phi(G_k(x), H_k(x))$ may be non-differentiable even if the functions G_k and H_k are sufficiently smooth, since the NCP-functions may be non-differentiable at the origin, such as $\phi(a, b) := \min(a, b)$ or $\phi(a, b) := a + b - \sqrt{a^2 + b^2}$. Finally, by introducing slack variables r and s, (MPCC) can be also reformulated as an (NLP) with abstract constraints as follows [165]:

$$
\begin{array}{lll}
\text{(MPCC)}_\Omega & \min & f(x) \\
& \text{s.t.} & g_i(x) \leq 0, \qquad\qquad\qquad\qquad\quad i \in I, \\
& & h_j(x) = 0, \qquad\qquad\qquad\qquad\quad j \in J, \\
& & G_k(x) - r_k = 0, H_k(x) - s_k = 0, \quad k \in K, \\
& & (r_k, s_k) \in \Omega, \qquad\qquad\qquad\qquad k \in K,
\end{array}
$$

where $\Omega = \{(a,b) \in R^2 \mid a \geq 0, b \geq 0, ab = 0\}$.

Denote by E, $\mathcal{F}_{\text{TNLP}}$, $\mathcal{F}_{\text{RNLP}}$, and $\mathcal{F}_{\text{NLP}(\beta_1,\beta_2)}$ the feasible sets of the problems (MPCC), (TNLP), (RNLP), and NLP(β_1, β_2), respectively. Then,

$$\mathcal{F}_{\text{TNLP}} = \bigcap_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{F}_{\text{NLP}(\beta_1,\beta_2)} \subset \mathcal{F}_{\text{NLP}(\beta_1,\beta_2)} \subset E = \bigcup_{(\beta_1,\beta_2)\in\mathcal{P}(\beta)} \mathcal{F}_{\text{NLP}(\beta_1,\beta_2)} \subset \mathcal{F}_{\text{RNLP}}. \tag{1.2.11}$$

holds on some neighborhood of $\bar{x}$, see [146]. Based on these relations, $\bar{x}$ is a local minimum of (MPCC) if and only if it is a local minimum of program NLP(β_1, β_2) for each $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$. If $\bar{x}$ is a local minimum of (RNLP) then it is a local minimum of (MPCC), and if $\bar{x}$ is a local minimum of (MPCC) then it is a local minimum of (TNLP). The reverse implications hold in general only if strict complementarity holds at $\bar{x}$ (i.e., $\beta = \emptyset$), see [146]. In this case, $\mathcal{P}(\beta) = \{(\emptyset, \emptyset)\}$, the problems (TNLP), (RNLP), and NLP$(\emptyset, \emptyset)$ are the same, and equality holds throughout (1.2.11) on some neighborhood of $\bar{x}$. In general, the (MPCC) with $\beta \neq \emptyset$ is more difficult to deal with than the (MPCC) with $\beta = \emptyset$.

Since (MPCC) has close connections with different NLPs, many concepts and methods known from the nonlinear programming literature have been used to study (MPCC). The various constraint qualifications ensuring the KKT conditions (also known as the strong stationary conditions in the MPCC literature) have been studied, see [16, 146]. The sequential quadratic programming method for (MPCC) can be found in [89, 61], and the sequential penalization approach has been investigated in [81]. More importantly and frequently, variants of these concepts and methods, which are tailored specifically for (MPCC), have been used in the (MPCC) community. In what follows, we will mainly focus on various constraint qualifications and stationarity concepts for (MPCC), and survey briefly on penalty methods for (MPCC).

Standard CQs for (MPCC)

As mentioned earlier, the (MPCC) is essentially a nonlinear programming problem. But it is a well-known fact that most of the familiar CQs known from nonlinear programming literature do not hold, see [34, 167, 56]. In particular, the linear CQ is violated because the constraints of (MPCC) cannot be all affine functions. The MFCQ is also violated at every feasible point of (MPCC), see [167, 146] and Proposition 2.15 of [52] for a detailed proof. Thus, any CQ stronger than the MFCQ, including in particular the LICQ, cannot be fulfilled at any feasible point of (MPCC). However, the LICQ and the MFCQ for (TNLP) may have a great chance to be fulfilled because of the absence of complementarity constraints, and they are respectively called the MPCC LICQ and the MPCC MFCQ, and were first considered by Scheel and Scholtes [146] for deriving optimality conditions for (MPCC). For various applications of the MPCC LICQ in connection with numerical methods for (MPCC), see [77, 78, 160, 81, 102].

If there exist some $k \in \beta$ and a sequence $x_\nu \to \bar{x}$ such that $G_k(x_\nu)H_k(x_\nu) \neq 0$ for all ν, then by definition the (QNCQ) does not hold at $\bar{x}$. This indicates that the (QNCQ) can be easily violated for (MPCC).

The ACQ holds at $\bar{x}$ for (MPCC) if by definition $T_E(\bar{x}) = T^{\text{lin}}(\bar{x})$, where $T_E(\bar{x})$ is the (Bouligand) tangent cone to E at $\bar{x}$, and $T^{\text{lin}}(\bar{x})$ is the first-order linearized cone of (MPCC) at $\bar{x}$ which can be explicitly expressed as

$$T^{\text{lin}}(\bar{x}) = \left\{ u \in R^n \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^T u \leq 0, & i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T u = 0, & j \in J \\ \nabla G_k(\bar{x})^T u = 0, & k \in \alpha \\ \nabla H_k(\bar{x})^T u = 0, & k \in \gamma \\ \nabla G_k(\bar{x})^T u \geq 0, & k \in \beta \\ \nabla H_k(\bar{x})^T u \geq 0, & k \in \beta \end{array} \right\}.$$

One drawback of the ACQ is that it can never be satisfied when $T_E(\bar{x})$ is non-convex, which is rather common for (MPCC) due to the existence of the complementarity constraints.

The ACQ is closely related with the notion of piecewise ACQ first introduced by Pang and Fukushima [125], which is satisfied at $\bar{x}$ by definition if the ACQ holds at $\bar{x}$ for each NLP(β_1, β_2) with $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$. For each $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$, let the first-order linearized cone of NLP(β_1, β_2) at $\bar{x}$ be given by

$$T^{\text{lin}}_{(\beta_1,\beta_2)}(\bar{x}) = \left\{ u \in R^n \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^T u \le 0, & i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T u = 0, & j \in J \\ \nabla G_k(\bar{x})^T u = 0, & k \in \alpha \cup \beta_1 \\ \nabla H_k(\bar{x})^T u = 0, & k \in \gamma \cup \beta_2 \\ \nabla G_k(\bar{x})^T u \ge 0, & k \in \beta_2 \\ \nabla H_k(\bar{x})^T u \ge 0, & k \in \beta_1 \end{array} \right\}.$$

Under the piecewise ACQ at $\bar{x}$, the ACQ holds at $\bar{x}$, if and only if, among all the subsets $T^{\text{lin}}_{(\beta_1,\beta_2)}(\bar{x})$ of $T^{\text{lin}}(\bar{x})$ with $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$, there exists at least one which is equal to $T^{\text{lin}}(\bar{x})$ and hence the biggest one, see Flegel and Kanzow [56]. As can be easily seen from the simple complementarity constraints: $x_1 \ge 0$, $x_2 \ge 0$, $x_1x_2 = 0$, the ACQ does hold at $\bar{x} := (0, 0)^T$ though the piecewise ACQ is satisfied at $\bar{x}$. Note that the piecewise ACQ is a very weak assumption, because each NLP(β_1, β_2) with $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$ is merely an ordinary nonlinear programming for which the ACQ is commonly believed [16] to be weak enough.

The ACQ is also closely related with the notion of nonsingularity for linear systems, which is also introduced by Pang and Fukushima [125]. Consider the linear system

$$Ax \le b, \quad Cx = d. \tag{1.2.12}$$

The inequality $A_i x \le b_i$, where A_i is the i-th row of the matrix A and b_i is the i-th component of the vector b, is said to be nonsingular if there exists a feasible solution of the system (1.2.12) which satisfies this inequality strictly. Denote by β^G (respectively, β^H) the subset of β consisting of all indices $k \in \beta$ such that the inequality $\nabla G_k(\bar{x})^T u \ge 0$ (respectively, $\nabla H_k(\bar{x})^T u \ge 0$) is nonsingular in the system defining $T^{\text{lin}}(\bar{x})$. If $\beta^G \cap \beta^H = \emptyset$ and the MPCC MFCQ holds at $\bar{x}$, then the ACQ holds at $\bar{x}$, see [125].

The GCQ holds at $\bar{x}$ for (MPCC) if by definition $T_E(\bar{x})^* = T^{\text{lin}}(\bar{x})^*$ or equivalently $T^{\text{lin}}(\bar{x}) = \text{clconv}T_E(\bar{x})$. Due to the close convex hull operation, the GCQ has a greater chance to be satisfied than the ACQ. Let the MPCC-linearized cone of (MPCC) at $\bar{x}$

be given by

$$T^{\rm lin}_{\rm MPCC}(\bar x) = \left\{ u \in R^n \;\middle|\; \begin{array}{ll} \nabla g_i(\bar x)^T u \le 0, & i \in I(\bar x) \\ \nabla h_j(\bar x)^T u = 0, & j \in J \\ \nabla G_k(\bar x)^T u = 0, & k \in \alpha \\ \nabla H_k(\bar x)^T u = 0, & k \in \gamma \\ \nabla G_k(\bar x)^T u \ge 0, & k \in \beta \\ \nabla H_k(\bar x)^T u \ge 0, & k \in \beta \\ (\nabla G_k(\bar x)^T u)(\nabla H_k(\bar x)^T u) = 0, & k \in \beta \end{array} \right\}.$$

This linearized cone was first introduced in [146, 125] and later studied extensively in [54, 165, 57]. It is shown by Corollary 3.20 of [52] that the GCQ holds at $\bar x$ if and only if $T_E(\bar x)^* = T^{\rm lin}_{\rm MPCC}(\bar x)^*$ and $T^{\rm lin}_{\rm MPCC}(\bar x)^* = T^{\rm lin}(\bar x)^*$. The latter equality was referred as to the intersection property (IP) for (MPCC) at $\bar x$ by Flegel in his PhD thesis [52], see also [58]. It is shown by Lemma 3.22 of [52] that the IP is implied by the assumption (A2) of [125], which is said to be satisfied at $\bar x$ by definition if there exists a partition $(\beta_1^{GH}, \beta_2^{GH}) \in \mathcal{P}(\beta^G \cap \beta^H)$ such that the equality

$$\sum_{i\in I(\bar x)} \lambda_i^g \nabla g_i(\bar x) + \sum_{j\in J} \lambda_j^h \nabla h_j(\bar x) - \sum_{k\in\alpha\cup\beta} \lambda_k^G \nabla G_k(\bar x) - \sum_{k\in\gamma\cup\beta} \lambda_k^H \nabla H_k(\bar x) = 0$$

implies that $\lambda_k^G = 0$ for all $k \in \beta_1^{GH}$ and $\lambda_k^H = 0$ for all $k \in \beta_2^{GH}$. The assumption (A2) of [125] is clearly implied by a stronger condition, called the partial MPCC LICQ [165], which is said to be satisfied at $\bar x$ if by definition the equality

$$\sum_{i\in I(\bar x)} \lambda_i^g \nabla g_i(\bar x) + \sum_{j\in J} \lambda_j^h \nabla h_j(\bar x) - \sum_{k\in\alpha\cup\beta} \lambda_k^G \nabla G_k(\bar x) - \sum_{k\in\gamma\cup\beta} \lambda_k^H \nabla H_k(\bar x) = 0$$

implies that $\lambda_k^G = \lambda_k^H = 0$ for every $k \in \beta$. It is shown by Theorem 4.6 of [56] that the GCQ holds at $\bar x$ if the MPCC LICQ holds at $\bar x$.

MPCC tailored CQs

Having checked several CQs known from the nonlinear programming literature, we have seen that most of the standard CQs including the LICQ and the MFCQ, are violated for the (MPCC). Therefore, the CQs tailored specifically to (MPCC) are needed, some of which are defined as follows.

The piecewise MFCQ [165, 121] holds at $\bar{x}$ if the MFCQ holds at $\bar{x}$ for each NLP(β_1, β_2) with $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$.

The MPCC generalized MFCQ (MPCC GMFCQ) [165] holds at $\bar{x}$ if there is no nonzero vector $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in R^{m+q+2l}$ such that

$$\begin{gathered}
\sum_{i \in I} \lambda_i^g \nabla g_i(\bar{x}) + \sum_{j \in J} \lambda_j^h \nabla h_j(\bar{x}) - \sum_{k \in K} [\lambda_k^G \nabla G_k(\bar{x}) + \lambda_k^H \nabla H_k(\bar{x})] = 0, \\
\forall i \in I, \lambda_i^g \geq 0, \lambda_i^g g_i(\bar{x}) = 0, \\
\forall k \in \gamma, \lambda_k^G = 0, \quad \forall k \in \alpha, \lambda_k^H = 0, \\
\forall k \in \beta, \text{ either } \lambda_k^G > 0, \lambda_k^H > 0 \text{ or } \lambda_k^G \lambda_k^H = 0.
\end{gathered} \tag{1.2.13}$$

The MPCC generalized pseudonormality (MPCC GPNCQ)[92] holds at $\bar{x}$ if there are no nonzero vector $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in R^{m+q+2l}$ and a sequence $x_\nu \to \bar{x}$ such that (1.2.13) holds and for all ν,

$$\sum_{i \in I} \lambda_i^g g_i(x_\nu) + \sum_{j \in J} \lambda_j^h h_j(x_\nu) - \sum_{k \in K} [\lambda_k^G G_k(x_\nu) + \lambda_k^H H_k(x_\nu)] > 0.$$

The MPCC generalized quasinormality (MPCC GQNCQ) [92] holds at $\bar{x}$ if there are no nonzero vector $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in R^{m+q+2l}$ and a sequence $x_\nu \to \bar{x}$ such that (1.2.13) holds and for all ν, $\lambda_i^g g_i(x_\nu) > 0$ for all i with $\lambda_i^g > 0$, $\lambda_j^h h_j(x_\nu) > 0$ for all j with $\lambda_j^h \neq 0$, $-\lambda_k^G G_k(x_\nu) > 0$ for all k with $\lambda_k^G \neq 0$, and $-\lambda_k^H H_k(x_\nu) > 0$ for all k with $\lambda_k^H \neq 0$.

The local MPPC error bound [55, 165] holds at $\bar{x}$ if there exist some $\mu > 0$, a neighborhood W of the origin of R^{m+q+2l} and a neighborhood V of $\bar{x}$ such that

$$d(x, E) \leq \mu \|(u, v, r, s)\| \quad \forall (u, v, r, s) \in W, \ \forall x \in \mathcal{Z}(u, v, r, s) \cap V,$$

where $\mathcal{Z} : R^{m+q+2l} \rightrightarrows R^n$ is a set-valued mapping defined by

$$\mathcal{Z}(u, v, r, s) := \left\{ x \in R^n \;\middle|\; \begin{array}{c} g_i(x) \leq u_i, i \in I, \quad h_j(x) = v_j, j \in J \\ G_k(x) + r_k \geq 0, H_k(x) + s_k \geq 0, \ k \in K \\ (G_k(x) + r_k)(H_k(x) + s_k) = 0, \ k \in K \end{array} \right\}, \tag{1.2.14}$$

which can be considered as a perturbation of the feasible set of the (MPCC) due to $\mathcal{Z}(0, 0, 0, 0) = E$.

The MPCC linear CQ [165, 54] holds if the functions g_i, h_j, G_k, H_k are all affine.

The MPCC-ACQ [54, 165] holds at $\bar{x}$ if $T_E(\bar{x}) = T^{\text{lin}}_{\text{MPCC}}(\bar{x})$.

The MPCC-GCQ [52, 57] holds at $\bar{x}$ if $T_E(\bar{x})^* = T^{\text{lin}}_{\text{MPCC}}(\bar{x})^*$.

The property of local MPCC error bound is actually defined via the calmness of the set-valued mapping $\mathcal{Z}$ at $(0, \bar{x})$, which in an alternative way [73] can be expressed as the existence of some $\tau > 0$ and a neighborhood V of $\bar{x}$ such that

$$\tau d_E(x) \leq \left(S(x) + \sum_{k \in K} |\min\{G_k(x), H_k(x)\}| \right) \quad \forall x \in V,$$

where $d_E(x)$ is the distance of the point x from E, and S is given by (1.1.3). Equivalently, the right-hand side (induced from the ℓ_∞ norm in R^{m+q+2l}) of the above inequality can be replaced by many other functions, such as

$$S(x) + \sum_{k \in K} \max\{-G_k(x), -H_k(x), -(G_k(x) + H_k(x)), \min\{G_k(x), H_k(x)\}\},$$

which is induced from the ℓ_1 norm in R^{m+q+2l}, see [92] for details.

The result that the local MPCC error bound holds at every feasible point of (MPCC) under the MPCC linear CQ, can be obtained by applying Robinson's well-known result ([134], Proposition 1) on a continuity property of polyhedral multifunctions, see Ye [165] and Flegel and Kanzow [55]. It can also be obtained in a direct way by exploited carefully the affine structure of (MPCC) under the MPCC linear CQ, see Lemma 3.1 of Meng and Yang [110]. It has been shown by Kanzow and Schwartz [92] that the local MPCC error bound holds at $\bar{x}$ if the MPCC GPNCQ holds at $\bar{x}$. However, by checking their proof ([92], Lemma 4.3), we find it that the local MPCC error bound actually holds at $\bar{x}$ under a weaker version of the MPCC GPNCQ, which happens to be equivalent with the MPCC GQNCQ by taking Proposition 3.2 of Bertsekas and Ozdaglar [22] into account. Therefore, the local MPCC error bound holds at $\bar{x}$ if the MPCC GQNCQ holds at $\bar{x}$.

According to [52, 54, 56, 165, 92] and the previous discussion, we have the following implications:

$$\begin{gathered} \text{MPCC LICQ} \Longrightarrow \text{MPCC MFCQ} \Longrightarrow \text{Piecewise MFCQ} \\ \Longrightarrow \text{MPCC GMFCQ} \Longrightarrow \text{MPCC GPNCQ} \Longrightarrow \text{MPCC GQNCQ} \\ \Longrightarrow \text{local MPPC error bound} \Longrightarrow \text{MPCC ACQ} \Longrightarrow \text{MPCC GCQ} \end{gathered}$$

and

$$\text{MPCC linear CQ} \Longrightarrow \text{local MPPC error bound.}$$

Stationarity conditions for (MPCC)

Stationarity (or first-order optimality) conditions of (MPCC) have been the subject of many recent papers and books, see [146, 125, 165, 105, 123] and references therein. Since there are several different approaches for deriving these conditions, various stationarity concepts arise, see a very recent PhD thesis [52] by Flegel for their definitions and connections.

According to [54], we say that $\bar{x}$ is a B-stationary point if $-\nabla f(\bar{x}) \in T_E(\bar{x})^*$, that it is an MPCC-linearized B-stationary point if $-\nabla f(\bar{x}) \in T_{\mathrm{MPCC}}^{\mathrm{lin}}(\bar{x})^*$, and that it is a linearized B-stationary point if $-\nabla f(\bar{x}) \in T^{\mathrm{lin}}(\bar{x})^*$. We have

$$\text{Linearized B-stationarity} \Longrightarrow \text{MPCC-linearized B-stationarity} \Longrightarrow \text{B-stationarity,}$$

due to

$$T^{\mathrm{lin}}(\bar{x})^* \subset T_{\mathrm{MPCC}}^{\mathrm{lin}}(\bar{x})^* \subset T_E(\bar{x})^*,$$

see [54]. B-stationarity was first proposed in [106], and studied in depth in [105, 125]. MPCC-linearized B-stationarity was first defined in [146], and later developed by many papers [165, 55]. Note that $\bar{x}$ is an MPCC-linearized B-stationary point if and only if the KKT condition holds at $\bar{x}$ for each NLP(β_1, β_2) with $(\beta_1, \beta_2) \in \mathcal{P}(\beta)$, see [146]. The notion of Linearized B-stationarity was introduced in [54] for the sake of completeness.

Besides these primal stationary conditions, there are several dual stationary conditions or KKT-type conditions developed for (MPCC), such as strong stationarity [125], M-stationarity [121], C-stationarity [146], A-stationarity [53, 54]), and weakly stationarity [146]. Here, we mainly focus on strong stationarity and M-stationarity. We say that $\bar{x}$ is a strongly stationary point [146, 125, 56] (respectively, an M-stationary point

[121, 122, 55, 165]) if, there is $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in R^{m+q+2l}$ such that

$$\nabla f(\bar{x}) + \sum_{i \in I} \lambda_i^g \nabla g_i(\bar{x}) + \sum_{j \in J} \lambda_j^h \nabla h_j(\bar{x}) - \sum_{k \in K} [\lambda_k^G \nabla G_k(\bar{x}) + \lambda_k^H \nabla H_k(\bar{x})] = 0,$$
$$\forall i \in I, \lambda_i^g \geq 0, \lambda_i^g g_i(\bar{x}) = 0,$$
$$\forall k \in \gamma, \lambda_k^G = 0, \quad \forall k \in \alpha, \lambda_k^H = 0,$$
$$\forall k \in \beta, \lambda_k^G \geq 0, \lambda_k^H \geq 0$$
$$(\text{respetively}, \forall k \in \beta, \text{ either } \lambda_k^G > 0, \lambda_k^H > 0 \text{ or } \lambda_k^G \lambda_k^H = 0).$$

Clearly, strong stationarity implies Mordukhovich stationarity. Note that $\bar{x}$ is a strongly stationary point if and only if the KKT condition holds at $\bar{x}$ for (MPCC) or (RNLP), see [54] for details. M-stationary condition (with 'M' standing for 'Mordukhovich') was first introduced in [166] for optimization problems with variational inequality constraints by using Mordukhovich's generalized differential calculus [113], and was further studied in [164] and [121]. It was shown by Theorem 2.3 of [165] that M-stationary condition is also sufficient for global or local optimality under certain MPCC generalized convexity condition.

In what follows, we assume that $\bar{x}$ is a local minimum of (MPCC). Then by the basic first-order conditions for optimality ([141], Theorem 6.12), we have

$$-\nabla f(\bar{x}) \in T_E(\bar{x})^*.$$

Therefore, any local minimum of (MPCC) is by definition a B-stationary point. If the GCQ holds at $\bar{x}$, then we have

$$-\nabla f(\bar{x}) \in T^{\text{lin}}(\bar{x})^*,$$

which implies by definition that $\bar{x}$ is a linearized B-stationary point, or equivalently a strongly stationary point by Farkas' lemma, see [54] for more details on this equivalence. Note that the GCQ is the weakest CQ for strong stationarity, as is the case for (NLP), see [64] and also Theorem 6.11 of [141].

To show the M-stationarity under the MPCC-GCQ, an MPCC variant of the notion of calmness has been considered in [164, 55]. (MPCC) is said to be MPCC-calm at $\bar{x}$ if there exist $\mu > 0$, a neighborhood W of the origin of R^{m+q+2l} and a neighborhood V of $\bar{x}$ such that

$$f(\bar{x}) \leq f(x) + \mu \|(u, v, r, s)\| \quad \forall (u, v, r, s) \in W, \ \forall x \in \mathcal{Z}(u, v, r, s) \cap V,$$

where $\mathcal{Z}$ is the set-valued mapping defined by (1.2.14). If (MPCC) is MPCC-calm at $\bar{x}$, then $\bar{x}$ is an M-stationary point, see [55] and also [165]. If the local MPCC-error bound holds at a feasible point x of (MPCC), then the (MPCC) is MPCC-calm at x provided that x is a local minimum of (MPCC). Recall that the MPCC linear CQ implies the local MPCC-error bound.

If the MPCC-GCQ holds at $\bar{x}$, then

$$-\nabla f(\bar{x}) \in T^{\text{lin}}_{\text{MPCC}}(\bar{x})^*, \tag{1.2.15}$$

which implies by definition that $\bar{x}$ is an MPCC-linearized B-stationary point. Note that (1.2.15) holds if and only if $u^* = 0$ is a local minimum of the problem

$$\begin{array}{ll} \min\limits_{u \in R^n} & \nabla f(\bar{x})^T u \\ \text{s.t.} & u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x}), \end{array} \tag{1.2.16}$$

which satisfies the MPCC linear CQ because all functions in defining $T^{\text{lin}}_{\text{MPCC}}(\bar{x})$ are affine. Thus, $u^* = 0$ is an M-stationary point of (1.2.16), which amounts to that $\bar{x}$ is an M-stationary point of the original (MPCC). This explains how M-stationarity can be derived from the MPCC-GCQ by means of MPCC-calmness, see [165, 57] for more details on the original idea and the proof.

Note that MPCC-linearized B-stationarity implies M-stationarity but not vice versa, as shown by the following MPCC instance with $m = q = 0$, $n = l = 1$, and $\bar{x} = 0$:

$$\begin{array}{ll} \min & f(x) = -x \\ \text{s.t.} & G(x) = x \geq 0, H(x) = x^2 \geq 0, G(x)H(x) = 0. \end{array} \tag{1.2.17}$$

The MPCC instance (1.2.17) also indicates that the MPCC-GCQ is not the weakest CQ for M-stationarity because it is violated at $\bar{x}$ which, though, is an M-stationary point. It was shown, though not explicitly, in Flegel and Kanzow [57] that, the weakest CQ for M-stationarity in the sense that it is both sufficient and necessary for M-stationarity, can be defined as follows:

$$T_E(\bar{x})^* \times \{0\} \times \{0\} \subset N_{\Omega_1}(0,0,0) + N_{\Omega_2}(0,0,0), \tag{1.2.18}$$

where $N_A(y)$ stands for the normal cone [141] (also known as the Mordukhovich limiting normal cone [113]) to A at $y \in A$, Ω_1 and Ω_2 are two cones defined respectively by

$$\Omega_1 = \left\{ (u, \xi_\beta, \eta_\beta) \in R^{n+2|\beta|} \mid \xi_k \geq 0, \eta_k \geq 0, \xi_k \eta_k = 0, \ \forall k \in \beta \right\},$$

and

$$\Omega_2 = \left\{ (u, \xi_\beta, \eta_\beta) \in R^{n+2|\beta|} \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^T u \leq 0, & i \in I(\bar{x}), \\ \nabla h_j(\bar{x})^T u = 0, & j \in J, \\ \nabla G_k(\bar{x})^T u = 0, & k \in \alpha, \\ \nabla H_k(\bar{x})^T u = 0, & k \in \gamma, \\ \nabla G_k(\bar{x})^T u - \xi_k = 0, & k \in \beta, \\ \nabla H_k(\bar{x})^T u - \eta_k = 0, & k \in \beta. \end{array} \right\}$$

Flegel and Kanzow [57] derived M-stationarity from the MPCC GCQ by actually showing that the MPCC GCQ implies condition (1.2.18). Again, the MPCC instance (1.2.17) can be used to demonstrate that the condition (1.2.18) may be strictly weaker than the MPCC GCQ.

Exact penalization results for MPCC

Exact penalty results for (MPCC) are known in the literature [105, 106, 106, 147, 168, 55, 52]. In particular, Flegel and Kanzow [55] showed that the (MPCC) is MPCC-calm at $\bar{x}$ if and only if the penalty function for (MPCC) defined by

$$\mathcal{H}(x) = f(x) + \mu \left(S(x) + \sum_{k \in K} |\min\{G_k(x), H_k(x)\}| \right)$$

has a local minimum at $\bar{x}$ with some $\mu \geq 0$, where S is given by (1.1.3). This equivalence indicates that any condition that can imply the MPCC-calmness will be sufficient for the exactness of the penalty function $\mathcal{H}(x)$. Examples of such conditions are the local MPCC-error bound property [55, 165], the MPCC linear CQ [55, 165], and the MPCC GMFCQ [102]. If the penalty function $\mathcal{H}(x)$ is exact at $\bar{x}$, then $\bar{x}$ is an M-stationary point, see [55, 165].

When the (MPCC) is treated as an ordinary (NLP), the classical l_1 penalty function for (MPCC) can be expressed as follows:

$$\mathcal{G}(x) = f(x) + \mu \left(S(x) + \sum_{k \in K} \{(-G_k(x))_+ + (-H_k(x))_+ + |G_k(x) H_k(x)|\} \right),$$

where S is given by (1.1.3). If the penalty function $\mathcal{G}(x)$ is exact at $\bar{x}$, then $\bar{x}$ is a strongly stationary point or equivalently the KKT condition holds at $\bar{x}$, see [36, 35, 26]. Note that the MFCQ is invalid at every feasible point of (MPCC) and that local minima of (MPCC) may be merely M-stationary points but not strongly stationary points. This indicates that the exactness of the penalty function $\mathcal{G}(x)$ may requires somewhat stronger regularity conditions, such as the LICQ.

1.3 Notation

The notation that we employ in this thesis is for the most part borrowed from the book [141] by Rockafellar and Wets. A partial list is provided for the reader's convenience.

We denote by R the set of all real numbers, and set

$$\overline{R} := R \cup \{\pm\infty\}, \quad R_+ := \{t \in R \mid t \geq 0\}, \quad R_{++} := \{t \in R \mid t > 0\}.$$

For $a, b \in R$ with $a \leq b$, we denote by $[a, b]$ the closed interval between a and b, and by $[a, b)$ the half-closed and half-open interval between a and b, and by (a, b) the open interval between a and b. For vectors x, y in R^n, we denote by x^T the transpose of x, by $x^T y$ or $\langle x, y\rangle$ the inner product of x and y, by $x^\perp := \{v \mid \langle v, x\rangle = 0\}$ the orthogonal complement of x, and by $\|x\|$ the Euclidean norm of x. For a given subset A of R^n, we denote the closure of A, the interior of A, the boundary of A and the convex hull of A respectively by cl$A$, int$A$, bd$A$ and conv$A$. We say that the function $f : R^n \to R$ is $\mathcal{C}^k$ with k being a positive integer if f is k times continuously differentiable, and that f is $\mathcal{C}^{1,1}$ if f is differentiable with the gradient being locally Lipschitz. For a set-valued mapping $F : R^n \rightrightarrows R^m$, the graph of F can be identified as a subset of $R^n \times R^m$, namely

$$\mathrm{gph}F := \{(x, y) \mid y \in F(x)\}.$$

For a nonnegative function $f : R^n \to R_+ \cup \{+\infty\}$ and a scalar $p > 0$, the p-th order function of f is defined by

$$f^p(x) := (f(x))^p \qquad \forall x \in R^n, \tag{1.3.19}$$

where the convention $(+\infty)^p := +\infty$ is used. When $p = 0$, we interpret the function f^p as the indicator of the set $\{x \in R^n \mid f(x) = 0\}$.

In what follows, let A be a nonempty subset of R^n. We say that A is locally closed at a point $\bar{x}$ (not necessarily in A) if $C \cap V$ is closed for some closed neighborhood V

of $\bar{x}$. The polar cone of A is defined by

$$A^* = \{v \in R^n \mid \langle v, x\rangle \leq 0 \quad \forall x \in A\}.$$

The positive hull of A is defined by

$$\text{pos}A = \{\lambda x \mid x \in A, \lambda \geq 0\}.$$

The horizon cone of A, representing the direction set of A, is defined by

$$A^\infty = \{x \in R^n \mid \exists x_k \in A,\ \exists \lambda_k \to 0+ \text{ with } \lambda_k x_k \to x\}.$$

The distance function to A, written as $d_A(\cdot)$ or $d(\cdot, A)$, is defined by

$$d_A(x) := \inf_{y\in A} \|x - y\|.$$

The projection mapping P_A that assigns to each $x \in R^n$ the point, or points, of A nearest to x, is defined by

$$P_A(x) := \{y \in A \mid \|y - x\| = d_A(x)\}.$$

The indicator function of A is defined by

$$\delta_A(x) := \begin{cases} 0 & \text{if } x \in A, \\ +\infty & \text{otherwise.} \end{cases}$$

If A is empty, we set by convention

$$A^* = R^n, \quad \text{pos}A = \{0\}, \quad A^\infty = \{0\}, \quad d_A(\cdot) = +\infty, \quad P_A(\cdot) = \emptyset, \quad \text{and} \quad \delta_A(\cdot) = \infty.$$

Let $\bar{x} \in A$. Variational geometry of A at $\bar{x}$ can be captured by a number of notions that have been investigated in great details in Chapters 6 and 13 of Rockafellar and Wets [141].

(i) A vector $w \in R^n$ belongs to the tangent cone $T_A(\bar{x})$ to A at $\bar{x}$, if there are sequences $t_k \to 0+$ and $w_k \to w$ such that $\bar{x} + t_k w_k \in A$ for all k.

(ii) A vector $v \in R^n$ belongs to the proximal normal cone $N_A^P(\bar{x})$ to A at $\bar{x}$, if there exists some $\tau > 0$ such that $\bar{x} \in P_A(\bar{x} + \tau v)$.

(iii) The regular normal cone $\widehat{N}_A(\bar{x})$ to A at $\bar{x}$ is the polar cone of $T_A(\bar{x})$.

(iv) A vector $v \in R^n$ belongs to the normal cone $N_A(\bar{x})$ to A at $\bar{x}$, if there are sequences $x_k \to \bar{x}$ and $v_k \to v$ with $x_k \in A$ and $v_k \in \widehat{N}_A(x_k)$ for all k.

(v) The set A is said to be regular at $\bar{x}$ in the sense of Clarke if it is locally closed at $\bar{x}$ and $\widehat{N}_A(\bar{x}) = N_A(\bar{x})$.

(vi) A vector $z \in R^n$ belongs to the second-order tangent set to A at $\bar{x}$ for a vector $w \in T_A(\bar{x})$, written as $z \in T_A^2(\bar{x} \mid w)$, if there are sequences $t_k \to 0+$ and $z_k \to z$ such that $\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k \in A$ for all k. When $w \notin T_A(\bar{x})$, we interpret $T_A^2(\bar{x} \mid w)$ as an empty set.

In harmony with the general theory of set-valued mappings, it is convenient to think of N_A^P, $\widehat{N}_A$ and N_A not just as mappings on C but of type $R^n \rightrightarrows R^n$ with

$$N_A^P(\bar{x}) = \widehat{N}_A(\bar{x}) = N_A(\bar{x}) := \emptyset \quad \text{when} \quad \bar{x} \notin A.$$

In general, we have $N_A^P(x) \subset \widehat{N}_A(x) \subset N_A(x)$ for every $x \in A$ and hence

$$\mathrm{gph} N_A^P \subset \mathrm{gph} \widehat{N}_A \subset \mathrm{gph} N_A. \tag{1.3.20}$$

In what follows, let $f : R^n \to \overline{R}$ be an extended real-valued function and let $\bar{x}$ be a point with $f(\bar{x})$ finite. The effective domain of f is the set

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

and the epigraph of f is the set

$$\mathrm{epi} f := \{(x, \alpha) \in R^n \times R \mid \alpha \geq f(x)\}.$$

For each $\alpha \in \overline{R}$, we will find it useful to have the notation for the lower level set

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

The function f is said to be lower semicontinuous if, $\mathrm{epi} f$ is closed in $R^n \times R$ or equivalently the level sets of type $\mathrm{lev}_{\leq \alpha} f$ are all closed in R^n; it is said to be upper semicontinuous if the function $-f$ is lower semicontinuous; and it is said to be locally lower semicontinuous at $\bar{x}$ if, $\mathrm{epi} f$ is locally closed at $(\bar{x}, f(\bar{x}))$.

The notions related with generalized differential and subdifferential that we need throughout the thesis are summarized as follows. They have been extensively studied in Chapters 8 and 13 of Rockafellar and Wets [141].

(i) The vector $v \in R^n$ is a regular subgradient of f at $\bar{x}$, written $v \in \widehat{\partial} f(\bar{x})$, if

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(\|x - \bar{x}\|).$$

(ii) The vector $v \in R^n$ is a (general) subgradient of f at $\bar{x}$, written $v \in \partial f(\bar{x})$, if there are sequences $x^k \to \bar{x}$ and $v^k \to v$ with $f(x^k) \to f(\bar{x})$ and $v^k \in \widehat{\partial} f(x^k)$.

(iii) The function f is said to be regular at $\bar{x}$ if epi$f$ is regular in the sense of Clarke at $(\bar{x}, f(\bar{x}))$ as a subset of $R^n \times R$.

(iv) The subderivative function $df(\bar{x}) : R^n \to \overline{R}$ is defined by

$$df(\bar{x})(w) := \liminf_{\tau \to 0+,\, w' \to w} \frac{f(\bar{x} + \tau w') - f(\bar{x})}{\tau}.$$

(v) For any $v \in R^n$, the second subderivative at $\bar{x}$ for v and w is

$$d^2 f(\bar{x} \mid v)(w) := \liminf_{\tau \to 0+,\, w' \to w} \frac{f(\bar{x} + \tau w') - f(\bar{x}) - \tau \langle v, w' \rangle}{\frac{1}{2}\tau^2}.$$

(vi) For any vector w with $df(\bar{x})(w)$ finite, the parabolic subderivative at $\bar{x}$ for w with respect to z is

$$d^2 f(\bar{x})(w \mid z) := \liminf_{\tau \to 0+,\, z' \to z} \frac{f(\bar{x} + \tau w + \frac{1}{2}\tau^2 z') - f(\bar{x}) - \tau df(\bar{x})(w)}{\frac{1}{2}\tau^2}.$$

(vii) The function f is said to be parabolically regular at $\bar{x}$ for a vector $v \in R^n$ (Definition 13.65 of [141]) if the equality

$$d^2 f(\bar{x} \mid v)(w) = \inf_{z \in R^n} \left\{ d^2 f(\bar{x})(w \mid z) - \langle v, z \rangle \right\}$$

holds for every w having $df(\bar{x})(w) = \langle v, w \rangle$, or in other words if for such w with $d^2 f(\bar{x} \mid v)(w) < \infty$ there exist, among the sequences $\tau^k \to 0+$ and $w^k \to w$ with

$$\frac{f(\bar{x} + \tau^k w^k) - f(\bar{x}) - \tau^k \langle v, w^k \rangle}{\frac{1}{2}(\tau^k)^2} \to d^2 f(\bar{x} \mid v)(w),$$

ones with the additional property that $\limsup_v \|w^k - w\|/\tau^k < \infty$.

$(viii)$ The function f is said to be calm at $\bar{x}$ from below with modulus $\tau \in R_+$ if there exists a neighborhood V of $\bar{x}$ such that

$$f(x) \geq f(\bar{x}) - \tau \|x - \bar{x}\| \qquad \forall x \in V.$$

It should be noted that the notion of calmness from below is closely related with the notion of strong slope introduced by De Giorgi et al. in [63], where the strong slope of f at $\bar{x}$ is defined by

$$|\nabla f|(\bar{x}) := \limsup_{x \to \bar{x},\, x \neq \bar{x}} \frac{(f(\bar{x}) - f(x))_+}{\|x - \bar{x}\|}.$$

1.4 Motivation and Outline of the Thesis

The study of exact penalty functions has grown and proliferated in many interesting areas within mathematical optimization society. In the literature, exact penalty functions have been employed to derive optimality conditions for constrained optimization problems, see [35, 36, 66, 105] and the milestone paper by Burke [26]. The technique used in these work is by first transforming constrained optimization problems into unconstrained ones via exact penalty functions, and then establishing the equivalence of optimality conditions for the constrained and unconstrained optimization problems. It should be noticed that not all exact penalty functions are qualified for deriving optimality conditions in this direct way by establishing such equivalences, as can be seen from Yang and Meng [161]. In this connection, a natural question arises as to whether and how general exact penalty functions can be employed to derive optimality conditions for constrained optimization problems. To a great extent, this thesis is motivated by Yang and Meng's work [161], and can be regarded as a further development of the idea hidden in [161]. In this thesis we aim to study the theory of deriving optimality conditions for constrained optimization problems from very general exact penalty functions, and intend to develop a unified theory from a modern perspective of variational analysis popularized by Rockafellar and Wets' book [141].

For simplicity in the present discussion, we proceed formally to sketch the main work of this thesis. Let C be a subset of R^n and let $\phi : R^n \to R_+ \cup \{+\infty\}$ be a lower semicontinuous function with the property that $x \in C$ if and only if $\phi(x) = 0$. Moreover, let $\bar{x} \in C$ be a fixed point and let the function $f_0 : R^n \to R$ be at least continuously differentiable.

Consider the constrained optimization problem

$$\text{(P)} \quad \min \quad f_0(x) \quad \text{s.t.} \quad x \in C,$$

and the unconstrained optimization problem

$$\min \quad f_0(x) + \mu\phi(x) \quad \text{s.t.} \quad x \in R^n,$$

where μ is a nonnegative real number. In harmony with the general theory of penalty functions, it is convenient to think of μ as a penalty parameter, ϕ as a penalty term, and $f_0 + \mu\phi$ as a penalty function. Note that when C is assumed to be the feasible set of (NLP) and f_0 is assumed to be the objective function of (NLP), the penalty function $f_0 + \mu\phi$ includes penalty functions of the form (1.1.7) and correspondingly the l_p $(0 \leq p \leq 1)$ penalty function as special cases.

We say that the penalty function $f_0+\mu\phi$ is *exact* at $\bar{x}$ if it has an unconstrained local minimum at $\bar{x}$ for a finite penalty parameter (and hence for all finite and larger values of the penalty parameter). It is to be note that if $f_0+\mu\phi$ is exact at $\bar{x}$, then by definition, (P) has a local minimum at $\bar{x}$. When ϕ happens to be the indicator function δ_C of C, it is clear to see that $f_0+\mu\phi$ is exact at $\bar{x}$ if and only if (P) has a local minimum at $\bar{x}$. In general, however, $f_0+\mu\phi$ may not be exact at $\bar{x}$ even if (P) has a local minimum at $\bar{x}$. In this thesis, we intend to address two basic questions concerning optimality conditions and exact penalty functions as to when penalty functions are exact at local minima of constrained optimization problems, and how optimality conditions of constrained optimization problems can be derived from exactness of penalty functions. Chapter 2 and part of Chapter 3 are devoted to the first question, while Chapters 4, 5 and part of Chapter 3 are devoted to the second one.

The outline of the thesis is as follow.

In Chapter 2, we study sufficient conditions for penalty terms to possess local error bounds. To be precise, we say that ϕ is a local error bound at $\bar{x}$ for C, if there exist some $\tau > 0$ and $\varepsilon > 0$ such that for all $x \in R^n$ with $\|x - \bar{x}\| < \varepsilon$,

$$\tau d_C(x) \leq \phi(x). \tag{1.4.21}$$

A stronger version of (1.4.21) asserts that there exist some $\tau > 0$ and $\varepsilon > 0$ such that for every $\alpha \geq 0$ and every $x \in R^n$ with $\|x - \bar{x}\| \leq \varepsilon$, it follows that

$$\tau d\left(x, \mathrm{lev}_{\leq\alpha}\phi\right) \leq (\phi(x) - \alpha)_+. \tag{1.4.22}$$

Note that $\mathrm{lev}_{\leq\alpha}\phi = C$ when $\alpha = 0$. Conditions of the type (1.4.22) were first studied by Azé and Corvellec [11]. A similar result can be found in ([37], Theorem 3.1).

In the first part of Chapter 2, we establish equivalent conditions for the stronger version (1.4.22) of the local error bound (1.4.21). These conditions are expressed in terms of the strong slopes, the subderivatives and regular subgradients of ϕ at points outside C. We observe that (1.4.21) not necessarily implies (1.4.22) even if ϕ is locally Lipschitz continuous. However, if ϕ is the max function for a finite collection of continuously differentiable functions, we show that (1.4.21) and (1.4.22) are equivalent. As a consequence of this equivalence, we show that the quasi-normality constraint qualification introduced by Hestenes [75] is sufficient for the existence of local error bounds. Moreover, we use an example to illustrate how to apply our result to identify when exactly local error bounds occur for a parameterized system.

In the second part of Chapter 2, we study sufficient conditions for (1.4.21) in a systematic way by checking limits defined on the boundary of C in the following way:

$$\lim_{k\to+\infty}\frac{\phi(x_k+t_kv_k)}{t_k}, \tag{1.4.23}$$

where $\{(x_k,v_k,t_k)\}\subset \mathrm{bd}C\times R^n\times R_{++}$ and $(x_k,v_k,t_k)\to(\bar{x},v,0)$ with $v\in N_C(\bar{x})\backslash\{0\}$. Studniarski and Ward ([151], Theorem 2.5) showed that (1.4.21) holds if all limits of the form (1.4.23) are positive. Ioffe and Outrata ([86], Theorem 2.1 (b)) obtained the same result under a weaker condition which requires positiveness of all limits of the form (1.4.23) but with an additional requirement that $v_k\in\widehat{N}_C(x_k)$ for all k. A similar result to that of Ioffe and Outrata has been given by Henrion and Outrata ([73], Theorem 1). In addition, we show under what circumstances on C these sufficient conditions are also necessary by considering limits of the kind (1.4.23) for the distance function d_C. We end this chapter by using an example to illustrate that the second class of sufficient conditions is applied when criteria studied in the first part fail.

In Chapter 3, we study first- and second-order necessary and sufficient conditions for $f_0+\mu\phi$ to be exact at $\bar{x}$. We present our conditions in terms of three epi-derivatives: subderivatives, second-order subderivatives, and parabolic second-order subderivatives. These notions have been successfully utilized to formulate very tight first- and second-order optimality conditions for an extended real-valued function to attain a local minimum, see in particular Theorems 10.1, 13.24 and 13.66 of Rockafellar and Wets [141]. We study some basic properties of the subderivative, the second subderivative and the parabolic subderivative of ϕ, and pay our attention to derivative kernels consisting of directions at which these subderivatives vanish. Because of the differentiability assumption on f_0, the subderivative, the second-order subderivative, and the parabolic second-order subderivatives of the function $f_0 + \mu\phi$ can be expressed respectively by that of f_0 and ϕ, which makes it possible for us to apply Theorems 10.1, 13.24 and 13.66 of [141] in a straightforward way.

In Chapter 4, we study KKT conditions and second-order necessary conditions of the form (1.1.6) for the nonlinear programming problem (NLP) via exactness of penalty functions associated with (NLP). By setting C to be the feasible set of (NLP), we have a general penalty function $f + \mu\phi$ for (NLP), which includes the l_p $(0 \leq p \leq 1)$ penalty functions as special cases. Beside the penalty term ϕ, we mainly focus on the penalty term S^p of the l_p penalty function. Explicitly, we have

$$S^p(x) = \left(\sum_{i \in I} g_i(x)_+ + \sum_{j \in J} |h_j(x)| \right)^p \quad \forall x \in R^n.$$

It is well known that both the KKT condition and the second-order necessary condition (1.1.6) holds at $\bar{x}$ if the l_1 penalty function is exact at $\bar{x}$, see in particular Han and Mangasarian ([66], Theorem 4.8) and Rockafellar ([139], Corollary 4.5). But for $0 < p < 1$, the KKT condition may not hold at $\bar{x}$ even if the l_p penalty function is exact at $\bar{x}$. This can be seen from the simple example: $\min -x$ s.t. $x^2 \leq 0$. However, Yang and Meng [161] showed that it is still possible to derive KKT conditions from lower order exact penalty functions, by requiring that the constraint functions of (NLP) satisfy some additional conditions in terms of (generalized) second-order derivatives. Yang and Meng formulated these conditions by applying Farkas' Lemma and by estimating Dini upper-directional derivatives of the l_p penalty function using the tools of (generalized) Taylor expansions.

We say that the penalty term ϕ is of KKT-type at $\bar{x}$ if the KKT condition holds at $\bar{x}$ whenever there is a continuously differentiable function f such that $f + \mu\phi$ is exact at $\bar{x}$. In Section 4.2, we study conditions under which penalty terms are of KKT-type. These conditions allow us to derive KKT conditions from exactness of penalty functions. The main results that we rely are Theorems 3.2.1 and 3.3.1, and the variational description of regular subgradients (Rockafellar and Wets [141], Proposition 8.5). In subsection 4.2.1, we give equivalent conditions for penalty terms ϕ and S^p to be of KKT-type. These equivalent conditions are expressed by either subderivative kernels or regular subgradients of ϕ and S^p. In subsection 4.2.2, we present several conditions in terms of the original data of (NLP), which are sufficient for S^p to be of KKT-type. In particular when (NLP) has one inequality only, we give full characterizations in terms of the original data for S^p to be of KKT-type. We end Section 4.2 by giving a class of parameterized problems to illustrate that our result can be applied to derive KKT conditions when all existing methods fail.

In Section 4.3, by applying the second-order necessary conditions presented in Theorem 3.3.3 and the duality theorem of linear programming, we derive second-order necessary conditions of the type (1.1.6) for (NLP) from exactness of $f + \mu\phi$ under some additional conditions in terms of the kernel of the parabolic subderivative of ϕ. When $\phi = S^p$, we give sufficient conditions for these conditions in terms of the original data of (NLP). We end this chapter by using an example to illustrate that even if neither the GCQ nor the SGCQ holds, our result obtained in this section can be applied to derive the second-order necessary condition (1.1.6).

In Chapter 5, we study strong stationarity and Mordukhovich stationarity for the mathematical program with complementarity constraints (MPCC) via exactness of penalty functions associated with (MPCC). Let $0 \leq p \leq 1$. We consider two l_p penalty functions for (MPCC) as follows:

$$\mathcal{H}_p(x) = f(x) + \mu \left(S(x) + \sum_{k \in K} |\min\{G_k(x), H_k(x)\}| \right)^p$$

and

$$\mathcal{G}_p(x) = f(x) + \mu \left(S(x) + \sum_{k \in K} \{(-G_k(x))_+ + (-H_k(x))_+ + |G_k(x)H_k(x)|\} \right)^p,$$

where S is given by (1.1.3). By setting C to be the feasible set of (MPCC), we have a general penalty function $f + \mu\phi$ for (MPCC), which includes the l_p penalty functions $\mathcal{G}_p$ and $\mathcal{H}_p$ as special cases. We say that the penalty term ϕ is of S-type (resp., M-type) at $\bar{x}$ if strong stationarity (resp., Mordukhovich stationarity) holds at $\bar{x}$ whenever there is a continuously differentiable function f such that $f + \mu\phi$ is exact at $\bar{x}$. In Section 5.2, we establish equivalent conditions for ϕ to be of S-type or M-type. In Section 5.3, we consider the l_p penalty functions $\mathcal{G}_p$ and $\mathcal{H}_p$. We give sufficient conditions in terms of the original data of (MPCC) for these two penalty functions to be of S-type and M-type. We also establish some relationships between these two penalty functions.

In Chapter 6, we conclude the thesis and give directions for future work.

Chapter 2

Characterizations of Local Error Bounds

2.1 Introduction

Throughout this chapter, let C be a subset of R^n, let $\bar{x} \in C$ be a fixed point, and let $\phi : R^n \to R_+ \cup \{+\infty\}$ be a lower semicontinuous function with the property that $x \in C$ if and only if $\phi(x) = 0$.

We say that ϕ is a local error bound at $\bar{x}$ for C, if there exist some $\tau > 0$ and $\varepsilon > 0$ such that for all $x \in R^n$ with $\|x - \bar{x}\| < \varepsilon$,

$$\tau d_C(x) \leq \phi(x). \tag{2.1.1}$$

Suppose that there is a function $f_0 : R^n \to R$ which is locally Lipschitz at $\bar{x}$ and has a local minimum at $\bar{x}$ relative to C. It follows from Clarke's elementary exact penalization principle ([36], Proposition 2.4.3) that if ϕ is a local error bound at $\bar{x}$ for C, then there exists some $\mu \geq 0$ such that $f_0 + \mu\phi$ has a local minimum at $\bar{x}$, see also Proposition 3.111 of [23] or Corollary 2.6 of [147]. That is, the local error bound (2.1.1) allows us to reduce the constrained problem of minimizing f_0 over C, in a local sense to that of minimizing $f_0 + \mu\phi$ over R^n. This fundamental result motivates us to study local error bounds and their characterizations in this chapter.

In general, there can be found in the literature two classes of sufficient conditions for (2.1.1). The first one is expressed in terms of various derivative-like objects defined for points outside C. Such a condition in terms of the Clarke subdifferential was first investigated by Ioffe [84], who used Ekeland's variational principle and the sum rule to provide a global error bound for a Lipschitz continuous equality system, under the assumption that the norm of any element in the Clarke subdifferential of the constraint function at each point outside the solution set is bounded away from zero. Ioffe's main idea has been developed in many aspects. One aspect of development is made by using other derivative-like objects, such as the limiting subdifferential (also known as the Mordukhovich subdifferential) [163, 151], a partial subdifferential in a general Banach space [91], the proximal subdifferential ([37], Theorem 3.1), the so-called abstract subdifferential [156], the lower Dini-directional derivatives [115, 116, 155], the subderivative (also known as the Hadamard directional derivative) [79, 157], and the strong slope [11, 10, 86, 119], a notion introduced by De Giorgi et al.[63]. Of particular note are conditions expressed by strong slopes, which are not only sufficient for (2.1.1), but also necessary if there exist some $\tau > 0$ and $\varepsilon > 0$ such that for every $\alpha \geq 0$ and every $x \in R^n$ with $\|x - \bar{x}\| \leq \varepsilon$, it follows that

$$\tau d\left(x, \mathrm{lev}_{\leq \alpha}\phi\right) \leq (\phi(x) - \alpha)_+. \tag{2.1.2}$$

Note that $\mathrm{lev}_{\leq \alpha}\phi = C$ when $\alpha = 0$. We confirm that (2.1.2) is a stronger version of (2.1.1). Conditions of the type (2.1.2) were first studied by Azé and Corvellec [11]. A similar result can be found in ([37], Theorem 3.1).

The second class of sufficient conditions for (2.1.1), which seems to attract less attention than the first one, relies on limits defined on the boundary of C in the following way:

$$\lim_{k \to +\infty} \frac{\phi(x_k + t_k v_k)}{t_k}, \tag{2.1.3}$$

where $\{(x_k, v_k, t_k)\} \subset \mathrm{bd}C \times R^n \times R_{++}$ and $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ with $v \in N_C(\bar{x})\backslash\{0\}$. Studniarski and Ward ([151], Theorem 2.5) showed that (2.1.1) holds if all limits of the form (2.1.3) are positive. Ioffe and Outrata ([86], Theorem 2.1 (b)) obtained the same result under a weaker condition which requires positiveness of all limits of the form (2.1.3) but with an additional requirement that $v_k \in \widehat{N}_C(x_k)$ for all k. A similar result to that of Ioffe and Outrata has been given by Henrion and Outrata ([73], Theorem 1).

The outline of this chapter is as follows. In Section 2.2, we study the first class of sufficient conditions for local error bounds, and establish equivalent conditions for the stronger version (2.1.2) of the local error bound (2.1.1). These conditions are expressed in terms of the strong slopes, the subderivatives and regular subgradients of ϕ at points outside C. We observe that (2.1.1) not necessarily implies (2.1.2) even if ϕ is locally Lipschitz continuous. However, if ϕ is the max function for a finite collection of continuously differentiable functions, we show that (2.1.1) and (2.1.2) are equivalent. As a consequence of this equivalence, we show that the quasi-normality constraint qualification introduced by Hestenes [75] is sufficient for the existence of local error bounds. We end Section 2.2 by using an example to illustrate how to apply our result to identify when exactly local error bounds occur for a parameterized system.

In Section 2.3, we study the second class of sufficient conditions for (2.1.1) in a systematic way, and show under what circumstances on C these sufficient conditions are also necessary by considering limits of the kind (2.1.3) for the distance function d_C. We end Section 2.3 by using an example to illustrate that the second class of sufficient conditions is applied when criteria obtained in Section 2.2 fail.

2.2 On Subdifferential, Subderivative and Strong Slope for Outside Points

In this section, we establish equivalent conditions for the stronger version (2.1.2) of the local error bound (2.1.1). These conditions are expressed in terms of the strong slopes, the subderivatives and regular subgradients of ϕ at points outside C. We observe that (2.1.1) not necessarily implies (2.1.2) even if ϕ is locally Lipschitz continuous. However, if ϕ is the max function for a finite collection of continuously differentiable functions, we show that (2.1.1) and (2.1.2) are equivalent. As a consequence of this equivalence, we show that the quasi-normality constraint qualification introduced by Hestenes [75] is sufficient for the existence of local error bounds. Finally, by an example, we illustrate how to apply our result to identify when exactly local error bounds occur for a parameterized system.

To begin with, we review a few results that are needed in the sequel.

Lemma 2.2.1 *(Ekeland variational principle). Let $f : R^n \to \overline{R}$ be lower semicontinu-*

ous with $\inf_{x\in R^n} f(x)$ *finite, and let* $\bar{x} \in R^n$ *be such that*

$$f(\bar{x}) \leq \inf_{x\in R^n} f(x) + \varepsilon,$$

where $\varepsilon > 0$*. Then, for any* $\delta > 0$*, there exists a point* $\tilde{x} \in R^n$ *such that* $\|\tilde{x} - \bar{x}\| \leq \dfrac{\varepsilon}{\delta}$*,* $f(\tilde{x}) \leq f(\bar{x})$*, and* $\tilde{x}$ *is the unique minimum of the function* $f(x) + \delta\|x - \tilde{x}\|$ *over* R^n*.*

Below are some basic results concerning the relations among strong slopes, subderivatives, and (regular) subgradients.

Lemma 2.2.2 *Consider a function* $f : R^n \to \overline{R}$ *and a point* $\bar{x}$ *with* $f(\bar{x})$ *finite. The following properties are equivalent:*

(i) *f is calm at $\bar{x}$ from below.*

(ii) $df(\bar{x})(0) = 0$*, or equivalently,* $df(\bar{x})(w) > -\infty$ *for all* $w \neq 0$*.*

(iii) $|\nabla f|(\bar{x}) < +\infty$*.*

Moreover, the following statements are true:

(i') $|\nabla f|(\bar{x}) = -\min_{\|w\|=1} df(\bar{x})(w) = \inf\{\tau \in R_+ \mid f \text{ is calm at } \bar{x} \text{ from below with modulus } \tau\}$*.*

(ii') $|\nabla f|(\bar{x}) \leq d(0, \widehat{\partial} f(\bar{x}))$*, and if f is locally lower semicontinuous at $\bar{x}$, then*

$$d(0, \partial f(\bar{x})) \leq |\nabla f|(\bar{x}).$$

(iii') *If f is regular at $\bar{x}$, then* $d(0, \partial f(\bar{x})) = |\nabla f|(\bar{x})$*, and*

$$\partial f(\bar{x}) \neq \emptyset \Longleftrightarrow f \text{ is calm at } \bar{x} \text{ from below}.$$

Proof. In view of Proposition 8.32 of [141] and the definition of strong slope, all the results follow readily. □

The following two examples demonstrate that the inequalities in statement (ii') of Lemma 2.2.2 can be strict when f is not regular at $\bar{x}$.

Example 2.2.1 *Let* $f(x) = -|x|$ *and let* $\bar{x} = 0$. f *is clearly locally lower semicontinuous at* $\bar{x}$ *since* f *is lower semicontinuous. We have* $\widehat{\partial} f(\bar{x}) = \emptyset$, $\partial f(\bar{x}) = \{\pm 1\}$, *and* $|\nabla f|(\bar{x}) = 1$. *Therefore,* f *is not regular at* $\bar{x}$, *and*

$$1 = d(0, \partial f(\bar{x})) = |\nabla f|(\bar{x}) < d(0, \widehat{\partial} f(\bar{x})) = +\infty.$$

Example 2.2.2 *Let* $f(x) = -\sqrt{x_+}$ *and let* $\bar{x} = 0$. f *is clearly locally lower semicontinuous at* $\bar{x}$ *since* f *is lower semicontinuous. We have* $\widehat{\partial} f(\bar{x}) = \emptyset$, $\partial f(\bar{x}) = -1$, *and* $|\nabla f|(\bar{x}) = +\infty$. *Therefore,* f *is not regular at* $\bar{x}$, *and*

$$1 = d(0, \partial f(\bar{x})) < |\nabla f|(\bar{x}) = d(0, \widehat{\partial} f(\bar{x})) = +\infty.$$

Now, we give characterizations for the stronger version (2.1.2) of the local error bound (2.1.1).

Theorem 2.2.1 *Consider the following conditions:*

(*i*) *There exist* $\varepsilon > 0$ *and* $\tau > 0$ *such that for every* $x \notin C$ *with* $\|x - \bar{x}\| \leq \varepsilon$ *and* $\phi(x) < +\infty$, *it holds that*

$$d(0, \partial \phi(x)) \geq \tau.$$

(*ii*) *There exist* $\varepsilon > 0$ *and* $\tau > 0$ *such that for every* $x \notin C$ *with* $\|x - \bar{x}\| \leq \varepsilon$ *and* $\phi(x) < +\infty$, *it holds that*

$$|\nabla \phi|(x) \geq \tau.$$

(*iii*) *There exist* $\varepsilon > 0$ *and* $\tau > 0$ *such that for every* $x \notin C$ *with* $\|x - \bar{x}\| \leq \varepsilon$ *and* $\phi(x) < +\infty$, *it holds that*

$$d\phi(x)(w) \leq -\tau$$

for some $w \in R^n$ *with* $\|w\| = 1$.

(*iv*) *There exist* $\varepsilon > 0$ *and* $\tau > 0$ *such that for every* $\alpha \geq 0$ *and every* $x \in R^n$ *with* $\|x - \bar{x}\| \leq \varepsilon$, *it holds that*

$$\tau d\left(x, \mathrm{lev}_{\leq \alpha} \phi\right) \leq (\phi(x) - \alpha)_+.$$

We have (i) $\Longrightarrow$ (ii) $\Longleftrightarrow$ (iii) $\Longleftrightarrow$ (iv). If, in addition, ϕ is regular at every point $x \notin C$ with $\phi(x) < +\infty$ and $\|x - \bar{x}\| \leq \varepsilon$ for some $\varepsilon > 0$, we have (ii) $\Longrightarrow$ (i).

Proof. In view of Lemma 2.2.2, it remains to show $(ii) \Longleftrightarrow (iv)$. First, we show $(ii) \Longrightarrow (iv)$. Suppose that ε and τ are the constants satisfying condition (ii). Let $\alpha \geq 0$ and $\|x - \bar{x}\| \leq \varepsilon'$ with $0 < \varepsilon' \leq \dfrac{\varepsilon}{2}$. To show condition (iv), it suffices to show

$$\tau d(x, \mathrm{lev}_{\leq \alpha}\phi) \leq (\phi(x) - \alpha)_+. \tag{2.2.4}$$

If $x \in \mathrm{lev}_{\leq \alpha}\phi$ or $\phi(x) = +\infty$, (2.2.4) holds trivially. In what follows, we assume that $\alpha < \phi(x) < +\infty$. In view of the lower semicontinuity of ϕ, the lower level set $\mathrm{lev}_{\leq \alpha}\phi$ is closed and we have

$$0 < d(x, \mathrm{lev}_{\leq \alpha}\phi) \leq \|x - \bar{x}\|, \tag{2.2.5}$$

where the second inequality follows because $\bar{x} \in \mathrm{lev}_{\leq \alpha}\phi$. Suppose by contradiction that (2.2.4) does not hold, i.e.,

$$(\phi(x) - \alpha)_+ < \tau d(x, \mathrm{lev}_{\leq \alpha}\phi).$$

Due to (2.2.5), we can find two constants $\bar{\tau}, \bar{\rho}$ with $0 < \bar{\tau} < \tau$ and $0 < \bar{\rho} < d(x, \mathrm{lev}_{\leq \alpha}\phi)$ satisfying $(\phi(x) - \alpha)_+ \leq \bar{\tau}\bar{\rho}$. Let $f(y) = (\phi(y) - \alpha)_+$ for all $y \in R^n$. It is easy to check that f is lower semicontinuous with $\inf_{y \in R^n} f(y) = f(\bar{x}) = 0$. Thus, we have

$$f(x) \leq \inf_{y \in R^n} f(y) + \bar{\tau}\bar{\rho}.$$

It follows from Lemma 2.2.1 that, there exists a point $\tilde{x} \in R^n$ such that $\|\tilde{x} - x\| \leq \bar{\rho}$, $f(\tilde{x}) \leq f(x)$, and, for all $y \in R^n$ with $y \neq \tilde{x}$,

$$f(\tilde{x}) < f(y) + \bar{\tau}\|y - \tilde{x}\|.$$

This implies, by the definition of the strong slope, that

$$|\nabla f|(\tilde{x}) \leq \bar{\tau}.$$

We claim that $f(\tilde{x}) > 0$ or equivalently $\tilde{x} \notin C$, because otherwise we have $\tilde{x} \in \mathrm{lev}_{\leq \alpha}\phi$, implying that $d(x, \mathrm{lev}_{\leq \alpha}\phi) \leq \|\tilde{x} - x\| \leq \bar{\rho} < d(x, \mathrm{lev}_{\leq \alpha}\phi)$, which is impossible. By the lower semicontinuity of f, we have for all y close to $\tilde{x}$,

$$f(y) = \phi(y) - \alpha > 0,$$

which implies that $|\nabla\phi|(\tilde{x}) = |\nabla f|(\tilde{x}) \leq \bar{\tau}$. In view of (2.2.5), we have

$$\|\tilde{x}-\bar{x}\| \leq \|\tilde{x}-x\| + \|x-\bar{x}\| \leq \bar{\rho} + \|x-\bar{x}\| \leq d(x, \mathrm{lev}_{\leq\alpha}\phi) + \|x-\bar{x}\| \leq 2\|x-\bar{x}\| \leq \varepsilon.$$

Therefore, we have found a point $\tilde{x} \notin C$ satisfying $\|\tilde{x}-\bar{x}\| \leq \varepsilon$, $\phi(\tilde{x}) = f(\tilde{x}) + \alpha < +\infty$, and $|\nabla\phi|(\tilde{x}) \leq \bar{\tau} < \tau$. This contradicts to condition (ii). Thus, (2.2.4) holds and the proof for $(ii) \Longrightarrow (iv)$ is completed.

Now, we show $(iv) \Longrightarrow (ii)$. Suppose that ε and τ are the constants satisfying condition (iv). Let $0 < \varepsilon' \leq \varepsilon$ and $x \notin C$ with $\|x-\bar{x}\| \leq \varepsilon'$ and $\phi(x) < +\infty$. To show condition (ii), it suffices to show

$$|\nabla\phi|(x) \geq \tau. \tag{2.2.6}$$

Let $\alpha_k := \phi(x) - \frac{1}{k}$ for all positive integers k. Due to $x \notin C$, we have $\phi(x) > 0$ and hence there exists some $\bar{k}$ such that $\alpha_k > 0$ for all $k \geq \bar{k}$. Since ϕ is lower semicontinuous, $\mathrm{lev}_{\leq\alpha_k}\phi$ is closed for all k. By condition (ii), we can find some $x_k \in \mathrm{lev}_{\leq\alpha_k}\phi$ such that

$$\tau\|x-x_k\| = \tau d(x, \mathrm{lev}_{\leq\alpha_k}\phi) \leq \phi(x) - \alpha_k \ \ \forall k \geq \bar{k}.$$

We have $x_k \neq x$ for all $k \geq \bar{k}$, because otherwise $\phi(x) = \phi(x_k) \leq \alpha_k = \phi(x) - \frac{1}{k}$, which is impossible. Therefore, we have

$$0 < \|x-x_k\| \leq \frac{\phi(x)-\alpha_k}{\tau} = \frac{1}{k\tau} \qquad \forall k \geq \bar{k}.$$

This implies that $x_k \to x$ as $k \to \infty$, and that x is not a local minimum of ϕ because

$$\phi(x_k) \leq \alpha_k = \phi(x) - \frac{1}{k} < \phi(x) \qquad \forall k \geq \bar{k}.$$

By the definition of the strong slope, we have

$$|\nabla\phi|(x) \geq \limsup_{k\to\infty} \frac{\phi(x)-\phi(x_k)}{\|x-x_k\|} \geq \limsup_{k\to\infty} \frac{\phi(x)-\alpha_k}{\|x-x_k\|} \geq \tau,$$

which shows (2.2.6). The proof is completed. □

Remark 2.2.1 *The equivalence of conditions* (ii) *and* (iv) *can be easily derived from Proposition 2.1 and Theorem 5.1 in [11]. Here, for completeness, we have given a direct proof in the context of finite dimensional spaces. Condition* (iv) *implies that* ϕ *is a local error bound for* C *at* $\bar{x}$, *but not vice versa even when* ϕ *is Lipschitzian continuous as shown by Example 2.2.3.*

Example 2.2.3 *(Example 5.2 of [151]) Let $C = -R_+$ and let $\phi : R \to R_+$ be defined by*

$$\phi(x) = \begin{cases} 0 & \text{if } x \le 0, \\ 2^{-n} & \text{if } 2^{-n-1} \le x \le 2^{-n} \text{ with } n \text{ being an odd integer}, \\ 3x - 2^{-n} & \text{if } 2^{-n-1} \le x \le 2^{-n} \text{ with } n \text{ being an even integer}, \\ x & \text{otherwise}. \end{cases}$$

It is clear to see that ϕ is Lipschitzian continuous and that $\phi(x) = 0$ if and only if $x \in C$. Consider a point $\bar{x} = 0 \in C$. It is easy to check that ϕ is a local error bound for C at $\bar{x}$, since $\phi(x) \ge x_+$ for all $x \in R$ and the function x_+ is a local error bound for C at $\bar{x}$. However, for $2^{-n-1} < x < 2^{-n}$ with n being any odd integer, we have $|\nabla\phi|(x) = 0$. Therefore, by Theorem 2.2.1, there exist no $\varepsilon > 0$ and $\tau > 0$ such that, for every $\alpha \ge 0$ and every $x \in R$ with $|x - \bar{x}| \le \varepsilon$, it holds that

$$\tau d\left(x, \text{lev}_{\le\alpha}\phi\right) \le (\phi(x) - \alpha)_+. \tag{2.2.7}$$

To see that in a direct way, let $x_k = 2^{-k}$ and $\alpha_k = 2^{-k} - 2^{-2k}$ for each odd integer k. Then, we have

$$\frac{(\phi(x_k) - \alpha_k)_+}{d\left(x, \text{lev}_{\le\alpha_k}\phi\right)} \le \frac{2^{-2k}}{2^{-k} - 2^{-k-1}},$$

where the right-hand side term tends to 0 as $k \to +\infty$. Therefore, we cannot find $\varepsilon > 0$ and $\tau > 0$ such that (2.2.7) holds for every $\alpha \ge 0$ and every $x \in R$ with $|x - \bar{x}| \le \varepsilon$.

Let $I = \{1, \cdots, m\}$ and let $f_i : R^n \to R$ be continuously differentiable for all $i \in I$. Consider the function defined by

$$\phi(x) = (\max_{i \in I} f_i(x))_+ \qquad \forall x \in R^n, \tag{2.2.8}$$

and the set defined by

$$C = \{x \in R^n \mid f_i(x) \le 0 \quad \forall i \in I\}. \tag{2.2.9}$$

The following proposition identifies exactly when ϕ is a local error bound for C at some $\bar{x} \in C$. In particular, we show that ϕ is a local error bound for C at $\bar{x}$ if and only if condition (iv) of Theorem 2.2.1 holds.

Proposition 2.2.1 *Let $\bar{x} \in C$ and let*

$$I(x) := \{i \in I \mid f_i(x) = \phi(x)\} \quad \forall x \in R^n, \tag{2.2.10}$$

where C is given by (2.2.9) and ϕ is given by (2.2.8). Then, the following conditions are equivalent:

(i) *ϕ is a local error bound for C at $\bar{x}$.*

(ii) *There exists $\tau > 0$ such that, for any sequences $\{x_k\} \subset R^n \backslash C$ and $\{p_k\} \subset C$ with $I(p_k) \equiv I'$, $p_k \in P_C(x_k)$, $x_k \to \bar{x}$, and $\frac{x_k - p_k}{\|x_k - p_k\|} \to \bar{v}$, it holds for some $i \in I'$ that*

$$\langle \nabla f_i(\bar{x}), \bar{v} \rangle \geq \tau.$$

(iii) *There exists $\tau > 0$ such that, for any $I' \in \mathcal{I}(\bar{x})$, there exists $\bar{v} \in R^n$ with $\|\bar{v}\| = 1$ such that*

$$\langle \nabla f_i(\bar{x}), \bar{v} \rangle \geq \tau \qquad \forall i \in I',$$

where

$$\mathcal{I}(\bar{x}) := \{ I' \subset I(\bar{x}) \mid \exists \{x_k\} \subset R^n \backslash C \text{ with } x_k \to \bar{x} \text{ and } I(x_k) \equiv I' \}. \tag{2.2.11}$$

(iv) *There exist $\varepsilon > 0$ and $\tau > 0$ such that, for every $x \notin C$ with $\|x - \bar{x}\| \leq \varepsilon$, one of the following equivalent conditions is satisfied: (a) $d(0, \partial\phi(x)) \geq \tau$; (b) $|\nabla \phi|(x) \geq \tau$; (c) $d\phi(x)(w) \leq -\tau$ for some $w \in R^n$ with $\|w\| = 1$.*

(v) *There exist $\varepsilon > 0$ and $\tau > 0$ such that, for every $\alpha \geq 0$ and every $x \in R^n$ with $\|x - \bar{x}\| \leq \varepsilon$, it holds that*

$$\tau d\left(x, \mathrm{lev}_{\leq \alpha} \phi\right) \leq (\phi(x) - \alpha)_+.$$

Proof. In what follows, we show step by step:

$$(i) \Longrightarrow (ii) \Longrightarrow (iii) \Longrightarrow (iv) \Longrightarrow (vi) \Longrightarrow (v) \Longrightarrow (i).$$

$[(i) \Longrightarrow (ii)]$: Since ϕ is a local error bound for C at $\bar{x}$, there exist $\varepsilon > 0$ and $\tau > 0$ such that for every $x \in R^n$ with $\|x - \bar{x}\| \leq \varepsilon$,

$$\tau d(x, C) \leq \phi(x). \tag{2.2.12}$$

We will show that this constant τ satisfies condition (ii). Suppose by contradiction that, there exist sequences $\{x_k\} \subset R^n \backslash C$ and $\{p_k\} \subset C$ with $I(p_k) \equiv I'$, $p_k \in P_C(x_k)$, $x_k \to \bar{x}$, and $\frac{x_k - p_k}{\|x_k - p_k\|} \to \bar{v}$, but it follows that

$$\langle \nabla f_i(\bar{x}), \bar{v} \rangle < \tau \qquad \forall i \in I'.$$

Since all functions f_i are continuous, there exists $\rho_k \in (0,1]$ such that $\tilde{x}_k = p_k + \rho_k(x_k - p_k)$ and $I(\tilde{x}_k) \subset I(p_k)$. By taking a subsequence if necessary, we can assume that $I(\tilde{x}_k) \equiv I''$. It is easy to check that $p_k \in P_C(\tilde{x}_k)$, $\tilde{x}_k \to \bar{x}$ and $\frac{\tilde{x}_k - p_k}{\|\tilde{x}_k - p_k\|} \to \bar{v}$. Since $I'' \subset I'$, we have, for every $i \in I''$

$$\lim_{k\to+\infty} \frac{\phi(\tilde{x}_k)}{d(\tilde{x}_k, C)} = \lim_{k\to+\infty} \frac{\phi(\tilde{x}_k) - \phi(p_k)}{\|\tilde{x}_k - p_k\|} = \lim_{k\to+\infty} \frac{f_i(\tilde{x}_k) - f_i(p_k)}{\|\tilde{x}_k - p_k\|} = \langle \nabla f_i(\bar{x}), \bar{v} \rangle < \tau,$$

which contradicts to (2.2.12).

$[(ii) \Longrightarrow (iii)]$: Assume that condition (ii) holds with some constant $\tau > 0$. Let $\{x_k\} \subset R^n \backslash C$ be such that $x_k \to \bar{x}$ and $I(x_k) \equiv I'$. It is clear that $I' \in \mathcal{I}(\bar{x})$. Since C is closed, we have $P_C(x_k) \neq \emptyset$ for all k. Let $p_k \in P_C(x_k)$ for each k. By taking a subsequence if necessary, we can assume that $I(p_k) \equiv I''$ and that $\frac{x_k - p_k}{\|x_k - p_k\|} \to \bar{v}$ with $\|\bar{v}\| = 1$. It follows from condition (ii) that, for some $j \in I''$,

$$\langle \nabla f_j(\bar{x}),\, \bar{v} \rangle \geq \tau.$$

Thus, we have for every $i \in I'$,

$$\begin{aligned}
\langle \nabla f_i(\bar{x}), \bar{v} \rangle &= \lim_{k\to+\infty} \frac{f_i(x_k) - f_i(p_k)}{\|x_k - p_k\|} \\
&\geq \liminf_{k\to+\infty} \frac{\phi(x_k) - \phi(p_k)}{\|x_k - p_k\|} \\
&\geq \liminf_{k\to+\infty} \frac{f_j(x_k) - f_j(p_k)}{\|x_k - p_k\|} \\
&= \langle \nabla f_j(\bar{x}),\, \bar{v} \rangle \\
&\geq \tau.
\end{aligned}$$

Therefore, condition (iii) holds with the same constant τ as that in condition (ii).

$[(iii) \Longrightarrow (iv)]$: Assume that condition (iii) holds with some constant $\tau > 0$. By Example 7.28 and Exercise 8.31 of [141], ϕ is regular at every $x \in R^n \backslash C$ with

$$d\phi(x)(w) = \max_{i \in I(x)} \langle \nabla f_i(x), w \rangle \qquad \forall w \in R^n.$$

Let $0 < \tau' < \tau$. Since all functions f_i are smooth, there exists $\varepsilon_1 > 0$ such that, for all $i \in I$, $x \in R^n$ with $\|x - \bar{x}\| \leq \varepsilon_1$, and $v \in R^n$ with $\|v\| = 1$, it holds that

$$|\langle \nabla f_i(x) - \nabla f_i(\bar{x}), v \rangle| \leq \max_{i \in I} \|\nabla f_i(x) - \nabla f_i(\bar{x})\| \leq \tau - \tau'. \tag{2.2.13}$$

We claim that there exists $\varepsilon_2 > 0$ such that $I(x) \in \mathcal{I}(\bar{x})$ for all $x \notin C$ with $\|x - \bar{x}\| \leq \varepsilon_2$. Otherwise, there exists a sequence $\{x_k\} \subset R^n \backslash C$ such that $x_k \to \bar{x}$ and $I(x_k) \notin \mathcal{I}(\bar{x})$

for all k. By taking a subsequence if necessary, we can assume that $I(x_k) \equiv I'$. Since $I(x_k) \subset I(\bar{x})$ for all sufficiently large k, we have $I' \subset I(\bar{x})$ and hence $I' \in \mathcal{I}(\bar{x})$, contradicting to the assumption that $I(x_k) \notin \mathcal{I}(\bar{x})$ for all k. Let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ and let $x \notin C$ with $\|x - \bar{x}\| \leq \varepsilon$. We have $I(x) \in \mathcal{I}(\bar{x})$. Thus, by condition (iii), there exists $w \in R^n$ with $\|w\| = 1$ such that

$$\langle \nabla f_i(\bar{x}), w \rangle \geq \tau \qquad \forall i \in I(x). \tag{2.2.14}$$

By (2.2.13) and (2.2.14), we have, for every $i \in I(x)$,

$$\langle \nabla f_i(x), w \rangle \geq \tau' + \langle \nabla f_i(\bar{x}), w \rangle - \tau \geq \tau'.$$

Thus, we have

$$d\phi(x)(-w) \leq -\tau'.$$

In view of Theorem 2.2.1 and Lemma 2.2.2, condition (iv) follows readily.

$[(iv) \Longrightarrow (v)]$: This implication follows directly from Theorem 2.2.1.

$[(v) \Longrightarrow (i)]$: This implication is trivial. $\square$

Remark 2.2.2 *Condition (ii) was considered by Studniarski [150]. It follows from Lemma 2.2.2 and Rockafellar and Wets ([141], Example 7.28 and Exercise 8.31) that ϕ is regular at every $x \in R^n \backslash C$, at which the subdifferential, the subderivative, and the strong slope appeared in condition (iv) can be calculated as follows:*

$$\partial\phi(x) = \operatorname{conv}\{\nabla f_i(x) \mid i \in I(x)\},$$
$$d\phi(x)(w) = \max_{i \in I(x)} \langle \nabla f_i(x), w \rangle \qquad \forall w \in R^n,$$
$$|\nabla\phi|(x) = \min_{\|w\|=1} \max_{i \in I(x)} \langle \nabla f_i(x), w \rangle = \min\{\|y\| \mid y \in \operatorname{conv}\{\nabla f_i(x) \mid i \in I(x)\}\}.$$

Condition (iii) holds if and only if one of the following equivalent conditions is satisfied:

(a) $0 \notin \operatorname{conv}\{\nabla f_i(\bar{x}) \mid i \in I'\} \quad \forall I' \in \mathcal{I}(\bar{x})$.

(b) *For any $I' \in \mathcal{I}(\bar{x})$, the inequality system $f_i(x) \leq 0$ with $i \in I'$ satisfies the* MFCQ *at $\bar{x}$, i.e., there exists $v \in R^n$ such that*

$$\langle \nabla f_i(\bar{x}), v \rangle < 0 \quad \forall i \in I'.$$

(c) *For any $I' \in \mathcal{I}(\bar{x})$, it follows that*

$$\sum_{i \in I'} \lambda_i \nabla f_i(\bar{x}) = 0 \ \text{ and } \ \lambda_i \geq 0 \quad \forall i \in I' \Longrightarrow \lambda_i = 0 \quad \forall i \in I'. \tag{2.2.15}$$

In applying the criteria (a), (b), and (c), the main difficulty lies in the identification of the index collection $\mathcal{I}(\bar{x})$ defined by (2.2.11). Incidentally, we are not certain at this time how to identify $\mathcal{I}(\bar{x})$ in an easier way. It is worth mentioning that the index collection $\mathcal{I}(\bar{x})$ is in spirit of the following index collection:

$$\mathcal{J}(\bar{x}) := \{I' \subset I(\bar{x}) \mid \exists \{x_k\} \subset \mathrm{bd}C \backslash \{\bar{x}\} \text{ with } x_k \to \bar{x} \text{ and } I(x_k) \equiv I'\},$$

which was introduced by Henrion and Outrata, who essentially showed in Theorem 3 of [73] that if, the ACQ *holds at $\bar{x}$, i.e.,*

$$T_C(\bar{x}) = \{w \in R^n \mid \langle \nabla f_i(\bar{x}), w \rangle \leq 0 \ \forall i \in I(\bar{x})\}), \tag{2.2.16}$$

and

$$0 \notin \mathrm{conv}\{\nabla f_i(\bar{x}) \mid i \in I'\} \quad \forall I' \in \mathcal{J}(\bar{x}), \tag{2.2.17}$$

then ϕ is a local error bound for C at $\bar{x}$. Note that Henrion and Outrata's result provides in general only sufficient conditions for local error bounds, as can be demonstrated by a simple example where $f_1(x) := x_1 - x_2$, $f_2(x) := x_2 - x_1$ and $\bar{x} = (0,0)^T$. For this example, ϕ is clearly a local error bound for C at $\bar{x}$, and we have by definition $\mathcal{J}(\bar{x}) = \{\{1,2\}\}$ and $\mathcal{I}(\bar{x}) = \{\{1\},\{2\}\}$. Therefore, the criteria (a) is fulfilled but condition (2.2.17) is not satisfied. Another thing worth a mention is that if ϕ is a local error bound for C at $\bar{x}$, then necessarily the ACQ *holds at $\bar{x}$, i.e., the equality (2.2.16) holds.*

Let ϕ and C be given respectively by (2.2.8) and (2.2.9). Recall that the quasi-normality constraint qualification (QNCQ), introduced by Hestenes [75], is satisfied for C at some $\bar{x} \in C$, if there exist no nonzero vector $\lambda \in R^m_+$ and no sequence $x_k \to \bar{x}$ such that

$$\sum_{i \in I} \lambda_i \nabla f_i(\bar{x}) = 0,$$

and for all k, $\lambda_i f_i(x_k) > 0$ for all i with $\lambda_i > 0$. Very recently, Minchenko and Tarakanov ([112], Theorem 2.1) showed that if the QNCQ holds at $\bar{x}$ and the gradients of f_i are locally Lipschitz continuous, then ϕ is a local error bound for C at $\bar{x}$. By applying the criteria (2.2.15), the assumption on Lipschitz continuity of the gradients can be dropped as shown in the following corollary.

Corollary 2.2.1 *Let ϕ and C be given respectively by (2.2.8) and (2.2.9). Suppose that the* (QNCQ) *holds at some $\bar{x} \in C$. Then, ϕ is a local error bound for C at $\bar{x}$.*

Proof. Suppose by contradiction that ϕ is not a local error bound for C at $\bar{x}$. By applying the criteria (2.2.15), we can find some $I' \in \mathcal{I}(\bar{x})$ and a vector $\lambda \in R^m_+ \backslash \{0\}$ such that $\lambda_i = 0$ for all $i \notin I'$, and $\sum_{i \in I} \lambda_i \nabla f_i(\bar{x}) = 0$. By the definition of $\mathcal{I}(\bar{x})$ given by (2.2.11), we can find a sequence $x_k \to \bar{x}$ satisfying $I(x_k) = I'$ and $x_k \notin C$ for all k, where $I(x_k)$ is defined by (2.2.10). Thus, for all k and all $i \in I'$ with $\lambda_i > 0$, we have $\lambda_i f_i(x_k) > 0$. This indicates by definition that the (QNCQ) is not satisfied at $\bar{x}$. The proof is completed. □

We end this chapter by demonstrating in the following example how to apply the criteria (2.2.15) to identify when exactly local error bounds occur for a parameterized system.

Example 2.2.4 *Consider a subset of R^3 defined by*

$$C := \left\{ x \in R^3 \;\middle|\; \begin{array}{l} f_1(x) := a^T x + a_4 x_3^4 \leq 0 \\ f_2(x) := b^T x + b_4 x_3^4 \leq 0 \\ f_3(x) := c^T x + c_4 x_3^4 \leq 0 \end{array} \right\}, \tag{2.2.18}$$

and a function defined by

$$\phi(x) := (\max\{f_1(x), f_2(x), f_3(x)\})_+ \qquad \forall x \in R^3,$$

where $a = (a_1, a_2, a_3)^T$, $b = (b_1, b_2, b_3)^T$, $c = (c_1, c_2, c_3)^T \in R^3$ and a_4, b_4, $c_4 \in R$. It is clear that C is closed and nonempty (since $0 \in C$), and that $x \in C$ if and only if $\phi(x) = 0$.

In what follows, let $\bar{x} = (0, 0, 0)^T \in R^3$ and let $e_3 = (0, 0, 1)^T \in R^3$. To identify when ϕ is a local error bound for C at $\bar{x}$ by applying the criteria (2.2.15), we consider all possible data settings except for those that can be treated symmetrically as follows:

(A) *The vectors a, b and c are linearly independent.*

(B) *The vectors* a *and* b *are linearly independent, and*

- (B1) $c = -k_1 a - k_2 b$ *with either* $k_1 < 0$ *or* $k_2 < 0$.
- (B2) $c = -k_1 a - k_2 b$ *with* $k_1 \geq 0$, $k_2 \geq 0$, *and* $k_1 a_4 + k_2 b_4 + c_4 \leq 0$.
- (B3) $c = -k_1 a - k_2 b$ *with* $k_1 \geq 0$, $k_2 \geq 0$, $k_1 k_2 = 0$, *and* $k_1 a_4 + k_2 b_4 + c_4 > 0$.
- (B4) $c = -k_1 a - k_2 b$ *with* $k_1 > 0$, $k_2 > 0$, $k_1 a_4 + k_2 b_4 + c_4 > 0$, *and*
  - (B4-1) *the vectors* a, b, *and* e_3 *are linearly dependent.*
  - (B4-2) *the vectors* a, b, *and* e_3 *are linearly independent.*

(C) $a \neq 0$, *and*

- (C1) $b = \rho_1 a$ *and* $c = \rho_2 a$ *with* $\rho_1 > 0$ *and* $\rho_2 > 0$.
- (C2) $b = \rho_1 a$ *and* $c = \rho_2 a$ *with* $\rho_1 \leq 0$, $\rho_2 < 0$, $b_4 - \rho_1 a_4 \leq 0$, *and* $c_4 - \rho_2 a_4 \leq 0$.
- (C3) $b = \rho_1 a$ *and* $c = \rho_2 a$ *with* $\rho_1 \leq 0$, $\rho_2 < 0$, *either* $b_4 - \rho_1 a_4 > 0$ *or* $c_4 - \rho_2 a_4 > 0$, *and*
  - (C3-1) *the vectors* a *and* e_3 *are linearly dependent.*
  - (C3-2) *the vectors* a *and* e_3 *are linearly independent.*
- (C4) $b = \rho_1 a$ *and* $c = \rho_2 a$ *with* $\rho_1 > 0$, $\rho_2 \leq 0$, *and* $c_4 - \rho_2 a_4 \leq 0$.
- (C5) $b = \rho_1 a$ *and* $c = \rho_2 a$ *with* $\rho_1 > 0$, $\rho_2 \leq 0$, $c_4 - \rho_2 a_4 > 0$, *and*
  - (C5-1) *the vectors* a *and* e_3 *are linearly dependent.*
  - (C5-2) *the vectors* a *and* e_3 *are linearly independent.*
- (C6) $b = 0$, $c = 0$, $b_4 \leq 0$, *and* $c_4 \leq 0$.
- (C7) $b = 0$, $c = 0$, *either* $b_4 > 0$ *or* $c_4 > 0$.

(D) $a = b = c = 0$, *and*

- (D1) $a_4 \leq 0$, $b_4 \leq 0$, *and* $c_4 \leq 0$.
- (D2) *one of the numbers* a_4, b_4, *and* c_4 *is positive.*

We confirm that for cases (A), (B1), (B2), (B4-1), (C1), (C2), (C3-1), (C4), (C5-1), (C6), *and* (D1), ϕ *is a local error bound for* C *at* $\bar{x}$. *But for the rest cases,* ϕ *is not a local error bound for* C *at* $\bar{x}$ *because in each of these cases, we have*

$$T_C(\bar{x}) = \mathrm{clconv} T_C(\bar{x}) \subsetneq \{w \in R^3 \mid a^T w \leq 0,\ b^T w \leq 0,\ c^T w \leq 0\}, \tag{2.2.19}$$

i.e., neither the ACQ *nor the* GCQ *is satisfied at $\bar{x}$. Recall that the* ACQ *at $\bar{x}$ is defined by the equality (2.2.16). While the* GCQ *at $\bar{x}$ is by definition the equality (2.2.16) with the left-hand side being replaced by its closed convex hull.*

In what follows, we give details on how to apply the criteria (2.2.15) to cases (B2), (B3), *and* (B4-1), *which are relatively more complicated cases than others. To begin with, we recall that the index collection $\mathcal{I}(\bar{x})$ defined by (2.2.11) has the property that $I' \in \mathcal{I}(\bar{x})$ if and only if $I' \subset \{1,2,3\}$ and there exists a sequence $x_k \to \bar{x}$ such that for each k, $f_i(x_k) = f_j(x_k) > 0$ for all $i, j \in I'$, and $f_i(x_k) > f_l(x_k)$ for all $i \in I'$ and $l \notin I'$.*

[(B2)]: *We further assume that $k_1 > 0$ and $k_2 > 0$ as the other cases can be investigated in a similar way. For any $I' \subset \{1,2,3\}$, it is easy to check that (2.2.15) holds if and only if $I' \neq \{1,2,3\}$. Suppose by contradiction that $\{1,2,3\} \in \mathcal{I}(\bar{x})$. Then we can find a sequence $x_k \to \bar{x}$ such that for each k, $f_1(x_k) = f_2(x_k) = f_3(x_k) > 0$. Thus,*

$$k_1 f_1(x_k) + k_2 f_2(x_k) + f_3(x_k) = (k_1 a_4 + k_2 b_4 + c_4) x_{3k}^4 > 0,$$

which is impossible because $k_1 a_4 + k_2 b_4 + c_4 \leq 0$. This indicates that $\{1,2,3\} \notin \mathcal{I}(\bar{x})$. Therefore, ϕ is a local error bound for C at $\bar{x}$.

[(B3)]: *For any $x \in C$, we have*

$$k_1 f_1(x) + k_2 f_2(x) + f_3(x) = (k_1 a_4 + k_2 b_4 + c_4) x_3^4,$$

which implies that $x_3 = 0$ because $k_1 a_4 + k_2 b_4 + c_4 > 0$. Thus, we have

$$C = \left\{ x \in R^3 \;\middle|\; \begin{array}{r} a^T x \leq 0 \\ b^T x \leq 0 \\ c^T x \leq 0 \\ x_3 = 0 \end{array} \right\}.$$

That is, C is a polyhedral cone in R^3. We thus have $C = T_C(\bar{x}) = \mathrm{clconv} T_C(\bar{x})$. Moreover, it is not hard to check that

$$C \subsetneq \{w \in R^3 \mid a^T w \leq 0,\ b^T w \leq 0,\ c^T w \leq 0\},$$

when $k_1 \geq 0$, $k_2 \geq 0$, $k_1 k_2 = 0$, $c = -k_1 a - k_2 b$ and the vectors a and b are linearly independent.

[(B4-1)]: *Let $e_3 = \lambda_1 a + \lambda_2 b$ for some $\lambda_1, \lambda_2 \in R$. For any $I' \subset \{1,2,3\}$, it is easy to check that (2.2.15) holds if and only if $I' \neq \{1,2,3\}$. Suppose by contradiction* *that* $\{1,2,3\} \in \mathcal{I}(\bar{x})$. *Then we can find a sequence* $x_k \to \bar{x}$ *such that for each* k, $f_1(x_k) = f_2(x_k) = f_3(x_k) > 0$. *Set* $\beta_k = f_1(x_k)$ *for each* k. *Then we have*

$$k_1\beta_k + k_2\beta_k + \beta_k = (k_1a_4 + k_2b_4 + c_4)x_{3k}^4,$$

and hence

$$\beta_k = \frac{k_1a_4 + k_2b_4 + c_4}{k_1 + k_2 + 1}x_{3k}^4 > 0.$$

This shows that $x_{3k} \neq 0$ *for all* k. *However, we have*

$$x_{3k} = e_3^T x_k = (\lambda_1 a + \lambda_2 b)^T x_k = \lambda_1(\beta_k - a_4x_{3k}^4) + \lambda_2(\beta_k - b_4x_{3k}^4),$$

and hence

$$x_{3k} = \left[(\lambda_1 + \lambda_2)\frac{k_1a_4 + k_2b_4 + c_4}{k_1 + k_2 + 1} - (\lambda_1a_4 + \lambda_2b_4)\right]x_{3k}^4,$$

which is impossible because $x_{3k} \to 0$ *and* $x_{3k} \neq 0$ *for all* k. *This indicates that* $\{1,2,3\} \notin \mathcal{I}(\bar{x})$. *Therefore,* ϕ *is a local error bound for* C *at* $\bar{x}$.

2.3 On Some Limits Defined on Boundary

It has been shown by Studniarski and Ward ([151], Theorem 2.5) that ϕ is a local error bound at $\bar{x}$ for C if, for any sequence $\{(x_k, v_k, t_k)\} \subset \mathrm{bd}C \times R^n \times R_{++}$ such that $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ with $v \in N_C(\bar{x})\backslash\{0\}$, it follows that

$$\liminf_{k\to+\infty} \frac{\phi(x_k + t_kv_k)}{t_k} > 0. \tag{2.3.20}$$

This result motivates us to study similar sufficient conditions in this section. Moreover, we also study conditions under which these sufficient conditions become necessary.

Theorem 2.3.1 *Consider the following conditions:*

(i) *For any sequence* $\{(x_k, v_k, t_k)\} \subset (\mathrm{bd}C \times R^n \times R_{++} \cup \mathrm{int}C \times \{0\} \times R_{++})$ *such that* $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ *with* $v \in N_C(\bar{x})\backslash\{0\}$, *the inequality (2.3.20) holds.*

(ii) *For any sequence* $\{(x_k, v_k, t_k)\} \subset \mathrm{gph}N_C \times R_{++}$ *such that* $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ *with* $v \neq 0$, *the inequality (2.3.20) holds.*

(*iii*) *For any sequence* $\{(x_k, v_k, t_k)\} \subset \mathrm{gph}\widehat{N}_C \times R_{++}$ *such that* $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ *with* $v \neq 0$*, the inequality (2.3.20) holds.*

(*iv*) *For any sequence* $\{(x_k, v_k, t_k)\} \subset \mathrm{gph}N_C^P \times R_{++}$ *such that* $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ *with* $v \neq 0$*, the inequality (2.3.20) holds.*

(*v*) *For any sequence* $\{(x_k, v_k, t_k)\} \subset \Omega_0 := \{(x, v, t) \in R^n \times R^n \times R_{++} \mid x \in P_C(x+tv)\}$ *such that* $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ *with* $v \neq 0$*, the inequality (2.3.20) holds.*

(*vi*) ϕ *is a local error bound at* $\bar{x}$ *for* C*.*

Then, we have

$$(i) \Longrightarrow (ii) \Longrightarrow (iii) \Longrightarrow (iv) \Longrightarrow (v) \Longleftrightarrow (vi).$$

If, in addition, ϕ *is continuous, then*

$$(ii) \Longleftrightarrow (iii) \Longleftrightarrow (iv).$$

Proof. Note that any v appeared in conditions $(ii)-(v)$ is an element in the set $N_C(\bar{x})$. Since

$$\Omega_0 \subset \mathrm{gph}N_C^P \times R_{++} \subset \mathrm{gph}\widehat{N}_C \times R_{++} \subset \mathrm{gph}N_C \times R_{++}$$

and

$$\mathrm{gph}N_C \times R_{++} \subset (\mathrm{bd}C \times R^n \times R_{++} \cup \mathrm{int}C \times \{0\} \times R_{++}),$$

we have

$$(i) \Longrightarrow (ii) \Longrightarrow (iii) \Longrightarrow (iv) \Longrightarrow (v).$$

By definition, ϕ is a local error bound at $\bar{x}$ for C, if and only if,

$$\liminf_{x \to \bar{x}, x \notin C} \frac{\phi(x)}{d(x, C)} > 0. \tag{2.3.21}$$

Let $\{(x_k, v_k, t_k)\} \subset \Omega_0$ be a sequence such that $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ with $v \neq 0$. By the definition of Ω_0, we have $x_k \in P_C(x_k + t_k v_k)$, implying that $d(x_k + t_k v_k, C) = t_k \|v_k\|$. Since $v \neq 0$, we have $d(x_k + t_k v_k, C) > 0$ or equivalently $x_k + t_k v_k \notin C$ for all k sufficiently large. Thus, from the inequalities (2.3.20) and (2.3.21), we can easily show that $(v) \Longleftrightarrow (vi)$.

By assuming that ϕ is continuous, we now show $(iv) \Longrightarrow (ii)$. Let $\{(x_k, v_k, t_k)\} \subset \mathrm{gph}N_C \times R_{++}$ be a sequence such that $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ with $v \neq 0$. Since ϕ is continuous and

$$N_C(x) = \limsup_{y \to x,\, y \in C} N_C^P(y) \qquad \forall x \in C,$$

there exists a sequence $\{(\bar{x}_k, \bar{v}_k)\} \subset \mathrm{gph}N_C^P$ such that $(\bar{x}_k, \bar{v}_k) \to (\bar{x}, v)$ and

$$|\phi(x_k + t_k v_k) - \phi(\bar{x}_k + t_k \bar{v}_k)| \leq (t_k)^2. \tag{2.3.22}$$

By condition (iv), we have

$$\liminf_{k \to +\infty} \frac{\phi(\bar{x}_k + t_k \bar{v}_k)}{t_k} > 0. \tag{2.3.23}$$

From (2.3.22) and (2.3.23), we have

$$\liminf_{k \to +\infty} \frac{\phi(x_k + t_k v_k)}{t_k} \geq \liminf_{k \to +\infty} \left(\frac{\phi(\bar{x}_k + t_k \bar{v}_k)}{t_k} - t_k\right) > 0.$$

This shows that $(iv) \Longrightarrow (ii)$ and hence $(ii) \Longleftrightarrow (iii) \Longleftrightarrow (iv)$. $\square$

Remark 2.3.1 *Condition (i) is identical to the condition used in Theorem 2.5 of [151].*

Theorem 2.3.2 *Consider the following conditions:*

(R_1) *For any sequence $\{(x_k, v_k, t_k)\} \subset (\mathrm{bd}C \times R^n \times R_{++} \cup \mathrm{int}C \times \{0\} \times R_{++})$ such that $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ with $v \in N_C(\bar{x}) \backslash \{0\}$, it holds that*

$$\liminf_{k \to +\infty} \frac{d(x_k + t_k v_k, C)}{t_k} > 0. \tag{2.3.24}$$

(R_2) *For any sequence $\{(x_k, v_k, t_k)\} \subset \mathrm{gph}N_C^P \times R_{++}$ such that $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ with $v \neq 0$, the inequality (2.3.24) holds.*

The following statements are true:

(a) If condition (R_1) holds or C is convex, then condition (R_2) holds.

(b) Condition (R_1) implies that any condition from (i) to (iv) of Theorem 2.3.1 is not only sufficient but also necessary for ϕ to be a local error bound for C at $\bar{x}$.

(c) *Condition* (R_2) *implies that any condition from* (ii) *to* (iv) *of Theorem 2.3.1 is not only sufficient but also necessary for* ϕ *to be a local error bound for* C *at* $\bar{x}$.

(d) *If condition* (R_1) *holds, then there exists* $\varepsilon > 0$ *such that*

$$T_C(x) \cap N_C(\bar{x}) = \{0\} \qquad \forall x \in \mathrm{bd}C \textit{ with } \|x - \bar{x}\| \leq \varepsilon. \tag{2.3.25}$$

(e) *If condition* (R_2) *holds, then there exists* $\varepsilon > 0$ *such that*

$$T_C(x) \cap N_C(x) = \{0\} \qquad \forall x \in \mathrm{bd}C \textit{ with } \|x - \bar{x}\| \leq \varepsilon. \tag{2.3.26}$$

Proof. Since $v \in N_C(\bar{x})$ for any v in condition (R_2), and

$$\mathrm{gph}N_C^P \times R_{++} \subset (\mathrm{bd}C \times R^n \times R_{++} \cup \mathrm{int}C \times \{0\} \times R_{++}),$$

we have $(R_1) \implies (R_2)$. Let $\{(x_k, v_k, t_k)\} \subset \mathrm{gph}N_C^P \times R_{++}$ be a sequence such that $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ with $v \neq 0$. Suppose that C is convex. By Proposition 6.17 of [141], we have for any $(x, v) \in \mathrm{gph}N_C^P$,

$$x \in P_C(x + tv) \qquad \forall t \geq 0.$$

Thus, we have

$$\liminf_{k\to+\infty} \frac{d(x_k + t_k v_k, C)}{t_k} = \liminf_{k\to+\infty} \frac{\|x_k + t_k v_k - x_k\|}{t_k} = \|v\| > 0,$$

implying that condition (R_2) holds. Thus, statement (a) is true. We now show statement (b) is true. Let $\{(x_k, v_k, t_k)\} \subset (\mathrm{bd}C \times R^n \times R_{++} \cup \mathrm{int}C \times \{0\} \times R_{++})$ be a sequence such that $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ with $v \in N_C(\bar{x})\backslash\{0\}$. If ϕ is a local error bound for C at $\bar{x}$, then there exists some $\tau > 0$ such that

$$\liminf_{k\to+\infty} \frac{\phi(x_k + t_k v_k)}{t_k} \geq \tau \liminf_{k\to+\infty} \frac{d(x_k + t_k v_k, C)}{t_k}. \tag{2.3.27}$$

In view of Theorem 2.3.1, it follows easily from (2.3.27) that statement (b) is true. We can show statement (c) in a similar way by noticing the fact that condition (R_2) remains the same when $\mathrm{gph}N_C^P$ in this condition is replaced by $\mathrm{gph}\widehat{N}_C$ or $\mathrm{gph}N_C$, see Theorem 2.3.1.

Suppose by contradiction that (2.3.25) does not hold, i.e., there exist some sequences $\{x_k\} \subset \mathrm{bd}C$ and $\{v_k\} \subset R^n$ such that $x_k \to \bar{x}$, $v_k \in T_C(x_k) \cap N_C(\bar{x})$, and $\|v_k\| = 1$.

Without loss of generality, we can assume that $v_k \to v$. It is clear that $v \in N_C(\bar{x})$ and $\|v\| = 1$. By definition, $v_k \in T_C(x_k)$ if and only if

$$\liminf_{t \to 0+} \frac{d(x_k + tv_k, C)}{t} = 0. \tag{2.3.28}$$

Let $\varepsilon_k \to 0+$. Due to (2.3.28), we can find a sequence $\{t_k\} \subset R_{++}$ such that $t_k \to 0$ and for each k,

$$\frac{d(x_k + t_k v_k, C)}{t_k} \leq \varepsilon_k.$$

Since $\varepsilon_k \to 0+$, we have

$$\liminf_{k \to +\infty} \frac{d(x_k + t_k v_k, C)}{t_k} = 0,$$

contradicting to condition (R_1). Therefore, statement (d) is true. Similarly, we can show that statement (e) is true. This completes the proof. $\square$

Example 2.2.3 can be used to illustrate that when criteria obtained in Section 2.2 fail, Theorem 2.3.1 can be utilized to verify the existence of local error bounds. Below is another example.

Example 2.3.1 *Let $p > 0$ be a constant and let $\phi : R \to R_+$ be defined by*

$$\phi(x) = \begin{cases} ((x^3 \sin \frac{1}{x})_+)^p & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Set $C := \{x \in R \mid \phi(x) = 0\}$. Explicitly we have

$$C = -R_+ \cup (\cup_{k=1}^{\infty} [a_k, b_k])$$

and

$$\text{bd}C = \{0\} \cup (\cup_{k=1}^{\infty} \{a_k, b_k\}),$$

where $a_k = \dfrac{1}{2k\pi}$ and $b_k = \dfrac{1}{(2k-1)\pi}$. Let $\bar{x} = 0 \in C$. Since ϕ is continuously differentiable on $R \backslash C$, each open interval (b_{k+1}, a_k) contains at least one local maximizer c_k, at which it is necessary that $|\nabla \phi|(c_k) = 0$. Since $a_k \to 0$ and $b_k \to 0$ as $k \to +\infty$, we have $c_k \to 0$ as $k \to +\infty$. Therefore, no condition in Theorem 2.2.1 can be applied to check whether or not ϕ is a local error bound for C at $\bar{x} = 0$.

In what follows, we apply condition (v) in Theorem 2.3.1 to verify that ϕ is a local error bound for C at $\bar{x}$ if and only if $p \leq \frac{2}{3}$. By direct calculation, we have $N_C(\bar{x}) = R$ and

$$\{(x, v, t) \in R^n \times R^n \times R_{++} \mid x \in P_C(x + tv)\} = A_1 \cup A_2 \cup A_3,$$

where $A_1 = C \times \{0\} \times R_{++}$, $A_2 = \cup_{k=1}^{\infty}\{(a_k, v, \frac{\beta}{-4k\pi v(2k+1)}) \mid v < 0,\ 0 < \beta \leq 1\}$, *and* $A_3 = \{(b_1, v, t) \mid v > 0, t > 0\} \bigcup \cup_{k=2}^{\infty}\{(b_k, v, \frac{\beta}{4\pi v(k-1)(2k-1)}) \mid v > 0,\ 0 < \beta \leq 1\}$. *Let* $\{(x_k, v_k, t_k)\} \subset A_1 \cup A_2 \cup A_3$ *be a sequence such that* $(x_k, v_k, t_k) \to (\bar{x}, v, 0)$ *with* $v \neq 0$. *Without loss of generality, we can assume in the case of* $v = -1$ *that,*

$$x_k = a_k, \quad v_k < 0, \quad t_k = \frac{\beta_k}{-4k\pi v_k(2k+1)} \text{ with some } \beta_k \in (0, 1].$$

Thus, we have in the case of $p = \frac{2}{3}$,

$$\begin{aligned}
& \liminf_{k\to+\infty} \frac{\phi(x_k + t_k v_k)}{t_k} \\
= \quad & \liminf_{k\to+\infty} \frac{\left(\frac{1}{2k\pi} - \frac{\beta_k}{4k\pi(2k+1)}\right)^2 \left(\sin \frac{1}{\frac{1}{2k\pi} - \frac{\beta_k}{4k\pi(2k+1)}}\right)^{\frac{2}{3}}}{\frac{\beta_k}{-4k\pi v_k(2k+1)}} \\
= \quad & \liminf_{k\to+\infty} \frac{\frac{1}{(4k\pi)^2}(2 - \frac{\beta_k}{2k+1})^2 \sin^{\frac{2}{3}}\left(2k\pi + \frac{2\beta_k\pi}{4 + \frac{2-\beta_k}{k}}\right)}{\frac{1}{-8v_k\pi}\frac{1}{k}\frac{1}{k+\frac{1}{2}}\beta_k} \\
= \quad & \liminf_{k\to+\infty} \frac{2}{\pi}\frac{\sin^{\frac{2}{3}}\left(\frac{2\beta_k\pi}{4 + \frac{2-\beta_k}{k}}\right)}{\beta_k} \geq \frac{2}{\pi}.
\end{aligned}$$

If $v = 1$, *we can similarly obtain that in the case of* $p = \frac{2}{3}$, $\liminf\limits_{k\to+\infty} \dfrac{\phi(x_k + t_k v_k)}{t_k} \geq \dfrac{2}{\pi}$. *Thus, by Theorem 2.2.1,* ϕ *is a local error bound for* C *at* $\bar{x}$ *when* $p \leq \frac{2}{3}$. *When* $p > \frac{2}{3}$ *and* $x_k = \dfrac{1}{2k\pi + \frac{\pi}{2}}$, *we have*

$$\frac{\phi(x_k)}{d(x_k, C)} = \frac{x_k^{3p-2}x_k^2}{x_k - b_{k+1}} = x_k^{3p-2}\frac{\frac{1}{4\pi^2}\frac{1}{(k+\frac{1}{4})^2}}{\frac{1}{8\pi}\frac{1}{k+\frac{1}{2}}\frac{1}{k+\frac{1}{4}}} \to 0,$$

indicating that ϕ *cannot be a local error bound for* C *at* $\bar{x}$.

Chapter 3

Necessary and Sufficient Conditions for Exact Penalty Functions

3.1 Introduction

Throughout this chapter, let C be a subset of R^n and let $\phi : R^n \to R_+ \cup \{+\infty\}$ be a lower semicontinuous function with the property that $x \in C$ if and only if $\phi(x) = 0$. Moreover, let $\bar{x} \in C$ be a fixed point and let $f_0 : R^n \to R$ be at least continuously differentiable.

Consider the constrained optimization problem

$$\text{(P)} \quad \min \quad f_0(x) \quad \text{s.t.} \quad x \in C,$$

and the unconstrained optimization problem

$$\min \quad f_0(x) + \mu\phi(x) \quad \text{s.t.} \quad x \in R^n,$$

where μ is a nonnegative real number. In harmony with the general theory of penalty functions, it is convenient to think of μ as a penalty parameter, ϕ as a penalty term, and $f_0 + \mu\phi$ as a penalty function. It is worth mentioning that penalty functions of the form $f_0 + \mu\phi$ include many penalty functions in the context of nonlinear programming as special cases, such as the classical l_1 penalty functions originated with Zangwill [169] and Eremin [48], and the lower order l_p $(0 < p < 1)$ penalty functions introduced by Luo et al. [105] and further studied in [80, 144, 145, 160, 161, 110, 162].

We say that the penalty function $f_0 + \mu\phi$ is *exact* at $\bar{x}$ if it has an unconstrained local minimum at $\bar{x}$ for a finite penalty parameter (and hence for all finite and larger values of the penalty parameter). It is to be note that if $f_0 + \mu\phi$ is exact at $\bar{x}$, then by definition, (P) has a local minimum at $\bar{x}$. When ϕ happens to be the indicator function δ_C of C, it is clear to see that $f_0 + \mu\phi$ is exact at $\bar{x}$ if and only if (P) has a local minimum at $\bar{x}$. In general, however, $f_0 + \mu\phi$ may not be exact at $\bar{x}$ even if (P) has a local minimum at $\bar{x}$, unless some conditions are satisfied. Conditions ensuring that ϕ is a local error bound for C are such conditions, which do not depend on the objective function f_0 and have been investigated in last chapter.

In this chapter, we focus on another type of conditions ensuring exactness of penalty functions, which relies not only on ϕ but also on the objective function f_0. More precisely, we study first- and second-order necessary as well as sufficient conditions for $f_0 + \mu\phi$ to be exact at $\bar{x}$. When ϕ can be formulated as the composition of a lower semicontinuous convex function and a continuously differentiable function, this problem has been extensively studied under the framework of convex composition optimization, see, e.g., Ioffe [83], Fletcher [60], Jeyakumar and Yang [88], Burke [24, 26], Burke and Poliquin [31]. By way of epi-derivatives, this problem has been also studied by Rockafellar [139, 138] who considered the case where ϕ is the composition of a piecewise linear-quadratic function and a continuously differentiable function. Because epi-derivatives relies on epigraphs and epigraphical convergence, they have a stronger basis in geometric approximation and correspondingly a greater potential of stability and robustness than classical derivatives relying on graphs and graphical convergence, as pointed out by Rockafellar [139]. Note that the classical l_1 penalty functions are covered by the mentioned references, but the lower order l_p $(0 < p < 1)$ penalty functions cannot be covered by any of the mentioned references because in these functions, ϕ is formulated as the composition of a non-Lipschitzian and non-convex function and a continuously differentiable function.

To cover at least the lower order l_p $(0 < p < 1)$ penalty functions in our consideration, we merely require that ϕ is a non-negative lower semi-continuous function. We present our conditions in terms of three epi-derivatives: subderivatives, second-order subderivatives, and parabolic second-order subderivatives. These notions have been successfully utilized to formulate very tight first- and second-order optimality conditions for an extended real-valued function to attain a local minimum, see in particular Theorems 10.1, 13.24 and 13.66 of Rockafellar and Wets [141].

The outline of this chapter is as follows. In Section 3.2, we study some basic properties of subderivatives, second subderivatives and parabolic subderivatives of ϕ, and pay our attention to derivative kernels consisting of directions at which these derivatives vanish. In Section 3.3, we give a number of first- and second-order necessary and sufficient conditions for $f_0 + \mu\phi$ to be exact at $\bar{x}$.

3.2 Derivative Kernels

In this section, we study subderivatives, second subderivatives, and parabolic subderivatives of the penalty terms for (NLP).

To begin with, we give some basic properties on these derivatives of ϕ as follows.

Lemma 3.2.1 *The following statements are true:*

(*i*) *Let $v \in R^n$ and let $w \in \mathrm{dom}d\phi(\bar{x})$. The functions*

$$d\phi(\bar{x})(\cdot), \quad d^2\phi(\bar{x} \mid v)(\cdot), \quad d^2\phi(\bar{x} \mid v)(\cdot) : R^n \to \overline{R}$$

are lower semi-continuous.

(*ii*) *Let $\tau > 0$, let $v, z \in R^n$, and let $w \in \mathrm{dom}d\phi(\bar{x})$. Then*

$$\begin{aligned}
&d(\tau\phi)(\bar{x}) = \tau d\phi(\bar{x});\\
&d^2(\tau\phi)(\bar{x} \mid v) = \tau d^2\phi(\bar{x} \mid \tfrac{v}{\tau});\\
&d^2(\tau\phi)(\bar{x})(w \mid \cdot) = \tau d^2\phi(\bar{x})(w \mid \cdot);\\
&d\phi(\bar{x})(\tau z) = \tau d\phi(\bar{x})(z);\\
&d^2\phi(\bar{x} \mid v)(\tau z) = \tau^2 d^2\phi(\bar{x} \mid v)(z);\\
&d^2\phi(\bar{x} \mid 0)(z) = 2[d\phi^{\frac{1}{2}}(\bar{x})(z)]^2.
\end{aligned}$$

Proof. Recall that the notation $\phi^{\frac{1}{2}}$ has been defined by (1.3.19). All the results are easily obtained from the definitions of subderivative, second-order subderivative, and parabolic subderivative, if Propositions 13.5 and 13.64 in [141] are taken into account. This completes the proof. □

It is clear to see that $\bar{x}$ is a global minimum of ϕ. Therefore, the necessary optimality conditions in Theorems 10.1, 13.24 and 13.66 of Rockafellar and Wets [141] can be applied immediately, which give the following lemma.

Lemma 3.2.2 *The following statements are true:*

(*i*) $d\phi(\bar{x})(w) \geq 0$ *for all* $w \in R^n$.

(*ii*) $d^2\phi(\bar{x} \mid 0)(w) \geq 0$ *for all* $w \in R^n$.

(*iii*) $d^2\phi(\bar{x})(w \mid z) \geq 0$ *for all* w *with* $d\phi(\bar{x})(w) = 0$, *and all* $z \in R^n$.

As can be seen from our latter developments, the subderivative kernels, consisting of the vanishing directions of subderivatives of ϕ at $\bar{x}$, will play a key role. We formally give their definitions as follows.

Definition 3.2.1 (*i*) *The kernel of the subderivative of* ϕ *at* $\bar{x}$ *is defined by*

$$\mathrm{ker} d\phi(\bar{x}) := \{w \in R^n \mid d\phi(\bar{x})(w) = 0\}.$$

(*ii*) *The kernel of the parabolic subderivative of* ϕ *at* $\bar{x}$ *for the vector* $w \in \mathrm{ker} d\phi(\bar{x})$ *is defined by*

$$\mathrm{ker} d^2\phi(\bar{x})(w \mid \cdot) := \{z \in R^n \mid d^2\phi(\bar{x})(w \mid z) = 0\}.$$

For convenience, we set $\mathrm{ker} d^2\phi(\bar{x})(w \mid \cdot) = \emptyset$ *when* $w \notin \mathrm{ker} d\phi(\bar{x})$.

(*iii*) *The kernel of the second subderivative of* ϕ *at* $\bar{x}$ *for the vector* 0 *is defined by*

$$\mathrm{ker} d^2\phi(\bar{x} \mid 0) := \{w \in R^n \mid d^2\phi(\bar{x} \mid 0)(w) = 0\}.$$

Remark 3.2.1 *Since* $d^2\phi(\bar{x} \mid 0)(w) = 2[d\phi^{\frac{1}{2}}(\bar{x})(w)]^2$ *for all* $w \in R^n$ *(see Lemma 3.2.1 (i)), we have*

$$\mathrm{ker} d^2\phi(\bar{x} \mid 0) = \mathrm{ker} d\phi^{\frac{1}{2}}(\bar{x}). \tag{3.2.1}$$

As such, the first two kernels in Definition 3.2.1 seem to be enough for applications. But the notation $\mathrm{ker} d\phi^{\frac{1}{2}}(\bar{x})$ *itself does not indicate any connection with the second subderivative, so we insist on using the notation* $\mathrm{ker} d^2\phi(\bar{x} \mid 0)$.

Let $0 \le p \le 1$. Recall that the p-th order function ϕ^p with $p > 0$ has been defined by (1.3.19), and that ϕ^p is interpreted as the indicator function δ_C when $p = 0$. Some basic results on $\mathrm{ker} d\phi^p(\bar{x})$ and $\mathrm{ker} d^2\phi^p(\bar{x})(w \mid \cdot)$ are summarized respectively in the next two propositions.

Proposition 3.2.1 *The following statements are true:*

(i) $\mathrm{ker} d\phi(\bar{x})$ *is a nonempty closed cone in* R^n *with the property that* $w \in \mathrm{ker} d\phi(\bar{x})$ *if and only if there exist* $t_k \to 0+$ *and* $w_k \to w$ *such that*

$$\frac{\phi(\bar{x} + t_k w_k)}{t_k} \to 0. \tag{3.2.2}$$

(ii) $T_C(\bar{x}) \subset \mathrm{ker} d\phi(\bar{x})$. *The equality holds if* ϕ *is a local error bound for* C *at* $\bar{x}$.

(iii) *Let* $p \ge 0$ *and let* $p' > p$. *Then*

$$\mathrm{ker} d\phi^p(\bar{x}) \subset \mathrm{dom} d\phi^p(\bar{x}) \subset \mathrm{ker} d\phi^{p'}(\bar{x}).$$

Proof. Since the subderivative function $d\phi(\bar{x})$ is lower semi-continuous and positively homogenous (see Lemma 3.2.1), $\ker d\phi(\bar{x})$ is clearly a nonempty closed cone in R^n. From the definition of subderivative, it is easy to verify that $w \in \ker d\phi(\bar{x})$ if and only if there exist $t_k \to 0+$ and $w_k \to w$ such that (3.2.2) holds. From statement (i) and the definition of the tangent cone, we can easily get

$$T_C(\bar{x}) \subset \ker d\phi(\bar{x}).$$

To get the converse inclusion, we suppose that ϕ is a local error bound for C at $\bar{x}$, i.e., there exist $\tau > 0$ and $\delta > 0$ such that for all $x \in R^n$ with $\|x - \bar{x}\| \le \delta$,

$$\tau d_C(x) \le \phi(x). \tag{3.2.3}$$

Let $w \in \ker d\phi(\bar{x})$. It follows from statement (i) and (3.2.3) that, there exist $t_k \to 0+$ and $w_k \to w$ such that

$$\frac{d_C(\bar{x} + t_k w_k)}{t_k} \to 0.$$

Since C is closed, there exists $y_k \in C$ for each k such that $\bar{x} + t_k w_k - y_k = o(t_k)$. That is, $\bar{x} + t_k(w_k - o(t_k)/t_k) = y_k \in C$ for all k. This implies by definition that $w \in T_C(\bar{x})$. Thus, statement (ii) is true. We now show statement (iii). Let $p \ge 0$ and let $p' \geq p$. It is clear to see that $\mathrm{ker}d\phi^p(\bar{x}) \subset \mathrm{dom}d\phi^p(\bar{x})$. Let $w \in \mathrm{dom}d\phi^p(\bar{x})$. We have $0 \leq d\phi^p(\bar{x})(w) < +\infty$. By the definition of subderivative, we can find sequences $t_k \to 0+$ and $w_k \to w$ such that

$$\frac{\phi^p(\bar{x}+t_k w_k) - \phi^p(\bar{x})}{t_k} = \frac{\phi^p(\bar{x}+t_k w_k)}{t_k} \to d\phi^p(\bar{x})(w),$$

which implies that $\phi(\bar{x}+t_k w_k) \to 0$. Due to $p'-p>0$ and the finiteness of $d\phi^p(\bar{x})(w)$, we have

$$\frac{\phi^{p'}(\bar{x}+t_k w_k)}{t_k} = \frac{\phi^p(\bar{x}+t_k w_k)\phi^{p'-p}(\bar{x}+t_k w_k)}{t_k} \to 0,$$

implying by statement (i) that $w \in \ker d\phi^{p'}(\bar{x})$. This completes the proof. $\square$

Proposition 3.2.2 *The following statements are true:*

(i) *Let $w \in \mathrm{ker}d\phi(\bar{x})$. The set $\mathrm{ker}d^2\phi(\bar{x})(w \mid \cdot)$ is a closed (possibly empty) subset of R^n with the property that $z \in \mathrm{ker}d^2\phi(\bar{x})(w \mid \cdot)$ if and only if there exist $t_k \to 0+$ and $z_k \to z$ such that*

$$\frac{\phi(\bar{x}+t_k w + \frac{1}{2}t_k^2 z_k)}{t_k^2} \to 0. \tag{3.2.4}$$

(ii) $\mathrm{ker}d^2\phi(\bar{x})(w \mid \cdot) = \mathrm{ker}d\phi(\bar{x})$ *when $w = 0$.*

(iii) $\mathrm{ker}d^2\phi(\bar{x})(w \mid \cdot) = \emptyset$ *when $w \in \mathrm{ker}d\phi(\bar{x})\backslash\mathrm{ker}d\phi^{\frac{1}{2}}(\bar{x})$.*

(iv) *For any $w \in T_C(\bar{x})$, $T_C^2(\bar{x} \mid w) \subset \mathrm{ker}d^2\phi(\bar{x})(w \mid \cdot)$. The equality holds if ϕ is a local error bound for C at $\bar{x}$.*

(v) *Let $p \geq 0$ and let $p' > p$. Then for any $w \in \mathrm{ker}d\phi^p(\bar{x})$,*

$$\mathrm{ker}d^2\phi^p(\bar{x})(w \mid \cdot) \subset \mathrm{dom}d^2\phi^p(\bar{x})(w \mid \cdot) \subset \mathrm{ker}d^2\phi^{p'}(\bar{x})(w \mid \cdot).$$

Proof. Let $w \in \mathrm{ker}d\phi(\bar{x})$. Since the function $d^2\phi(\bar{x})(w \mid \cdot) : R^n \to \overline{R}$ is lower semicontinuous (see Lemma 3.2.1), the set $\mathrm{ker}d^2\phi(\bar{x})(w \mid \cdot)$, though possibly empty, is a closed subset of R^n. By the definition of parabolic subderivative, it is easy to verify that $z \in \mathrm{ker}d^2\phi(\bar{x})(w \mid \cdot)$ if and only if there exist $t_k \to 0+$ and $z_k \to z$ such that (3.2.4) holds. In view of this characterization, the rest results can be obtained in a similar way as we have done in the proof of Proposition 3.2.1. So the details are omitted. $\square$

It is interesting to establish the equivalence of the polar cone of $\ker df(\bar{x})$ with the positive hull of the regular subgradient set $\widehat{\partial} f(\bar{x})$ under the circumstance that the function $f : R^n \to \overline{R}$ has a local minimum at $\bar{x}$ with $f(\bar{x})$ finite, where

$$\ker df(\bar{x}) := \{w \in R^n \mid df(\bar{x})(w) = 0\}.$$

Theorem 3.2.1 *Suppose that $f : R^n \to \overline{R}$ has a local minimum at $\bar{x}$ with $f(\bar{x})$ finite. Then we have*

(i) $[\ker df(\bar{x})]^* = \text{pos}(\widehat{\partial} f(\bar{x}))$.

(ii) $\widehat{\partial} f(\bar{x})$ *is a cone if and only if* $[\text{dom} df(\bar{x})]^* = [\ker df(\bar{x})]^*$.

Proof. It follows from Theorem 10.1 of [141] that $df(\bar{x}) \geq 0$ or equivalently $0 \in \widehat{\partial} f(\bar{x})$. By Theorem 8.9 of [141], we have

$$v \in \widehat{\partial} f(\bar{x}) \Longleftrightarrow (v, -1) \in \widehat{N}_{\text{epi}f}(\bar{x}, f(\bar{x})) = T_{\text{epi}f}(\bar{x}, f(\bar{x}))^*. \tag{3.2.5}$$

According to Theorem 8.2 of [141], we have

$$\text{epi}\, df(\bar{x}) = T_{\text{epi}f}(\bar{x}, f(\bar{x})), \tag{3.2.6}$$

which implies that

$$\ker df(\bar{x}) \times \{0\} \subset T_{\text{epi}f}(\bar{x}, f(\bar{x})). \tag{3.2.7}$$

It follows from (3.2.5) and (3.2.7) that, if $v \in \widehat{\partial} f(\bar{x})$, then $v \in [\ker df(\bar{x})]^*$ because $\langle v, u\rangle = \langle (v, -1), (u, 0)\rangle \leq 0$ for all $u \in \ker df(\bar{x})$. This implies that $\widehat{\partial} f(\bar{x}) \subset [\ker df(\bar{x})]^*$. Since $[\ker df(\bar{x})]^*$ is a cone, we have $\text{pos}(\widehat{\partial} f(\bar{x})) \subset [\ker df(\bar{x})]^*$.

To show the converse inclusion, let $A = \{u \in R^n \mid df(\bar{x})(u) \leq 1\}$ and $E = \{u \in R^n \mid 0 < df(\bar{x})(u) \leq 1\}$. It is clear that $\ker df(\bar{x}) \cap E = \emptyset$. Since $df(\bar{x}) \geq 0$, we have $A = \ker df(\bar{x}) \cup E$. It is easy to see from (3.2.6) that $df(\bar{x})$ is lower semicontinuous and positively homogeneous. This implies that $\ker df(\bar{x})$ and A are closed sets in R^n with $\ker df(\bar{x})$ being a cone, and together with (3.2.6) also implies that

$$T_{\text{epi}f}(\bar{x}, f(\bar{x})) = \{\lambda(u, 1) \mid u \in A, \lambda > 0\} \cup \{(u, 0) \mid u \in \ker df(\bar{x})\}. \tag{3.2.8}$$

If there exists some $\tau > 0$ such that for all $v \in [\ker df(\bar{x})]^*$ and $u \in E$,

$$\langle v, u\rangle \leq \tau \|v\|, \tag{3.2.9}$$

then by (3.2.5) and (3.2.8), we have $[\ker df(\bar{x})]^* \subset \mathrm{pos}(\widehat{\partial} f(\bar{x}))$, because for every $v \in [\ker df(\bar{x})]^* \backslash \{0\}$, we have

$$\langle (\frac{v}{\tau\|v\|}, -1), (u, 0)\rangle = \langle \frac{v}{\tau\|v\|}, u\rangle \le 0 \quad \forall\, u \in \ker df(\bar{x}),$$

and by (3.2.9), we also have

$$\langle (\frac{v}{\tau\|v\|}, -1), \lambda(u, 1)\rangle = \lambda[\langle \frac{v}{\tau\|v\|}, u\rangle - 1] \le 0 \quad \forall\, u \in A,\ \forall\, \lambda > 0.$$

Now, it remains to show the existence of some $\tau > 0$ such that (3.2.9) holds. Without loss of generality, we can assume that E is unbounded, otherwise (3.2.9) holds automatically when $\tau > 0$ is large enough. We claim that there exist two positive numbers ρ and σ such that for all $x \in E$ with $\|x\| > \rho$,

$$d(x, \ker df(\bar{x})) \le \sigma. \tag{3.2.10}$$

Suppose by contradiction that (3.2.10) does not hold. Then, we can find a sequence $x_\nu \in E$ such that $\|x_\nu\| \to +\infty$ and $d_{\ker df(\bar{x})}(x_\nu) \to +\infty$. Without loss of generality, we can assume that $\frac{x_\nu}{\|x_\nu\|} \to u$ for some $u \in R^n$ with $\|u\| = 1$. Since $df(\bar{x})$ is positively homogeneous, E is a subset of R^n with the properties that $\lambda x \in E$ for all $0 < \lambda \le 1$ and $x \in E$, and that for every $x \in R^n$, there exists some $\lambda_0 > 0$ such that $\lambda x \notin E$ for all $\lambda \ge \lambda_0$. Therefore, for any positive number λ, we have $\lambda \frac{x_\nu}{\|x_\nu\|} \in E$ for all sufficiently large ν, implying that $\lambda u \in clE$. Since $E \subset A$ and A is closed, we have $\lambda u \in A$ for all $\lambda > 0$. Observing that $\lambda u \notin E$ when $\lambda \ge \lambda_0$ for some $\lambda_0 > 0$, we have $\lambda u \in \ker df(\bar{x})$ for all $\lambda \ge \lambda_0$ because $\ker df(\bar{x}) \cap E = \emptyset$ and $\ker df(\bar{x}) \cup E = A$. This implies that $\|x_\nu\| u \in \ker df(\bar{x})$ since $\ker df(\bar{x})$ is a cone. Thus, we have $d_{\ker df(\bar{x})}(x_\nu) \le \|x_\nu - \|x_\nu\| u\| \to 0$ which contradicts to the assumption that $d_{\ker df(\bar{x})}(x_\nu) \to +\infty$. Let $\tau = \max\{\rho, \sigma\}$, where $\rho > 0$ and $\sigma > 0$ are such that (3.2.10) holds. Fix a point v in $[\ker df(\bar{x})]^*$ and a point u in E. If $\|u\| \le \rho$, we have

$$\langle v, u\rangle \le \|v\|\|u\| \le \rho\|v\| \le \tau\|v\|.$$

Alternatively, if $\|u\| > \rho$, then by the closedness of $\ker df(\bar{x})$, there exists $\bar{u} \in \ker df(\bar{x})$ such that $d_{\ker df(\bar{x})}(u) = \|u - \bar{u}\|$, and by using the inequality (3.2.10), we also have

$$\langle v, u\rangle \le \langle v, u\rangle - \langle v, \bar{u}\rangle = \langle v, u - \bar{u}\rangle \le \|v\|\|u - \bar{u}\| \le \sigma\|v\| \le \tau\|v\|,$$

where the first inequality follows from the fact that $v \in [\ker df(\bar{x})]^*$ and $\bar{u} \in \ker df(\bar{x})$. By now, we have shown the existence of some $\tau > 0$ such that (3.2.9) holds. Therefore, statement (i) is true.

We now show statement (ii). According to Theorem 8.9 of [141] and its proof therein, we have

$$\widehat{\partial} f(\bar{x})^{\infty} = [\mathrm{dom} df(\bar{x})]^{*}. \tag{3.2.11}$$

Observing that $\widehat{\partial} f(\bar{x})$ a closed and convex set containing the origin (see Theorem 8.6 of [141]), we get from Theorem 3.6 of [141],

$$\widehat{\partial} f(\bar{x})^{\infty} \subset \widehat{\partial} f(\bar{x}). \tag{3.2.12}$$

If $[\mathrm{dom} df(\bar{x})]^{*} = [\ker df(\bar{x})]^{*}$, it then follows from statement (i), (3.2.11), and (3.2.12) that $\mathrm{pos}(\widehat{\partial} f(\bar{x})) \subset \widehat{\partial} f(\bar{x})$, implying that $\widehat{\partial} f(\bar{x})$ is a cone. Conversely, if $\widehat{\partial} f(\bar{x})$ is a cone, then $\widehat{\partial} f(\bar{x})^{\infty} = \widehat{\partial} f(\bar{x}) = \mathrm{pos}(\widehat{\partial} f(\bar{x}))$, which together with (3.2.11) and statement (i) implies that $[\mathrm{dom} df(\bar{x})]^{*} = [\ker df(\bar{x})]^{*}$. This completes the proof. □

3.3 Necessary and Sufficient Conditions

In this section, we study first- and second-order necessary and sufficient conditions for $f_0 + \mu\phi$ to be exact at $\bar{x}$. When second-order conditions are discussed, we assume that f_0 is twice continuously differentiable. Our conditions are presented in terms of subderivatives, second-order subderivatives, and parabolic subderivatives. Because of the differentiability assumption on f_0, the subderivative, the second-order subderivative, and the parabolic subderivatives of $f_0 + \mu\phi$ can be expressed respectively by that of f_0 and ϕ, which makes it possible for us to apply Theorems 10.1, 13.24 and 13.66 of Rockafellar and Wets [141] in a straightforward way.

Theorem 3.3.1 (a) (necessity) *If $f_0 + \mu\phi$ is exact at $\bar{x}$, then*

$$\langle \nabla f_0(\bar{x}), w \rangle \geq 0 \qquad \forall w \in \ker d\phi(\bar{x}). \tag{3.3.13}$$

(b) (sufficiency) *Conversely, if the inequality (3.3.13) is strict when $w \neq 0$, then $f_0+\mu\phi$ is exact at $\bar{x}$.*

Proof. Applying the definition of subderivative, we can easily verify that

$$d(f_0 + \mu\phi)(\bar{x})(w) = \langle \nabla f_0(\bar{x}), w \rangle + \mu d\phi(\bar{x})(w) \qquad \forall \mu \geq 0,\ \forall w \in R^n. \tag{3.3.14}$$

[(necessity)]: By assumption, there exists some $\mu_0 \geq 0$ such that the function $f_0 + \mu_0\phi$ has a local minimum at $\bar{x}$. By Theorem 10.1 of [141], we have

$$d(f_0 + \mu_0\phi)(\bar{x})(w) \geq 0 \qquad \forall w \in R^n.$$

In view of (3.3.14), we have $\langle \nabla f_0(\bar{x}), w\rangle \geq 0$ for all $w \in \ker d\phi(\bar{x})$.

[(sufficiency)]: Suppose that the inequality (3.3.13) is strict when $w \neq 0$, which clearly implies that $-\nabla f_0(\bar{x}) \in [\ker d\phi(\bar{x})]^*$. By Theorem 3.2.1, we have

$$-\nabla f_0(\bar{x}) \in \text{pos}(\widehat{\partial}\phi(\bar{x})).$$

Thus, we can find some $\tau > 0$ such that $-\frac{1}{\tau}\nabla f_0(\bar{x}) \in \widehat{\partial}\phi(\bar{x})$. It follows from Exercise 8.4 of [141] that

$$\langle -\frac{1}{\tau}\nabla f_0(\bar{x}), w\rangle \leq d\phi(\bar{x})(w) \qquad \forall w \in R^n.$$

Let $\bar{\mu} > \tau$. Since the inequality (3.3.13) is strict when $w \neq 0$, we thus have

$$\langle \nabla f_0(\bar{x}), w\rangle + \bar{\mu} d\phi(\bar{x})(w) > 0 \qquad \forall w \neq 0.$$

In view of (3.3.14), we have

$$d(f_0 + \bar{\mu}\phi)(\bar{x})(w) > 0 \qquad \forall w \neq 0.$$

This implies that $f_0 + \bar{\mu}\phi$ admits an unconstrained local minimum at $\bar{x}$. Thus, $f_0 + \mu\phi$ is exact at $\bar{x}$. This completes the proof. □

Theorem 3.3.2 *Suppose that f_0 is twice continuously differentiable.*

(a) (necessity) *If $f_0 + \mu\phi$ is exact at $\bar{x}$, then*

$$-\nabla f_0(\bar{x}) \in [\ker d\phi(\bar{x})]^*, \tag{3.3.15}$$

and there exists $\tau > 0$ such that for all $w \in \ker d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^\perp$,

$$\langle w, \nabla^2 f_0(\bar{x})w\rangle + \tau d^2\phi(\bar{x} \mid -\frac{\nabla f_0(\bar{x})}{\tau})(w) \geq 0. \tag{3.3.16}$$

(b) (sufficiency) *Conversely, if (3.3.15) holds and (3.3.16) holds with a strict inequality when $w \neq 0$, then there exist $\mu \geq 0$, $\varepsilon > 0$ and $\delta > 0$ such that for all $x \in R^n$ with $\|x - \bar{x}\| \leq \delta$,*

$$f_0(x) + \mu\phi(x) \geq f_0(\bar{x}) + \varepsilon\|x - \bar{x}\|^2.$$

Proof. Applying the definition of second-order subderivative, we can easily verify that for any $\mu > 0$ and any $w \in R^n$,

$$d^2(f_0 + \mu\phi)(\bar{x} \mid 0)(w) = \langle w, \nabla^2 f_0(\bar{x})w\rangle + \mu d^2\phi(\bar{x} \mid -\frac{\nabla f_0(\bar{x})}{\mu})(w). \tag{3.3.17}$$

[(necessity)]: Let $\tau > 0$ be such that $f_0 + \tau\phi$ has a local minimum at $\bar{x}$. Theorem 3.3.1 gives that $-\nabla f_0(\bar{x}) \in [\mathrm{ker} d\phi(\bar{x})]^*$. From Theorem 13.24 of [141], we have

$$d^2(f_0 + \tau\phi)(\bar{x} \mid 0)(w) \geq 0 \qquad \forall w \in R^n.$$

This together with (3.3.17) implies that the inequality (3.3.16) holds not only for w in the set $\mathrm{ker} d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^\perp$, but for all $w \in R^n$.

[(sufficiency)]: Suppose that $-\nabla f_0(\bar{x}) \in [\mathrm{ker} d\phi(\bar{x})]^*$ and that there exists $\tau > 0$ such that for all $w \in \mathrm{ker} d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^\perp$ with $w \neq 0$,

$$\langle w, \nabla^2 f_0(\bar{x})w\rangle + \tau d^2\phi(\bar{x} \mid -\frac{\nabla f_0(\bar{x})}{\tau})(w) > 0. \tag{3.3.18}$$

By Theorem 3.2.1, we have $-\nabla f_0(\bar{x}) \in \mathrm{pos}(\widehat{\partial}\phi(\bar{x}))$. It then follows from Exercise 8.4 of [141] that, there exists $\tau_0 > 0$ such that

$$\langle \nabla f_0(\bar{x}), w\rangle + \tau_0 d\phi(\bar{x})(w) \geq 0 \qquad \forall w \in R^n. \tag{3.3.19}$$

Let $\mu > \max\{\tau, \tau_0\}$ and let $w \in R^n \backslash \{0\}$. Since $\mu > \tau_0$ and $d\phi(\bar{x}) \geq 0$, it follows from (3.3.14) and (3.3.19) that

$$d(f_0 + \mu\phi)(\bar{x})(z) \geq 0 \qquad \forall z \in R^n. \tag{3.3.20}$$

If $w \notin \mathrm{ker} d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^\perp$, then, from (3.3.14) and (3.3.19), we have

$$d(f_0 + \mu\phi)(\bar{x})(w) = \langle \nabla f_0(\bar{x}), w\rangle + \mu d\phi(\bar{x})(w) > 0,$$

which implies by Proposition 13.5 of [141] that

$$d^2\phi(\bar{x} \mid -\frac{\nabla f_0(\bar{x})}{\mu})(w) = +\infty.$$

This together with (3.3.17) implies that

$$d^2(f_0 + \mu\phi)(\bar{x} \mid 0)(w) = +\infty > 0 \qquad \forall w \notin \mathrm{ker} d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^\perp.$$

Since $\mu > \tau$, by applying the definition of second-order subderivative, we have

$$\mu d^2\phi(\bar{x} \mid -\frac{\nabla f_0(\bar{x})}{\mu}) = d^2(\mu\phi)(\bar{x} \mid -\nabla f_0(\bar{x})) \geq d^2(\tau\phi)(\bar{x} \mid -\nabla f_0(\bar{x})) = \tau d^2\phi(\bar{x} \mid -\frac{\nabla f_0(\bar{x})}{\tau}).$$

Thus, it follows from (3.3.17) and (3.3.18) that

$$d^2(f_0 + \mu\phi)(\bar{x} \mid 0)(w) > 0 \qquad \forall w \in (\text{ker}d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^\perp)\backslash\{0\}.$$

Therefore, we have

$$d^2(f_0 + \mu\phi)(\bar{x} \mid 0)(w) > 0 \qquad \forall w \neq 0. \tag{3.3.21}$$

In view of (3.3.20) and (3.3.21), all results now follow readily from Theorem 13.24 of [141]. This completes the proof. □

Corollary 3.3.1 *Suppose that f_0 is twice continuously differentiable with $\nabla f_0(\bar{x}) = 0$.*

(a) (necessity) *If $f_0 + \mu\phi$ is exact at $\bar{x}$, then there exists $\tau > 0$ such that for all $w \in \text{ker}d\phi(\bar{x})$,*

$$\langle w, \nabla^2 f_0(\bar{x})w\rangle + \tau d^2\phi(\bar{x} \mid 0)(w) \geq 0. \tag{3.3.22}$$

(b) (sufficiency) *Conversely, if (3.3.22) holds with a strict inequality when $w \neq 0$, then there exist $\mu \geq 0$, $\varepsilon > 0$ and $\delta > 0$ such that for all $x \in R^n$ with $\|x - \bar{x}\| \leq \delta$,*

$$f_0(x) + \mu\phi(x) \geq f_0(\bar{x}) + \varepsilon\|x - \bar{x}\|^2.$$

Remark 3.3.1 *In view of (3.2.1), the necessary condition in Corollary 3.3.1* (a) *can be weakened as each of the following equivalent conditions:*

(a_1) $\langle w, \nabla^2 f_0(\bar{x})w\rangle \geq 0 \quad \forall w \in \text{ker}d^2\phi(\bar{x} \mid 0)$.

(a_2) $\langle w, \nabla^2 f_0(\bar{x})w\rangle \geq 0 \quad \forall w \in \text{ker}d^{\frac{1}{2}}\phi(\bar{x})$.

Theorem 3.3.3 *Suppose that f_0 is twice continuously differentiable.*

(a) (necessity) *If $f_0 + \mu\phi$ is exact at $\bar{x}$, then*

$$-\nabla f_0(\bar{x}) \in [\text{ker}d\phi(\bar{x})]^*, \tag{3.3.23}$$

and there exists $\tau > 0$ such that for all $w \in \text{ker}d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^\perp$,

$$\langle w, \nabla^2 f_0(\bar{x})w\rangle + \inf_z \left\{\langle \nabla f_0(\bar{x}), z\rangle + \tau d^2\phi(\bar{x})(w \mid z)\right\} \geq 0. \tag{3.3.24}$$

(b) (sufficiency) *Suppose that (3.3.23) holds and (3.3.24) holds with a strict inequality when $w \neq 0$. If there exists $\tau_0 \geq \tau$ such that ϕ is parabolically regular at $\bar{x}$ for the vector $-\frac{1}{\tau_0}\nabla f_0(\bar{x})$, then there exist $\mu \geq 0$, $\varepsilon > 0$ and $\delta > 0$ such that for all $x \in R^n$ with $\|x - \bar{x}\| \leq \delta$,*

$$f_0(x) + \mu\phi(x) \geq f_0(\bar{x}) + \varepsilon\|x - \bar{x}\|^2.$$

Proof.[(necessity)]: It follows from Proposition 13.64 of [141] that for any $\mu > 0$ and any $w \in \ker d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^{\perp}$,

$$\inf_z \left\{ \langle \nabla f_0(\bar{x}), z \rangle + \mu d^2\phi(\bar{x})(w \mid z) \right\} \geq \mu d^2\phi(\bar{x} \mid -\frac{\nabla f_0(\bar{x})}{\mu})(w).$$

In view of this inequality and Theorem 3.3.2, the results in (a) follow readily.

[(sufficiency)]: Suppose that $-\nabla f_0(\bar{x}) \in [\ker d\phi(\bar{x})]^*$, and that there exists $\tau > 0$ such that for all $w \in \ker d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^{\perp}$ with $w \neq 0$,

$$\langle w, \nabla^2 f_0(\bar{x})w \rangle + \inf_z \left\{ \langle \nabla f_0(\bar{x}), z \rangle + \tau d^2\phi(\bar{x})(w \mid z) \right\} > 0. \tag{3.3.25}$$

Let $\tau_0 \geq \tau$ be such that ϕ is parabolically regular at $\bar{x}$ for the vector $-\frac{1}{\tau_0}\nabla f_0(\bar{x})$. Then for any $w \in \ker d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^{\perp}$, the equality

$$\inf_z \left\{ \langle \nabla f_0(\bar{x}), z \rangle + \tau_0 d^2\phi(\bar{x})(w \mid z) \right\} = \tau_0 d^2\phi(\bar{x} \mid -\frac{\nabla f_0(\bar{x})}{\tau_0})(w) \tag{3.3.26}$$

holds. Let $w \in (\ker d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^{\perp})\backslash\{0\}$. From the definition of parabolic subderivative, we have

$$d^2\phi(\bar{x})(w \mid z) \geq 0 \qquad \forall z \in R^n.$$

In view of (3.3.25) and $\tau_0 \geq \tau$, we thus have

$$\langle w, \nabla^2 f_0(\bar{x})w \rangle + \inf_z \left\{ \langle \nabla f_0(\bar{x}), z \rangle + \tau_0 d^2\phi(\bar{x})(w \mid z) \right\} > 0,$$

which, together with (3.3.26), implies that

$$\langle w, \nabla^2 f_0(\bar{x})w \rangle + \tau_0 d^2\phi(\bar{x} \mid -\frac{\nabla f_0(\bar{x})}{\tau_0})(w) > 0.$$

Since $w \in (\ker d\phi(\bar{x}) \cap \nabla f_0(\bar{x})^{\perp})\backslash\{0\}$ is arbitrarily chosen, the results now follow readily from Theorem 3.3.2 (b). This completes the proof. $\square$

Remark 3.3.2 *The necessary conditions in Theorem 3.3.3 (a) can be weakened as follows:*

$$\left.\begin{array}{r}\langle \nabla f_0(\bar{x}),\ w\rangle = 0\\ z \in \mathrm{ker} d^2\phi(\bar{x})(w \mid \cdot)\end{array}\right\} \Longrightarrow \langle \nabla f_0(\bar{x}), z\rangle + \langle w, \nabla^2 f_0(\bar{x})w\rangle \geq 0,$$

or equivalently

$$\left.\begin{array}{r}\langle \nabla f_0(\bar{x}),\ w\rangle = 0\\ z \in \mathrm{clconv}[\mathrm{ker} d^2\phi(\bar{x})(w \mid \cdot)]\end{array}\right\} \Longrightarrow \langle \nabla f_0(\bar{x}), z\rangle + \langle w, \nabla^2 f_0(\bar{x})w\rangle \geq 0. \tag{3.3.27}$$

We end this chapter by mentioning that the kernels of subderivatives and parabolic subderivatives appeared in Theorems 3.2.1 and 3.3.1, and Remark 3.3.2, will be used very often in the next two chapters.

Chapter 4

First- and Second-order Necessary Conditions in Nonlinear Programming via Exact Penalty Functions

4.1 Introduction

In this chapter, we consider first- and second-order necessary condition for the nonlinear programming problem

$$\begin{array}{lll} \text{(NLP)} & \min & f(x) \\ & \text{s.t.} & g_i(x) \le 0, \quad i \in I, \\ & & h_j(x) = 0, \quad j \in J, \end{array}$$

where $I = \{1, 2, \cdots, m\}$, $J = \{m+1, m+2, \cdots, m+q\}$, and the functions $f, g_i, h_j : R^n \to R$ are assumed to be at least continuously differentiable. In particular when second-order conditions are discussed in this chapter, all functions in defining (NLP) are assumed to be twice continuously differentiable. Associated with (NLP), the Lagrange function $L : R^n \times R^{m+q} \to R$ is given by

$$L(x, \lambda) = f(x) + \sum_{i \in I} \lambda_i g_i(x) + \sum_{j \in J} \lambda_j h_j(x).$$

Throughout this chapter, let C be the feasible set of (NLP) and let $\bar{x} \in C$. The following index sets are useful in the sequel:

$$I(\bar{x}) := \{i \in I \mid g_i(\bar{x}) = 0\},$$
$$I(\bar{x}, w) := \{i \in I(\bar{x}) \mid \langle w, \nabla g_i(\bar{x})\rangle = 0\} \quad \forall w \in R^n,$$
$$I(\bar{x}, w, z) := \{i \in I(\bar{x}, w) \mid \langle z, \nabla g_i(\bar{x})\rangle + \langle w, \nabla^2 g_i(\bar{x}) w\rangle = 0\} \quad \forall w, z \in R^n.$$

The first-order linearized tangent cone to C at $\bar{x}$ is given by

$$L_C(\bar{x}) = \left\{ w \in R^n \;\middle|\; \begin{array}{ll} \langle w, \nabla g_i(\bar{x})\rangle \leq 0 & \forall i \in I(\bar{x}) \\ \langle w, \nabla h_j(\bar{x})\rangle = 0 & \forall j \in J \end{array} \right\},$$

while the second-order linearized tangent set to C at $\bar{x}$ in the direction $w \in L_C(\bar{x})$ is given by

$$L_C^2(\bar{x} \mid w) := \left\{ z \in R^n \;\middle|\; \begin{array}{ll} \langle \nabla g_i(\bar{x}), z\rangle + \langle w, \nabla^2 g_i(\bar{x}) w\rangle \leq 0 & \forall i \in I(\bar{x}, w) \\ \langle \nabla h_j(\bar{x}), z\rangle + \langle w, \nabla^2 h_j(\bar{x}) w\rangle = 0 & \forall j \in J \end{array} \right\}.$$

When $w \notin L_C(\bar{x})$, we interpret $L_C^2(\bar{x} \mid w)$ as an empty set. The critical cone of (NLP) at $\bar{x}$ is defined by

$$\mathcal{V}(\bar{x}) := \left\{ w \in R^n \;\middle|\; \begin{array}{ll} \langle w, \nabla f(\bar{x})\rangle \leq 0 & \\ \langle w, \nabla g_i(\bar{x})\rangle \leq 0 & \forall i \in I(\bar{x}) \\ \langle w, \nabla h_j(\bar{x})\rangle = 0 & \forall j \in J \end{array} \right\}.$$

The first-order necessary conditions for (NLP) that we shall study in this chapter are the well-known Karush-Kuhn-Tucker conditions (also known as KKT conditions). We say that the KKT condition holds at $\bar{x}$ if there exists a vector $\lambda \in R^{m+q}$ such that

$$\nabla_x L(\bar{x}, \lambda) = 0 \quad \lambda_i \geq 0,\; \lambda_i g_i(\bar{x}) = 0 \quad \forall i \in I.$$

We call such a vector λ a KKT multiplier, and denote by $\mathrm{KKT}(\bar{x})$ the set of all KKT multipliers of (NLP) at $\bar{x}$. By Farkas' lemma, $\mathrm{KKT}(\bar{x}) \neq \emptyset$ if and only if

$$-\nabla f(\bar{x}) \in L_C(\bar{x})^*.$$

As can be seen from Chapter 1, there are several classes of second-order necessary conditions for (NLP). In this chapter, we focus our attention on the ones depending the entire KKT multiplier set. Explicitly, we say that the second-order necessary condition holds at $\bar{x}$ if

$$\sup_{\lambda \in \mathrm{KKT}(\bar{x})} \langle w, \nabla^2_{xx} L(\bar{x}, \lambda) w\rangle \geq 0 \qquad \forall w \in \mathcal{V}(\bar{x}). \tag{4.1.1}$$

It should be noticed that if (4.1.1) holds, then necessarily $\mathrm{KKT}(\bar{x}) \neq \emptyset$. Note that (4.1.1) is exactly the same as the (SON2) described in Chapter 1. This class of second-order necessary conditions was initially studied by Ioffe [83], and has been extensively studied in [19, 20, 94, 138, 139, 140, 18, 31]. It is well-known that KKT conditions or second-order necessary conditions of the type (4.1.1) do not necessarily hold at local minima of (NLP) unless some conditions are satisfied.

One type of such conditions, often referred to as constraint qualifications (for short, CQs), relies on the constraints of (NLP) only, or in other words, is independent of the objective function of (NLP). We refer to the literature review presented in Chapter 1 for various CQs ensuring KKT conditions or second-order necessary conditions of the type (4.1.1). Among all possible CQs, Guignard constraint qualification (for short, GCQ) is the weakest one in the sense that it is both necessary and sufficient for KKT conditions to hold at local minima of (NLP), see [64]. The GCQ holds at $\bar{x}$ if by definition

$$T_C(\bar{x})^* = L_C(\bar{x})^*$$

or equivalently,

$$\mathrm{clconv} T_C(\bar{x}) = L_C(\bar{x}).$$

As for the CQs ensuring second-order necessary conditions of the type (4.1.1), it is worth mentioning the so-called second-order Guignard constraint [94] (for short, SGCQ), which is defined in spirit of the GCQ as follows:

$$\mathrm{clconv} T_C^2(\bar{x} \mid w) = L_C^2(\bar{x} \mid w) \qquad \forall w \in \mathcal{V}(\bar{x}).$$

Note that the GCQ can always be recovered from the SGCQ by taking $w = 0$.

Another type of conditions, which in contrast relies on both the constraints and the objective function of (NLP), is often expressed by virtue of exact penalty functions. See the survey paper by Burke [26] for a comprehensive investigation on the central roles that the classical l_1 exact penalty functions play in connection with optimality conditions. To be precise, we consider the l_p $(0 \leq p \leq 1)$ penalty function for (NLP) defined as follows:

$$\mathcal{F}_p(x) := f(x) + \mu S^p(x) \qquad \forall x \in R^n,$$

where $\mu \geq 0$ is the penalty parameter and the function S is defined by

$$S(x) := \sum_{i \in I} (g_i(x))_+ + \sum_{j \in J} |h_j(x)| \qquad \forall x \in R^n. \tag{4.1.2}$$

When $p = 0$, we interpret S^p as the indicator function δ_C, and thus have $\mathcal{F}_0 = f + \delta_C$. From this construction, it follows that $\bar{x}$ is a local minimum of (NLP) if and only if $\mathcal{F}_0$ has a local minimum at $\bar{x}$. When $p = 1$, $\mathcal{F}_p(x)$ reduces to the classical l_1 penalty function, which dates back to Eremin [48] and Zangwill [169], and has been investigated by many researchers, e.g., Pietrzykowski [128], Howe [76], Han and Mangasarian [66], Burke [25, 26], and Rockafellar [139]. When $0 < p < 1$, $\mathcal{F}_p(x)$ is often referred to as the lower order l_p penalty function, which was introduced by Luo et al. [105] for the study of mathematical programs with equilibrium constraints, and has been studied extensively in [80, 144, 145, 160, 161, 110, 162].

It is well known that both the KKT condition and the second-order necessary condition (4.1.1) hold at $\bar{x}$ if the l_1 penalty function is exact at $\bar{x}$, see in particular Han and Mangasarian ([66], Theorem 4.8) and Rockafellar ([139], Corollary 4.5). But for $0 < p < 1$, the KKT condition may not hold at $\bar{x}$ even if the l_p penalty function is exact at $\bar{x}$. This can be seen from the simple example: $\min -x$ s.t. $x^2 \leq 0$. However, Yang and Meng [161] showed that it is still possible to derive KKT conditions from lower order exact penalty functions by requiring that the constraint functions of (NLP) satisfy some additional conditions in terms of (generalized) second-order derivatives. Yang and Meng formulated these conditions by applying Farkas' Lemma and by estimating Dini upper-directional derivatives of $\mathcal{F}_p(x)$ using the tools of (generalized) Taylor expansions.

In what follows, let $\phi : R^n \to R_+ \cup \{+\infty\}$ be a lower semicontinuous function with the property that $\phi(x) = 0$ if and only if $x \in C$. To a great extent, ϕ plays the role as a general penalty term for (NLP). Note that the functions S^p with $0 \leq p \leq 1$ are particular instances of ϕ. We say that ϕ is a KKT-type penalty term if it has the ability of detecting KKT conditions in the sense as described in the following definition.

Definition 4.1.1 *We say that ϕ is of* KKT-type *at $\bar{x}$ if the* KKT *condition holds at $\bar{x}$ whenever there is a continuously differentiable function $f : R^n \to R$ such that the penalty function $f + \mu\phi$ is exact at $\bar{x}$ (i.e., for some finite penalty parameter $\mu \geq 0$, this penalty function admits an unconstrained local minimum at $\bar{x}$).*

According to the previous literature review, S is of KKT-type at $\bar{x}$, and S^0 is of KKT-type at $\bar{x}$ if and only if the GCQ holds at $\bar{x}$.

The outline of this chapter is as follows. In Section 4.2, we study conditions under which penalty terms are of KKT-type. These conditions allow us to derive KKT conditions from exactness of penalty functions. The main results that we rely are Theorems 3.2.1 and 3.3.1, and the variational description of regular subgradients (Rockafellar and Wets [141], Proposition 8.5). In subsection 4.2.1, we give equivalent conditions for penalty terms ϕ and S^p to be of KKT-type. These equivalent conditions are expressed by either subderivative kernels or regular subgradients of ϕ and S^p. In subsection 4.2.2, we present several conditions in terms of the original data of (NLP), which are sufficient for S^p to be of KKT-type. In particular when (NLP) has one inequality only, we give full characterizations in terms of the original data for S^p to be of KKT-type. We end Section 4.2 by giving a class of parameterized problems to illustrate that our result can be applied to derive KKT conditions when all existing methods fail.

In Section 4.3, by applying the second-order necessary conditions presented in Theorem 3.3.3 and the duality theorem of linear programming, we derive second-order necessary conditions of the type (4.1.1) for (NLP) from exactness of $f + \mu\phi$ under some additional conditions in terms of the kernel of the parabolic subderivative of ϕ. When $\phi = S^p$, we give sufficient conditions for these conditions by virtue of the original data of (NLP). We end this chapter by using an example to illustrate that even if neither the GCQ nor the SGCQ holds, our result obtained in this section can be applied to derive second-order necessary conditions of the type (4.1.1) for (NLP).

4.2 First-order Necessary Conditions via Exact Penalty Functions

In this section, we study conditions under which penalty terms are of KKT-type. These conditions allow us to derive KKT conditions from exactness of penalty functions. In subsection 4.2.1, we give equivalent conditions for penalty functions ϕ and S^p to be of KKT-type. These equivalent conditions are expressed by either subderivative kernels or regular subgradients of ϕ and S^p . In subsection 4.2.2, we present several conditions in terms of the original data of (NLP), which are sufficient for the equivalent conditions obtained in Section 4.2.1.

4.2.1 Equivalent Conditions for KKT-type Penalty Terms

In this subsection, we apply Theorems 3.2.1 and 3.3.1, and the variational description of regular subgradients (Rockafellar and Wets [141], Proposition 8.5) to establish some equivalent conditions for penalty terms to be of KKT-type.

We begin with a review of the variational description of regular subgradients.

Lemma 4.2.1 *(Rockafellar and Wets [141], Proposition 8.5). Consider a function $f : R^n \to \overline{R}$ and a point $\bar{x}$ with $f(\bar{x})$ finite. A vector v belongs to $\widehat{\partial} f(\bar{x})$ if and only if, on some neighborhood of $\bar{x}$, there is a function $h \leq f$ with $h(\bar{x}) = f(\bar{x})$ such that h is differentiable at $\bar{x}$ with $\nabla h(\bar{x}) = v$. Moreover h can be taken to be continuously differentiable with $h(x) < f(x)$ for all $x \neq \bar{x}$ near $\bar{x}$.*

Now we give equivalent conditions for ϕ to be of KKT-type at $\bar{x}$.

Theorem 4.2.1 *The following conditions are equivalent:*

(i) $[\ker d\phi(\bar{x})]^* \subset L_C(\bar{x})^*$.

(ii) $\mathrm{pos}(\widehat{\partial}\phi(\bar{x})) \subset L_C(\bar{x})^*$.

(iii) ϕ *is a* KKT-type *penalty term at* $\bar{x}$.

Proof. $[(i) \iff (ii)]$: Observing that $\bar{x}$ is a global minimum of ϕ, the equivalence follows immediately from Theorem 3.2.1.

$[(i) \Longrightarrow (iii)]$: Suppose that there is a continuously differentiable function $f : R^n \to R$ such that the penalty function

$$f(x) + \mu\phi(x)$$

has a local minimum at $\bar{x}$ for some finite penalty parameter $\mu \geq 0$. It follows from Theorem 3.3.1 that

$$-\nabla f(\bar{x}) \in [\ker d\phi(\bar{x})]^*.$$

By condition (i), we have

$$-\nabla f(\bar{x}) \in L_C(\bar{x})^*.$$

This implies by Farkas' Lemma that the KKT condition holds at $\bar{x}$. Thus, by Definition 4.1.1, ϕ is a KKT-type penalty term at $\bar{x}$.

[$(iii) \implies (ii)$]: Let $v \in \widehat{\partial}\phi(\bar{x})$. According to the variational description of regular subgradients in Lemma 4.2.1, there exist a neighborhood V of $\bar{x}$ and a continuously differentiable function $\psi : R^n \to R$ with $\psi(\bar{x}) = \phi(\bar{x}) = 0$ and $\nabla\psi(\bar{x}) = v$ such that

$$\psi(x) \leq \phi(x) \qquad \forall x \in V.$$

Set $f = -\psi$. It is clear to see that

$$f(x) + \phi(x) = -\psi(x) + \phi(x) \geq 0 = f(\bar{x}) + \phi(\bar{x}) \qquad \forall x \in V.$$

That is, the function $f + \phi$ has a local minimum at $\bar{x}$. Since ϕ is a KKT-type penalty term at $\bar{x}$, we have the KKT condition at $\bar{x}$. By Farkas' Lemma again, we have

$$-\nabla f(\bar{x}) \in L_C(\bar{x})^*.$$

Since $\nabla f(\bar{x}) = -\nabla\psi(\bar{x}) = -v$, we have

$$v \in L_C(\bar{x})^*.$$

Therefore, we have shown $\widehat{\partial}\phi(\bar{x}) \subset L_C(\bar{x})^*$ and hence

$$\mathrm{pos}(\widehat{\partial}\phi(\bar{x})) \subset L_C(\bar{x})^*.$$

This completes the proof. $\square$

In what follows, we establish equivalent conditions for S^p with $0 \leq p \leq 1$ to be a KKT-type penalty term at $\bar{x}$. We begin with the calculation of the subderivative (kernel) and the subdifferential of S at $\bar{x}$.

Lemma 4.2.2 *Let S be given by (4.1.2). Then*

$$dS(\bar{x})(w) = \sum_{i \in I(\bar{x})} (\langle \nabla g_i(\bar{x}), w \rangle)_+ + \sum_{j \in J} |\langle \nabla h_j(\bar{x}), w \rangle| \qquad \forall w \in R^n, \tag{4.2.3}$$

and

$$\begin{aligned} &\widehat{\partial}S(\bar{x}) = \partial S(\bar{x}) \\ &= \left\{ \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) + \sum_{j \in J} \lambda_j \nabla h_j(\bar{x}) \mid 0 \leq \lambda_i \leq 1 \; \forall i \in I(\bar{x}), -1 \leq \lambda_j \leq 1 \; \forall j \in J \right\}. \end{aligned} \tag{4.2.4}$$

In particular, one has

$$\ker dS(\bar{x}) = L_C(\bar{x}) \tag{4.2.5}$$

and

$$\begin{aligned} &[\ker dS(\bar{x})]^* = [L_C(\bar{x})]^* = \mathrm{pos}(\widehat{\partial} S(\bar{x})) = \mathrm{pos}(\partial S(\bar{x})) \\ = &\left\{ \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) + \sum_{j \in J} \lambda_j \nabla h_j(\bar{x}) \mid \lambda_i \geq 0 \quad \forall i \in I(\bar{x}),\ \lambda_j \in R \quad \forall j \in J \right\}. \end{aligned} \tag{4.2.6}$$

Proof. Applying the basic chain rule and the sum rule given respectively in Theorem 10.6 and Corollary 10.9 of [141], we can easily get (4.2.3) and (4.2.4). Once these formulas are obtained, (4.2.5) and (4.2.6) follow readily. This completes the proof. □

Theorem 4.2.2 *Let S be given by (4.1.2). Then S is a* KKT-type *penalty term at $\bar{x}$, and for any $0 \leq p < 1$, the following conditions are equivalent:*

(i) $[\ker dS^p(\bar{x})]^* = L_C(\bar{x})^*$.

(ii) $\widehat{\partial} S^p(\bar{x}) = L_C(\bar{x})^*$.

(iii) S^p *is a* KKT-type *penalty term at* $\bar{x}$.

Proof. Let $0 \leq p < 1$. It follows from (4.2.5) and Proposition 3.2.1 (*iii*) that

$$\ker dS^p(\bar{x}) \subset \mathrm{dom}\, dS^p(\bar{x}) \subset \ker dS(\bar{x}) = L_C(\bar{x}), \tag{4.2.7}$$

which implies that

$$L_C(\bar{x})^* \subset [\ker dS^p(\bar{x})]^*.$$

In view of Lemma 4.2.2 and Theorem 4.2.1, it suffices to show that the regular subdifferential $\widehat{\partial} S^p(\bar{x})$ of S^p at $\bar{x}$ is a cone when condition (*i*) is satisfied. In fact, it follows from condition (*i*) and (4.2.7) that

$$[\ker dS^p(\bar{x})]^* = [\mathrm{dom}\, dS^p(\bar{x})]^*.$$

Observing that the function S^p has a global minimum at $\bar{x}$, we get from Theorem 3.2.1 that $\widehat{\partial} S^p(\bar{x})$ is a cone. This completes the proof. □

Remark 4.2.1 *Theorem 4.2.2 includes two existing results as special cases:*

(i) *The* GCQ *is the weakest* CQs *in the sense that it is both necessary and sufficient for local minima of* (NLP) *to possess* KKT *conditions. To see that, we observe that every local minimum of* (NLP) *is a local minimum of the* l_0 *penalty function* $\mathcal{F}_0$*, and that condition (i) in Theorem 4.2.2 holds with* $p = 0$ *if and only if the* GCQ *holds.*

(ii) *The* KKT *condition holds at* $\bar{x}$ *if the* l_1 *penalty function is exact at* $\bar{x}$*. This is because* S *is a* KKT-type *penalty function at* $\bar{x}$ *without any other condition.*

It should be noticed that a similar condition to condition (i) of Theorem 4.2.2 has been given by Meng and Yang ([110], Theorem 2.5). This condition is expressed by the notion of contingent derivatives of set-valued mappings, see [8] for more details on various derivatives for set-valued mappings. The contingent derivative of a set-valued mapping $M : R^n \rightrightarrows R^m$ at $(x, y) \in \text{gph}M$ is defined by the set-valued map $DM(x, y) : R^n \rightrightarrows R^s$ such that

$$\text{gph}(DM(x, y)) = T_{\text{gph}M}(x, y).$$

In particular, when M is single-valued at x, i.e., $M(x) = \{y\}$, we use $DM(x)$ to denote $DM(x, y)$ for simplicity, and define the kernel of $DM(x)$ by

$$KerDM(x) = \{u \in R^n \mid 0 \in DM(x)(u)\}.$$

Let $0 \leq p \leq 1$. The kernel of the contingent derivative of S^p at $(\bar{x}, 0)$ is then given by

$$\text{ker}DS^p(\bar{x}) := \{u \in R^n \mid 0 \in DS^p(\bar{x})(u)\}.$$

By virtue of this kernel set, we obtain a sufficient condition for S^p to be a KKT-type penalty term as follows, which is a reformulation of Theorem 2.5 of Meng and Yang [110].

Theorem 4.2.3 *([110], Theorem 2.5). Let* $0 \leq p \leq 1$ *and let* S *be given by (4.1.2). If*

$$[\text{ker}DS^p(\bar{x})]^* = L_C(\bar{x})^*, \tag{4.2.8}$$

then S^p *is a* KKT-type *penalty term at* $\bar{x}$*.*

Proof. Let f be a continuously differentiable function such that the penalty function $\mathcal{F}_p(x)$ has a local minimum at $\bar{x}$ with a finite penalty parameter $\mu \geq 0$. By the definition of contingent derivative, we thus have

$$D\mathcal{F}_p(\bar{x})(w) \subset R_+ \quad \forall w \in R^n.$$

Noting that f is assumed to be continuously differentiable, it follows easily from the sum rule of contingent derivative (see [8] and [99]) that

$$D\mathcal{F}_p(\bar{x})(w) = \nabla f(\bar{x})^T w + \mu DS^p(\bar{x})(w).$$

Therefore, we have

$$-\nabla f(\bar{x})^T w \leq 0 \quad \forall w \in \ker DS^p(\bar{x}),$$

that is, $-\nabla f(\bar{x}) \in \ker DS^p(\bar{x})^*$. In view of (4.2.8), we have $-\nabla f(\bar{x}) \in L_C(\bar{x})^*$, which by Farkas' lemma amounts to the KKT condition at $\bar{x}$. This completes the proof. □

Remark 4.2.2 *By definition, it is straightforward to verify that*

$$\ker dS^p(\bar{x}) = \ker DS^p(\bar{x}).$$

In view of this equality and Theorem 4.2.2, we confirm that condition (4.2.8) is not only sufficient but also necessary for S^p to be a KKT-type *penalty term. Incidentally, we are not aware of the necessity at the time when our paper [110] got published. We emphasize that this equivalence cannot be easily obtained without the help of Theorem 3.2.1 and the variational description of regular subgradients (Rockafellar and Wets [141], Proposition 8.5), both of which are closely related with the notion of subderivative instead of contingent derivative. This is why we do not use contingent derivatives in this thesis anymore as we have done in the paper [110]. The same reason applies in next chapter.*

4.2.2 Sufficient Conditions by Original Data

In this subsection, in terms of the original data of (NLP), we give several conditions which are sufficient for the equality

$$\ker dS^p(\bar{x}) = L_C(\bar{x}). \tag{4.2.9}$$

According to Theorem 4.2.2, if (4.2.9) is satisfied, then S^p is clearly a KKT-type penalty term at $\bar{x}$.

To begin with, we consider the implications of condition (4.2.9) when p varies. In view of Proposition 3.2.1 and (4.2.5), we have

Lemma 4.2.3 *Concerning condition (4.2.9) with different values of p, we have*

(i**)** $\ker dS^1(\bar{x}) = L_C(\bar{x})$.

(ii**)** $\ker dS^0(\bar{x}) = L_C(\bar{x})$ *if and only if* $T_C(\bar{x}) = L_C(\bar{x})$ *(i.e., the* ACQ *holds at* $\bar{x}$*).*

(iii**)** *If* $\ker dS^p(\bar{x}) = L_C(\bar{x})$ *for some* $0 \le p \le 1$*, then*

$$\ker dS^{p'}(\bar{x}) = L_C(\bar{x}) \quad \forall p' \in [p, 1].$$

Proposition 4.2.1 *For* $0 < p \le 1$*, (4.2.9) holds if and only if, for every* $u \in L_C(x)$*, there exist* $t_k \to 0+$ *and* $u_k \to u$ *such that*

$$\max\{\frac{g_i(\bar{x} + t_k u_k)}{t_k^{1/p}}, 0\} \to 0 \quad \forall i \in I(\bar{x}, u), \tag{4.2.10}$$

and

$$\frac{h_j(\bar{x} + t_k u_k)}{t_k^{1/p}} \to 0 \quad \forall j \in J. \tag{4.2.11}$$

Proof. Let $0 < p \le 1$. In view of Lemma 4.2.3, the relation $\ker dS^p(\bar{x}) \subset L_C(\bar{x})$ holds automatically. It follows from Proposition 3.2.1 (i) that, $u \in \ker dS^p(\bar{x})$ if and only if there exist $t_k \to 0+$ and $u_k \to u$ such that

$$\frac{S^p(\bar{x} + t_k u_k)}{t_k} \to 0. \tag{4.2.12}$$

By the definition of S^p, (4.2.12) can be reformulated as follows:

$$\max\{\frac{g_i(\bar{x} + t_k u_k)}{t_k^{1/p}}, 0\} \to 0 \quad \forall i \in I, \tag{4.2.13}$$

and

$$|\frac{h_j(\bar{x} + t_k u_k)}{t_k^{1/p}}| \to 0 \quad \forall j \in J. \tag{4.2.14}$$

Clearly, (4.2.11) is equivalent to (4.2.14). To show the equivalence of (4.2.10) and (4.2.13), it suffices to show that, for any $u \in L_C(\bar{x})$ with any sequences $t_k \to 0+$ and $u_k \to u$,

$$\max\{\frac{g_i(\bar{x}+t_k u_k)}{t_k^{1/p}}, 0\} \to 0 \qquad \forall i \notin I(\bar{x}, u). \tag{4.2.15}$$

If $i \notin I(\bar{x})$, (4.2.15) follows easily from the continuity of g_i at $\bar{x}$. If $i \in I(\bar{x})$ but $i \notin I(\bar{x}, u)$, we then have $\nabla g_i(\bar{x})^T u < 0$ since $u \in L_C(\bar{x})$. Observing that g_i is assumed to be continuously differentiable, we have by the first-order Taylor expansion

$$g_i(\bar{x}+t_k u_k) = t_k \nabla g_i(\bar{x})^T u_k + o(t_k),$$

and by noticing that $\nabla g_i(\bar{x})^T u < 0$ and $1/p - 1 > 0$, we further have

$$\liminf_{k\to+\infty} \frac{g_i(\bar{x}+t_k u_k)}{t_k^{1/p}} = \liminf_{k\to+\infty} \frac{\nabla g_i(\bar{x})^T u_k + \frac{o(t_k)}{t_k}}{t_k^{1/p-1}} = -\infty.$$

Therefore, we have shown (4.2.15). This completes the proof. $\square$

The following generalized lower and upper second-order directional derivatives for a continuously differentiable function can be found in [38] and [159].

Definition 4.2.1 *Let $f : R^n \to R$ be a continuously differentiable function. The generalized lower and upper second-order directional derivatives for f at $x \in R^n$ in the direction $u \in R^n$ are defined, respectively, by*

$$f_{oo}(x; u) = \liminf_{y\to x,\, t\to 0+} \frac{\nabla f(y+tu)^T u - \nabla f(y)^T u}{t},$$

and

$$f^{oo}(x; u) = \limsup_{y\to x,\, t\to 0+} \frac{\nabla f(y+tu)^T u - \nabla f(y)^T u}{t}.$$

The following results on the generalized lower and upper second-order directional derivatives can be found in [38].

Lemma 4.2.4 *Let $f : R^n \to R$ be a continuously differentiable function. The following statements are true:*

(a) $f_{oo}(x; u)$ and $f^{oo}(x; u)$ are positively homogeneous of degree 2 in the second argument u, i.e., for $t \geq 0$,

$$f_{oo}(x; tu) = t^2 f_{oo}(x; u), \quad f^{oo}(x; tu) = t^2 f^{oo}(x; u);$$

(b) *The function* $(x,u) \to f_{oo}(x;u)$ *is lower semicontinuous and the function* $(x,u) \to f^{oo}(x;u)$ *is upper semicontinuous;*

(c) *Let* $x, y \in R^n$*. The following generalized Taylor expansions hold* ($\beta_1, \beta_2 \in (0,1)$)*:*

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{1}{2} f_{oo}(x + \beta_1(y-x); y-x), \tag{4.2.16}$$

$$f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{1}{2} f^{oo}(x + \beta_2(y-x); y-x). \tag{4.2.17}$$

The following proposition is originally due to Yang and Meng [161]. For the sake of completeness, we give its detailed proof from a slightly different perspective.

Proposition 4.2.2 *Each of the following conditions is sufficient for (4.2.9) by fulfilling (4.2.10) and (4.2.11) for each* $u \in L_C(\bar{x})$ *with an arbitrary sequence* $t_k \to 0+$ *and a constant sequence* $u_k \equiv u$*.*

(*i*) $\frac{1}{2} < p < 1$ *and the functions* g_i*,* $i \in I(\bar{x})$ *and* h_j*,* $j \in J$ *are* $\mathcal{C}^{1,1}$*.*

(*ii*) $p = \frac{1}{2}$*, and for every* $u \in L_C(\bar{x})$*, it follows that*

$$g_i^{oo}(\bar{x};u) \leq 0 \quad \forall i \in I(\bar{x},u), \quad h_j^{oo}(\bar{x};u) = h_{j_{oo}}(\bar{x};u) = 0 \quad \forall j \in J.$$

(*iii*) $p = \frac{1}{2}$*, the functions* g_i*,* $i \in I(\bar{x})$ *and* h_j*,* $j \in J$ *are assumed to be twice continuously differentiable, and for every* $u \in L_C(\bar{x})$*, it follows that*

$$\langle u, \nabla^2 g_i(\bar{x}) u \rangle \leq 0 \quad \forall i \in I(\bar{x},u), \quad \langle u, \nabla^2 h_j(\bar{x}) u \rangle = 0 \quad \forall j \in J. \tag{4.2.18}$$

(*iv*) $p = 0$*,* $q = 0$ *(i.e., there is no equality constraint), and for every* $u \in L_C(\bar{x}) \backslash \{0\}$*, it follows that*

$$g_i^{oo}(\bar{x};u) < 0 \quad \forall i \in I(\bar{x},u).$$

Proof. Fix arbitrarily $u \in L_C(\bar{x})$, $i \in I(\bar{x},u)$, and $j \in J$. Let $\{t_k\}$ be an arbitrary sequence such that $t_k \to 0+$. To show conditions (*i*) to (*iii*), it suffices to show (according to Proposition 4.2.1)

$$\max\{\frac{g_i(\bar{x}+t_k u)}{t_k^{1/p}}, 0\} \to 0, \tag{4.2.19}$$

and

$$\frac{h_j(\bar{x}+t_k u)}{t_k^{1/p}} \to 0. \tag{4.2.20}$$

[condition (i)]: It follows from the generalized Taylor expansion (4.2.17) that,

$$\begin{aligned}\frac{g_i(\bar{x}+t_k u)}{t_k^{1/p}} &\le \frac{g_i(\bar{x})+t_k\nabla g_i(\bar{x})^T u+\frac{1}{2}g_i^{oo}(\bar{x}+\beta_2 t_k u; t_k u)}{t_k^{1/p}}\\ &= \frac{\frac{1}{2}g_i^{oo}(\bar{x}+\beta_2 t_k u; t_k u)}{t_k^{1/p}}\\ &= \tfrac{1}{2}g_i^{oo}(\bar{x}+\beta_2 t_k u; u)t_k^{2-1/p},\end{aligned} \tag{4.2.21}$$

where $\beta_2 \in (0,1)$, and the second equality follows from the fact that $g_i^{oo}(x;u)$ is positively homogeneous of degree 2 in the second argument u. Since g_i is assumed to be $\mathcal{C}^{1,1}$, $g_i^{oo}(\bar{x};u)$ is clearly finite. Thus, by the upper semicontinuity of $(x,u) \to g_i^{oo}(x;u)$ and by noticing that $2-1/p>0$ because $\frac{1}{2}<p<1$, we have $\frac{1}{2}g_i^{oo}(\bar{x}+\beta_2 t_k u; u)t_k^{2-1/p} \to 0$. This together with (4.2.21) implies (4.2.19). Along the same line of deriving (4.2.21) and by applying the generalized Taylor expansions (4.2.16) and (4.2.17), we have

$$\frac{1}{2}h_{j_{oo}}(\bar{x}+\beta_1 t_k u; u)t_k^{2-1/p} \le \frac{h_j(\bar{x}+t_k u)}{t_k^{1/p}} \le \frac{1}{2}h_j^{oo}(\bar{x}+\beta_2 t_k u; u)t_k^{2-1/p}, \tag{4.2.22}$$

where $\beta_1, \beta_2 \in (0,1)$. Since h_j is assumed to be $\mathcal{C}^{1,1}$, both $h_{j_{oo}}(\bar{x};u)$ and $h_j^{oo}(\bar{x};u)$ are finite. Thus, by the lower semicontinuity of $(x,u) \to h_{j_{oo}}(x;u)$ and the upper semicontinuity of $(x,u) \to h_j^{oo}(x;u)$ and by noticing again that $2-1/p>0$, (4.2.20) can be easily obtained from (4.2.22).

[condition (ii)]: Along the same line of deriving (4.2.21), we have

$$\frac{g_i(\bar{x}+t_k u)}{t_k^2} \le \frac{1}{2}g_i^{oo}(\bar{x}+\beta_2 t_k u; u) \tag{4.2.23}$$

where $\beta_2 \in (0,1)$. Since $g_i^{oo}(\bar{x};u) \le 0$, (4.2.19) with $p=\frac{1}{2}$ follows readily from (4.2.23) and the upper semicontinuity of $(x,u) \to g_i^{oo}(x;u)$. Along the same line of deriving (4.2.22), we have

$$\frac{1}{2}h_{j_{oo}}(\bar{x}+\beta_1 t_k u; u) \le \frac{h_j(\bar{x}+t_k u)}{t_k^2} \le \frac{1}{2}h_j^{oo}(\bar{x}+\beta_2 t_k u; u), \tag{4.2.24}$$

where $\beta_1, \beta_2 \in (0,1)$. Since $h_j^{oo}(\bar{x};u) = h_{j_{oo}}(\bar{x};u) = 0$, (4.2.20) with $p=\frac{1}{2}$ follows easily from (4.2.24), the lower semicontinuity of $(x,u) \to h_{j_{oo}}(x;u)$ and the upper semicontinuity of $(x,u) \to h_j^{oo}(x;u)$.

[condition (iii)]: It follows from the second-order Taylor expansion for twice continuously differentiable functions that,

$$\begin{aligned}\frac{g_i(\bar{x}+t_k u)}{t_k^2} &= \frac{g_i(\bar{x})+t_k\nabla g_i(\bar{x})^T u+\frac{1}{2}t_k^2\langle u,\nabla^2 g_i(\bar{x})u\rangle+o(t_k^2)}{t_k^2}\\ &= \tfrac{1}{2}\langle u,\nabla^2 g_i(\bar{x})u\rangle+\tfrac{o(t_k^2)}{t_k^2}\to\tfrac{1}{2}\langle u,\nabla^2 g_i(\bar{x})u\rangle,\end{aligned}$$

which implies (4.2.19). Similarly, we can obtain (4.2.20).

[condition (iv)]: By Lemma 4.2.3, it suffices to show $T_C(\bar{x}) = L_C(\bar{x})$ or equivalently $L_C(\bar{x}) \subset T_C(\bar{x})$. Let $w \in L_C(\bar{x})\backslash\{0\}$ and let $i \in I(\bar{x})$. It follows from the generalized Taylor expansion (4.2.17) that for any $t_k \to 0+$,

$$g_i(\bar{x}+t_k w) \le t_k\nabla g_i(\bar{x})^T w+\frac{1}{2}g_i^{oo}(\bar{x}+\beta_2 t_k w;t_k w) \tag{4.2.25}$$

If $\nabla g_i(\bar{x})^T w < 0$, then (4.2.25) implies that $g_i(\bar{x}+t_k w) \le 0$ for sufficiently large k. If $\nabla g_i(\bar{x})^T w = 0$, condition (iv) implies that $g_i^{oo}(x;w) < 0$. Due to the upper semicontinuity of $(x,w) \to g_i^{oo}(x;w)$, we get from (4.2.25) that $g_i(\bar{x}+t_k w) \le 0$ for sufficiently large k. Therefore, we actually have $g_i(\bar{x}+t_k w) \le 0$ for sufficiently large k and for all $i \in I$. This shows that $w \in T_C(\bar{x})$. Thus, we have shown $L_C(\bar{x}) \subset T_C(\bar{x})$. This completes the proof. □

The following proposition gives another sufficient condition for $\mathrm{ker} dS^{\frac{1}{2}}(\bar{x}) = L_C(\bar{x})$, which is strictly weaker than (4.2.18) as can be seen soon.

Proposition 4.2.3 *Assume that g_i, $i \in I$ and h_j, $j \in J$ are twice continuously differentiable functions. Set*

$$\mathrm{KKT}_0(\bar{x}) := \left\{ \lambda \in R^{m+q} \;\middle|\; \begin{array}{l} \sum_{i\in I}\lambda_i\nabla g_i(\bar{x})+\sum_{j\in J}\lambda_j\nabla h_j(\bar{x})=0 \\ \lambda_i \ge 0 \quad \forall i \in I(\bar{x}), \quad \lambda_i = 0 \quad \forall i \in I\backslash I(\bar{x}) \end{array} \right\}.$$

If for every $u \in L_C(\bar{x})$, it follows that

$$\max_{\lambda\in\mathrm{KKT}_0(\bar{x})}\left\{\sum_{i\in I}\lambda_i\langle u,\nabla^2 g_i(\bar{x})u\rangle+\sum_{j\in J}\lambda_j\langle u,\nabla^2 h_j(\bar{x})u\rangle\right\}=0, \tag{4.2.26}$$

then $\mathrm{ker} dS^{\frac{1}{2}}(\bar{x}) = L_C(\bar{x})$.

Proof. Let $u \in L_C(\bar{x})$. It is easy to verify that

$$\mathrm{KKT}_0(\bar{x}) = \left\{ \lambda \in R^{m+q} \;\middle|\; \begin{array}{l} \sum_{i\in I}\lambda_i\nabla g_i(\bar{x})+\sum_{j\in J}\lambda_j\nabla h_j(\bar{x})=0 \\ \lambda_i \ge 0 \quad \forall i \in I(\bar{x},u), \quad \lambda_i = 0 \quad \forall i \in I\backslash I(\bar{x},u) \end{array} \right\}. \tag{4.2.27}$$

Thus, by a nonhomogeneous Farkas' Lemma (see p.32 of [107] or Lemma 4.2 of [149]), (4.2.26) holds if and only if there exists $w \in R^n$ such that

$$\begin{aligned} \nabla g_i(\bar{x})^T w + \langle u, \nabla^2 g_i(\bar{x}) u\rangle \le 0 \qquad & \forall i \in I(\bar{x}, u), \\ \nabla h_j(\bar{x})^T w + \langle u, \nabla^2 h_j(\bar{x}) u\rangle = 0 \qquad & \forall j \in J. \end{aligned} \tag{4.2.28}$$

For any $t_k \to 0+$ and $w_k \to w$, it is not hard to verify that

$$\max\{\frac{g_i(\bar{x} + t_k u + \frac{1}{2}t_k^2 w_k)}{t_k^2}, 0\} \to 0 \qquad \forall i \in I(\bar{x}, u), \tag{4.2.29}$$

and

$$\frac{h_j(\bar{x} + t_k u + \frac{1}{2}t_k^2 w_k)}{t_k^2} \to 0 \qquad \forall j \in J. \tag{4.2.30}$$

In fact, by the second order Taylor expansion, we have for each $i \in I(\bar{x}, u)$

$$g_i(\bar{x} + t_k u + \frac{1}{2}t_k^2 w_k) = \frac{1}{2}t_k^2 \left[\nabla g_i(\bar{x})^T w_k + \langle u + \frac{1}{2}t_k w_k, \nabla^2 g_i(\bar{x})(u + \frac{1}{2}t_k w_k)\rangle\right] + o(t_k^2),$$

which, together with (4.2.28), implies (4.2.29). (4.2.30) can be verified in the same way. It follows from (4.2.29) and (4.2.30) that the sequences $\{t_k\}$ and $\{u_k := u + \frac{1}{2}t_k w_k\}$ fulfill (4.2.10) and (4.2.11). By Proposition 4.2.1, we thus have $\ker dS^{\frac{1}{2}}(\bar{x}) = L_C(\bar{x})$. This completes the proof. □

It is well-known that the MFCQ holds at $\bar{x}$ if and only if $\mathrm{KKT}_0(\bar{x}) = \{0\}$. Thus, (4.2.26) holds automatically when the MFCQ holds at $\bar{x}$. However, even if the LICQ holds at $\bar{x}$, the condition (4.2.18) may not hold. Take the feasible set of the form

$$C = \{x \in R^2 \mid x_2^2 - x_1 \le 0\}$$

and $\bar{x} = 0$ for example. Therefore, condition (4.2.26) is in general strictly weaker than condition (4.2.18).

Below is another example which illustrates that condition (4.2.26) can be strictly weaker than condition (4.2.18) when the GCQ does not hold at $\bar{x}$.

Example 4.2.1 *In* (NLP), *let* $n = m = 2$, $q = 0$, $g_1(x) = x_1^2 x_2$, $g_2(x) = x_2^2 - x_1$, *and let* $\bar{x} = 0$. *Condition (4.2.18) is not satisfied, because for any* $u = (0, u_2)^T$ *with* $u_2 \neq 0$, *we have* $u \in L_C(\bar{x}) = R_+ \times R$ *and* $I(\bar{x}, u) = \{1, 2\}$, *but*

$$\langle u, \nabla^2 g_2(\bar{x}) u\rangle = 2u_2^2 > 0.$$

However, we can show that condition (4.2.26) is satisfied. By definition, we have

$$\mathrm{KKT}_0(\bar{x}) = \{\lambda \in R^2_+ \mid \lambda_1 \nabla g_1(\bar{x}) + \lambda_2 \nabla g_2(\bar{x}) = 0\} = R_+ \times \{0\}.$$

Then, for each $u \in L_C(\bar{x})$, we have

$$\max_{\lambda \in \mathrm{KKT}_0(\bar{x})} \left\{\lambda_1 \langle u, \nabla^2 g_1(\bar{x}) u\rangle + \lambda_2 \langle u, \nabla^2 g_2(\bar{x}) u\rangle\right\} = 0.$$

Therefore, condition (4.2.26) is satisfied and hence $\mathrm{ker}dS^{\frac{1}{2}}(\bar{x}) = L_C(\bar{x})$. *In fact, by Lemma 4.2.3 and the criterion in Proposition 4.2.1, we can calculate* $\mathrm{ker}dS^p(\bar{x})$ *for all* $p \in [0,1]$, *which gives that*

$$\mathrm{ker}dS^p(\bar{x}) = \begin{cases} R_+ \times (-R_+) & \text{if } 0 \le p \le \dfrac{1}{5}, \\ R_+ \times (-R_+) \cup \{0\} \times R_+ & \text{if } \dfrac{1}{5} < p \le \dfrac{1}{3}, \\ R_+ \times R & \text{if } \dfrac{1}{3} < p \le 1. \end{cases}$$

Thus, we have

$$\mathrm{ker}dS^p(\bar{x}) = L_C(\bar{x}) \qquad \forall p \in (\frac{1}{3}, 1],$$

$$[\mathrm{ker}dS^p(\bar{x})]^* = L_C(\bar{x})^* \qquad \forall p \in (\frac{1}{5}, \frac{1}{3}],$$

and

$$[\mathrm{ker}dS^p(\bar{x})]^* \ne L_C(\bar{x})^* \qquad \forall p \in [0, \frac{1}{5}].$$

In particular, the GCQ *does not hold at $\bar{x}$, and the function S^p with $p \in (\frac{1}{5}, 1]$ cannot be a local error bound for C at $\bar{x}$.*

When the feasible set of (NLP) is defined by one single inequality, the question as to whether S^p with any $0 \le p \le 1$ is a KKT-type penalty term is quite clear as shown by the following proposition.

Proposition 4.2.4 *Let $g : R^n \to R$ be a twice continuously differentiable function. Assume that* (NLP) *has only one inequality constraint $g(x) \le 0$. In this case, we have $S(x) = \max\{g(x), 0\}$ and $C = \{x \in R^n \mid g(x) \le 0\}$. Let $\bar{x} \in R^n$ be such that $g(\bar{x}) \le 0$. The following statements are true:*

(a) If $g(\bar{x}) < 0$ or $g(\bar{x}) = 0$ with $\nabla g(\bar{x}) \ne 0$, then

$$\mathrm{ker}dS^p(\bar{x}) = L_C(\bar{x}) \qquad \forall p \in [0, 1].$$

(*b*) *If* $g(\bar{x}) = 0$, $\nabla g(\bar{x}) = 0$ *and* $\nabla^2 g(\bar{x})$ *is negative semi-definite with* $\nabla^2 g(\bar{x}) \neq 0$, *then*

$$\mathrm{ker} dS^p(\bar{x}) = L_C(\bar{x}) \qquad \forall p \in [0,1].$$

(*c*) *If* $g(\bar{x}) = 0$, $\nabla g(\bar{x}) = 0$ *and* $\nabla^2 g(\bar{x})$ *is positive semi-definite with* $\nabla^2 g(\bar{x}) \neq 0$, *then*

$$[\mathrm{ker} dS^p(\bar{x})]^* \neq L_C(\bar{x})^* \qquad \forall p \in [0, \frac{1}{2}], \tag{4.2.31}$$

and

$$\mathrm{ker} dS^p(\bar{x}) = L_C(\bar{x}) \qquad \forall p \in (\frac{1}{2}, 1]. \tag{4.2.32}$$

(*d*) *If* $g(\bar{x}) = 0$, $\nabla g(\bar{x}) = 0$ *and* $\nabla^2 g(\bar{x}) = 0$, *then*

$$\mathrm{ker} dS^p(\bar{x}) = L_C(\bar{x}) \qquad \forall p \in [\frac{1}{2}, 1].$$

(*e*) *If* $g(\bar{x}) = 0$, $\nabla g(\bar{x}) = 0$ *and* $\nabla^2 g(\bar{x})$ *is indefinite, then*

$$[\mathrm{ker} dS^p(\bar{x})]^* = L_C(\bar{x})^* \qquad \forall p \in [0, \frac{1}{2}], \tag{4.2.33}$$

and

$$\mathrm{ker} dS^p(\bar{x}) = L_C(\bar{x}) \qquad \forall p \in (\frac{1}{2}, 1]. \tag{4.2.34}$$

Proof. Statement (*a*) holds trivially. In what follows, we assume that $\bar{x} \in C$ is such that $g(\bar{x}) = 0$ and $\nabla g(\bar{x}) = 0$. By applying the spectral decomposition theorem for real symmetric matrices, we can find an orthogonal matrix P such that

$$\nabla^2 g(\bar{x}) = P \Lambda P^T, \tag{4.2.35}$$

where Λ is a diagonal matrix with diagonal entries being the eigenvalues of $\nabla^2 g(\bar{x})$. Moreover, from a direct calculation, we have $L_C(\bar{x}) = R^n$, $L_C(\bar{x})^* = \{0\}$ and

$$O \subset \mathrm{ker} dS^0(\bar{x}) \subset \mathrm{ker} dS^{\frac{1}{2}}(\bar{x}) = Q, \tag{4.2.36}$$

where $O := \{w \in R^n \mid \langle w, \nabla^2 g(\bar{x}) w \rangle < 0\}$ and $Q := \{w \in R^n \mid \langle w, \nabla^2 g(\bar{x}) w \rangle \leq 0\}$. In fact, from the second-order Taylor expansion, we have for any $t_k \to 0+$ and any $w_k \to w$,

$$g(\bar{x} + t_k w_k) = t_k^2 [\frac{1}{2} \langle w_k, \nabla^2 g(\bar{x}) w_k \rangle + \frac{o(t_k^2)}{t_k^2}],$$

by virtue of which, we can easily show (4.2.36).

$[(b)]$: Since $\nabla^2 g(\bar{x})$ is negative semi-definite with $\nabla^2 g(\bar{x}) \neq 0$, we can find $\tilde{w} \in O$, i.e., $\langle \tilde{w}, \nabla^2 g(\bar{x})\tilde{w}\rangle < 0$. Let $w \in Q \backslash O$. We have $\langle w, \nabla g(\bar{x}) w\rangle = 0$. Since $\nabla^2 g(\bar{x})$ is negative semi-definite, we have

$$\langle w + \tau\tilde{w}, \nabla^2 g(\bar{x})(w + \tau\tilde{w})\rangle < 0 \qquad \forall \tau > 0.$$

Therefore, $w + \tau\tilde{w} \in O$ for $\tau > 0$. This implies that $w \in \mathrm{cl}O$ and hence $\mathrm{cl}O = Q$. Since $L_C(\bar{x}) = R^n$ and $Q = R^n$, we get from (4.2.36) that $\ker dS^0(\bar{x}) = L_C(\bar{x})$. In view of Lemma 4.2.2, statement (b) follows readily.

$[(c)]$: It follows from Proposition 4.2.2 (i) that (4.2.32) holds. Let $\tilde{Q} = \{y \in R^n \mid y^T \Lambda y \leq 0\}$. In view of (4.2.35), we have $Q = P\tilde{Q}$ and $Q^* = P\tilde{Q}^*$. Since $\nabla^2 g(\bar{x})$ is positive semi-definite with $\nabla^2 g(\bar{x}) \neq 0$, we can assume without loss of generality that, there exists a positive integer n_1 such that $\Lambda(i,i) > 0$ with $1 \leq i \leq n_1$, and $\Lambda(i,i) = 0$ with $n_1 + 1 \leq i \leq n$. Then, we have $\tilde{Q} = \{0_{R^{n_1}}\} \times R^{n-n_1}$ and hence $\tilde{Q}^* = R^{n_1} \times \{0_{R^{n-n_1}}\}$. Since $Q^* = P\tilde{Q}^*$ and P is an orthogonal matrix, we have $Q^* \neq \{0\}$. Since $L_C(\bar{x})^* = \{0\}$, it follows from (4.2.36) that

$$[\ker dS^{\frac{1}{2}}(\bar{x})]^* \neq L_C(\bar{x})^*.$$

In view of Proposition 3.2.1 (iii), (4.2.31) holds.

$[(d)]$: The result follows directly from (4.2.36) and Lemma 4.2.2.

$[(e)]$: It follows from Proposition 4.2.2 (i) that (4.2.34) holds. Let $\tilde{O} = \{y \in R^n \mid y^T \Lambda y < 0\}$. In view of (4.2.35), we have $O = P\tilde{O}$ and $O^* = P\tilde{O}^*$. Since $\nabla^2 g(\bar{x})$ is indefinite, we can assume without loss of generality that, there exist positive integers n_1 and n_2 with $n_1 + n_2 \leq n$ such that $\Lambda(i,i) > 0$ with $1 \leq i \leq n_1$, $\Lambda(i,i) < 0$ with $n_1 + 1 \leq i \leq n_1 + n_2$ and $\Lambda(i,i) = 0$ with $n_1 + n_2 + 1 \leq i \leq n$. Then we have $\tilde{O} = \tilde{O}_1 \times R^{n-n_1-n_2}$, where

$$\tilde{O}_1 = \left\{ y \in R^{n_1+n_2} \mid \sum_{i=1}^{n_1} \Lambda(i,i) y_i^2 - \sum_{i=n_1+1}^{n_1+n_2} |\Lambda(i,i)| y_i^2 < 0 \right\}.$$

For every $n_1 + 1 \leq i \leq n_1 + n_2$, let $y^{(i)} \in R^{n_1+n_2}$ be such that $y_i^{(i)} = 1$ and $y_j^{(i)} = 0$ for all $j \neq i$. For every $1 \leq i \leq n_1$, let $y^{(i)} \in R^{n_1+n_2}$ be such that $y_j^{(i)} = 1$ for all $n_1 + 1 \leq j \leq n_1 + n_2$, $y_i^{(i)} = \sqrt{\frac{\Delta}{2\Lambda(i,i)}}$, and $y_j^{(i)} = 0$ for all $1 \leq j \leq n_1$ but $j \neq i$, where $\Delta = \displaystyle\sum_{i=n_1+1}^{n_1+n_2} |\Lambda(i,i)|$. It is straightforward to check that $\pm y^{(i)} \in \tilde{O}_1$ for all $1 \leq i \leq n_1 + n_2$, and that the $n_1 + n_2$ vectors $y^{(i)}$ with $1 \leq i \leq n_1 + n_2$ are linearly independent. Thus, we have $\tilde{O}_1^* = \{0\}$ and $\tilde{O}^* = \{0\}$. Since $O^* = P\tilde{O}^*$, we have $O^* = \{0\}$. Since $L_C(\bar{x}^*) = \{0\}$, we have $O^* = L_C(\bar{x})^* = \{0\}$. In view of (4.2.36), (4.2.33) follows readily. This completes the proof. $\square$

We end this section by giving a class of parameterized problems to illustrate the application of Theorem 4.2.2. We identify exactly when KKT conditions can be verified by one of the existing constraint qualifications, and when they can be verified only by applying Theorem 4.2.2.

Example 4.2.2 *Consider the* (NLP) *defined as follows:*

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in C, \end{aligned} \tag{4.2.37}$$

where

$$C = \left\{ x \in R^3 \;\middle|\; \begin{array}{l} g_1(x) := a^T x + a_4 x_3^4 \leq 0 \\ g_2(x) := b^T x + b_4 x_3^4 \leq 0 \\ g_3(x) := c^T x + c_4 x_3^4 \leq 0 \end{array} \right\}$$

is given as in Example 2.2.4. Let $\bar{x} = 0 \in R^3$ and let

$$S(x) := \max\{g_1(x), 0\} + \max\{g_2(x), 0\} + \max\{g_2(x), 0\} \qquad \forall x \in R^3.$$

Consider all the cases for the data on C as specified in Example 2.2.4. For the cases (A), (B1), (B2), (B4-1), (C1), (C2), (C3-1), (C4), (C5-1), (C6), *and* (D1), *it follows from Example 2.2.4 that S is a local error bound for C at $\bar{x}$ indicating that the* ACQ *holds at $\bar{x}$. Therefore, in these cases, the* KKT *condition holds at $\bar{x}$ whenever $\bar{x}$ is a local minimum of (4.2.37).*

But in the remaining cases, it follows from the discussion given in Example 2.2.4 that the GCQ *does not hold at $\bar{x}$. That is, the* KKT *condition may or may not hold at $\bar{x}$ even if $\bar{x}$ is a local minimum of (4.2.37).*

Fortunately, we observe from Proposition 4.2.2 (iii) that S^p with $0.5 \leq p \leq 1$ is a KKT-type *penalty term at $\bar{x}$. As a result, if we can show that the l_p penalty function*

$$\mathcal{F}_p(x) = f(x) + \mu S^p(x)$$

is exact at $\bar{x}$ for some $p \in [0.5, 1]$, then we can obtain the KKT *condition at $\bar{x}$.*

In what follows, we consider in particular the case (B3) *for the feasible set C. That is, the vectors a and b are linearly independent, and the vector $c = -k_1 a - k_2 b$ with $k_1 \geq 0$, $k_2 \geq 0$, $k_1 k_2 = 0$ and $k_1 a_4 + k_2 b_4 + c_4 > 0$. In addition, we assume that the vectors a, b and $e_3 = (0, 0, 1)^T$ are linearly independent. Suppose that the objective function f takes the form*

$$f(x) = w^T x + w_4 x_3^2 \qquad \forall x \in R^3, \tag{4.2.38}$$

where $w = -\rho_1 a - \rho_2 b$ with $\rho_1 \geq 0$, $\rho_2 \geq 0$ and $w_4 < 0$.

First, we show that the l_p penalty function cannot be exact at $\bar{x}$ when $p > 0.5$. Since the vectors a, b and e_3 are linearly independent, we can find a sequence $x_k = (x_{1k}, x_{2k}, x_{3k})^T \in R^3$ such that $x_k \to \bar{x}$, $x_{3k} \not\equiv 0$, $a^T x_k = 0$ and $b^T x_k = 0$. For such a sequence, it is easy to check that for any $\mu > 0$ and $p > 0.5$, there exists k_0 such that the inequality

$$\mathcal{F}_p(x_k) = w_4 x_{3k}^2 + \mu \left[(a_4 x_{3k}^4)_+ + (b_4 x_{3k}^4)_+ + (c_4 x_{3k}^4)_+ \right]^p < 0 = \mathcal{F}_p(\bar{x})$$

holds for all $k \geq k_0$. Thus, the l_p penalty function is not exact at $\bar{x}$ when $p > 0.5$.

Next, we show that the $l_{0.5}$ penalty function is exact at $\bar{x}$. It is easy to see that

$$\begin{aligned} \mathcal{F}_{0.5}(x) := \; & [w_4 + (\rho_1 a_4 + \rho_2 b_4) x_3^2] x_3^2 - \rho_1 (a^T x + a_4 x_3^4) - \rho_2 (b^T x + b_4 x_3^4) \\ & + \mu \sqrt{(a^T x + a_4 x_3^4)_+ + (b^T x + b_4 x_3^4)_+ + (c^T x + c_4 x_3^4)_+}. \end{aligned}$$

Let $\delta = \min\{1, \frac{1}{\|a\| + |a_4|}, \frac{1}{\|b\| + |b_4|}\}$ and

$$\tilde{\mu} = 2 \max \left\{ 2\rho_1, 2\rho_2, \frac{|\rho_1 a_4 + \rho_2 b_4| - w_4}{\sqrt{\min\{\frac{1}{k_1}, \frac{1}{k_2}, 1\}(k_1 a_4 + k_2 b_4 + c_4)}} \right\},$$

where the convention $\frac{1}{0} := \infty$ is used when $k_1 = 0$ or $k_2 = 0$. Let $\mu \geq \tilde{\mu}$ and $\|x\| \leq \delta$. By the definition of δ, we have

$$(\rho_1 a_4 + \rho_2 b_4) x_3^2 \geq -|\rho_1 a_4 + \rho_2 b_4| \delta^2 \geq -|\rho_1 a_4 + \rho_2 b_4|, \tag{4.2.39}$$

and

$$|a^T x + a_4 x_3^4| \leq \|a\| \|x\| + |a_4| x_3^4 \leq \|a\| \delta + |a_4| \delta^4 \leq (\|a\| + |a_4|) \delta \leq 1, \tag{4.2.40}$$

and similarly,

$$|b^T x + b_4 x_3^4| \leq 1. \tag{4.2.41}$$

Thus, we obtain from (4.2.40) and (4.2.41)

$$\begin{array}{rl} & \sqrt{(a^T x + a_4 x_3^4)_+ + (b^T x + b_4 x_3^4)_+ + (c^T x + c_4 x_3^4)_+} \\ \geq & \sqrt{(a^T x + a_4 x_3^4)_+ + (b^T x + b_4 x_3^4)_+} \\ \geq & \frac{1}{2}\sqrt{(a^T x + a_4 x_3^4)_+} + \frac{1}{2}\sqrt{(b^T x + b_4 x_3^4)_+} \\ \geq & \frac{1}{2}(a^T x + a_4 x_3^4)_+ + \frac{1}{2}(b^T x + b_4 x_3^4)_+, \end{array} \tag{4.2.42}$$

where the second inequality follows from Lemma 4.1 in [80]. Since $k_1 a_4 + k_2 b_4 + c_4 > 0$, we have

$$\begin{array}{rl} & \sqrt{(a^T x + a_4 x_3^4)_+ + (b^T x + b_4 x_3^4)_+ + (c^T x + c_4 x_3^4)_+} \\ \geq & \sqrt{\min\{\frac{1}{k_1}, \frac{1}{k_2}, 1\}}\sqrt{k_1(a^T x + a_4 x_3^4)_+ + k_2(b^T x + b_4 x_3^4)_+ + (c^T x + c_4 x_3^4)_+} \\ \geq & \sqrt{\min\{\frac{1}{k_1}, \frac{1}{k_2}, 1\}}\sqrt{[k_1(a^T x + a_4 x_3^4) + k_2(b^T x + b_4 x_3^4) + (c^T x + c_4 x_3^4)]_+} \\ = & \left(\sqrt{\min\{\frac{1}{k_1}, \frac{1}{k_2}, 1\}}\sqrt{k_1 a_4 + k_2 b_4 + c_4}\right) x_3^2. \end{array} \tag{4.2.43}$$

In view of (4.2.39), (4.2.42), (4.2.43), and the definition of $\tilde{\mu}$, we have

$$\begin{aligned} \mathcal{F}_{0.5}(x) \; & \geq (w_4 - |\rho_1 a_4 + \rho_2 b_4|) x_3^2 - \rho_1(a^T x + a_4 x_3^4) - \rho_2(b^T x + b_4 x_3^4) \\ & \quad + \tfrac{\mu}{4}(a^T x + a_4 x_3^4)_+ + \tfrac{\mu}{4}(b^T x + b_4 x_3^4)_+ + \tfrac{\mu}{2}\left(\sqrt{\min\{\tfrac{1}{k_1}, \tfrac{1}{k_2}, 1\}}\sqrt{k_1 a_4 + k_2 b_4 + c_4}\right) x_3^2 \\ & \geq 0, \end{aligned}$$

which implies that the $l_{0.5}$ penalty function is exact at $\bar{x}$.

4.3 Second-order Necessary Conditions via Exact Penalty Functions

In this section, by applying the second-order necessary conditions presented in Theorem 3.3.3 for exact penalty functions and by applying the duality theorem of linear programming, we derive second-order necessary conditions of the type (4.1.1) for (NLP) from exactness of $f + \mu\phi$ under some additional condition in terms of the kernel of the parabolic subderivative of ϕ. By applying the third order Taylor expansions and in terms of the original data of (NLP), we give some sufficient conditions for these conditions when the l_p penalty functions are involved.

To begin with, we establish a condition which allows us to derive the second-order necessary condition (4.1.1) for (NLP) from exactness of $f + \mu\phi$.

Theorem 4.3.1 *Suppose that $f + \mu\phi$ is exact at $\bar{x}$. If*

$$L_C^2(\bar{x} \mid w) \subset \operatorname{clconv}[\ker d^2\phi(\bar{x})(w \mid \cdot)] \qquad \forall w \in \mathcal{V}(\bar{x}), \tag{4.3.44}$$

then

$$\sup_{\lambda \in \mathrm{KKT}(\bar{x})} \langle w, \nabla_{xx}^2 L(\bar{x}, \lambda) w\rangle \geq 0 \qquad \forall w \in \mathcal{V}(\bar{x}). \tag{4.3.45}$$

In particular when $L_C^2(\bar{x} \mid w) = \emptyset$, the supremum in (4.3.45) is $+\infty$.

Proof. Observe that $0 \in \mathcal{V}(\bar{x})$. It follows from the definition of $L_C^2(\bar{x} \mid 0)$ and Proposition 3.2.2 (ii) that, (4.3.44) holds with $w = 0$ if and only if

$$L_C(\bar{x}) \subset \operatorname{clconv}[\ker d\phi(\bar{x})],$$

or equivalently

$$[\ker d\phi(\bar{x})]^* \subset L_C(\bar{x})^*.$$

This implies by Theorem 4.2.1 that $\mathrm{KKT}(\bar{x}) \neq \emptyset$. By the definition of $\mathcal{V}(\bar{x})$, it is easy to verify that for any $w \in \mathcal{V}(\bar{x})$,

$$\mathrm{KKT}(\bar{x}) = \left\{ \lambda \in R^{m+q} \;\middle|\; \begin{array}{ll} \nabla_x L(\bar{x}, \lambda) = 0 & \\ \lambda_i \geq 0 & \forall i \in I(\bar{x}, w) \\ \lambda_i = 0 & \forall i \in I \backslash I(\bar{x}, w) \end{array} \right\}. \tag{4.3.46}$$

According to Theorem 3.3.3 or the weakened necessary condition (3.3.27) in Remark 3.3.2, there exists no vector $(w, z) \in R^n \times R^n$ such that

$$\langle \nabla f(\bar{x}), w\rangle = 0,\ \langle \nabla f(\bar{x}), z\rangle + \langle w, \nabla^2 f(\bar{x}) w\rangle < 0,\ z \in \operatorname{clconv}[\ker d^2\phi(\bar{x})(w \mid \cdot)]. \tag{4.3.47}$$

Let $w \in \mathcal{V}(\bar{x})$. First we assume that $L_C^2(\bar{x} \mid w) \neq \emptyset$. It follows from (4.3.44), the definition of $L_C^2(\bar{x} \mid w)$, and the inconsistency of the system (4.3.47) that, the optimal value of the linear program

$$\begin{array}{lll} \min\limits_{z \in R^n} & \langle \nabla f(\bar{x}), z\rangle + \langle w, \nabla^2 f(\bar{x}) w\rangle & \\ \text{s.t.} & \langle \nabla g_i(\bar{x}), z\rangle + \langle w, \nabla^2 g_i(\bar{x}) w\rangle \leq 0 & \forall i \in I(\bar{x}, w), \\ & \langle \nabla h_j(\bar{x}), z\rangle + \langle w, \nabla^2 h_j(\bar{x}) w\rangle = 0 & \forall j \in J, \end{array} \tag{4.3.48}$$

is nonnegative. Applying the duality theorem of linear programming (see [107]), we confirm that the optimal value of the linear program

$$\begin{aligned}\max_{\lambda \in R^{m+q}} \quad & \langle w, \nabla^2_{xx} L(\bar{x}, \lambda) w \rangle \\ \text{s.t.} \quad & \nabla_x L(\bar{x}, \lambda) = 0, \\ & \lambda_i \geq 0 \quad \forall i \in I(\bar{x}, w), \\ & \lambda_i = 0 \quad \forall i \in I \backslash I(\bar{x}, w),\end{aligned}$$

is also nonnegative. This together with (4.3.46) implies that $\mathrm{KKT}(\bar{x}) \neq \emptyset$ and

$$\max_{\lambda \in \mathrm{KKT}(\bar{x})} \langle w, \nabla^2_{xx} L(\bar{x}, \lambda) w \rangle \geq 0.$$

Next we assume that $L^2_C(\bar{x} \mid w) = \emptyset$. Since $\mathcal{V}(\bar{x}) \subset L_C(\bar{x})$, we have $w \in L_C(\bar{x})$. It follows from the definition of $L^2_C(\bar{x} \mid w)$ that, there exists no $z \in R^n$ such that

$$\begin{aligned}\langle \nabla g_i(\bar{x}), z \rangle + \langle w, \nabla^2 g_i(\bar{x}) w \rangle \leq 0 \quad & \forall i \in I(\bar{x}, u), \\ \langle \nabla h_j(\bar{x}), z \rangle + \langle w, \nabla^2 h_j(\bar{x}) w \rangle = 0 \quad & \forall j \in J.\end{aligned}$$

The duality theorem of linear programming (see [107]) guarantees the existence of some $\tilde{\lambda} \in R^{m+q}$ with $\tilde{\lambda}_i \geq 0$ for all $i \in I(\bar{x}, w)$ and $\tilde{\lambda}_i = 0$ for all $i \in I \backslash I(\bar{x}, w)$ such that

$$\sum_{i \in I} \tilde{\lambda}_i \nabla g_i(\bar{x}) + \sum_{j \in J} \tilde{\lambda}_j \nabla h_j(\bar{x}) = 0, \tag{4.3.49}$$

and

$$\sum_{i \in I} \tilde{\lambda}_i \langle w, \nabla^2 g_i(\bar{x}) w \rangle + \sum_{j \in J} \tilde{\lambda}_j \langle w, \nabla^2 h_j(\bar{x}) w \rangle > 0. \tag{4.3.50}$$

Let $\bar{\lambda} \in \mathrm{KKT}(\bar{x})$ and let $\lambda_t = \bar{\lambda} + t\tilde{\lambda}$ for all $t \geq 0$. It follows from (4.3.46), (4.3.49), and (4.3.50) that $\lambda_t \in \mathrm{KKT}(\bar{x})$ for all $t \geq 0$, and that

$$\sup_{\lambda \in \mathrm{KKT}(\bar{x})} \langle w, \nabla^2_{xx} L(\bar{x}, \lambda) w \rangle \geq \sup_{t \geq 0} \langle w, \nabla^2_{xx} L(\bar{x}, \lambda_t) w \rangle = +\infty.$$

This completes the proof. □

Let $0 \leq p \leq 1$ and let S be given by (4.1.2). Since all functions in defining (NLP) are assumed to be twice continuously differentiable, it is easy to verify by applying the second-order Taylor expansion that

$$L^2_C(\bar{x} \mid w) = \mathrm{clconv}[\mathrm{ker} d^2 S(\bar{x})(w \mid \cdot)] \qquad \forall w \in \mathcal{V}(\bar{x}),$$

which indicates that condition (4.3.44) holds automatically when ϕ happens to be the function S. This recovers a well-known result that the second-order necessary condition (4.3.45) holds when the l_1 penalty function is exact at $\bar{x}$, see Corollary 4.5 of Rockafellar [139]. Moreover, in view of Proposition 3.2.2 (v) and the fact that $L^2_C(\bar{x} \mid w)$ is closed and convex due to being a polyhedron by definition, condition (4.3.44) holds when $\phi = S^p$ with $0 \leq p < 1$ if and only if

$$L^2_C(\bar{x} \mid w) = \text{clconv}[\text{ker}d^2 S^p(\bar{x})(w \mid \cdot)] \qquad \forall w \in \mathcal{V}(\bar{x}). \tag{4.3.51}$$

Note that when $p = 0$, we interpret S^p as the indicator function δ_C of the feasible set C of (NLP). Thus, in view of Proposition 3.2.2 (iv), condition (4.3.51) with $p = 0$ reduces to the SGCQ. Therefore, Theorem 4.3.1 also recovers a known result that, under the SGCQ, any local minimum of (NLP) satisfies the second-order necessary condition (4.3.45), see Kawasaki [94].

In what follows, by assuming that all functions in defining (NLP) are three times continuously differentiable, we shall given sufficient conditions in terms of the original data for

$$L^2_C(\bar{x} \mid w) = \text{ker}d^2 S^p(\bar{x})(w \mid \cdot) \qquad \forall w \in L_C(\bar{x}),$$

which is slightly stronger than (4.3.51).

Assume that $\psi : R^n \to R$ is a three times continuously differentiable function. Let $\{t_k\} \subset R_+$ be a sequence such that $t_k \to 0+$, and let $(x, u) \in R^n \times R^n$ be such that $\psi(x) = 0$ and $\langle \nabla\psi(x), w\rangle = 0$. For each $z \in R^n$, it follows from the third order Taylor expansion that,

$$
\begin{aligned}
& \psi(x + t_k w + \frac{1}{2}t_k^2 z) \\
= \;& \psi(x) + t_k\langle \nabla\psi(x), w\rangle + \frac{1}{2}t_k^2\langle \nabla\psi(x), z\rangle + \frac{1}{2}t_k^2\langle w + \frac{1}{2}t_k z, \nabla^2\psi(x)(w + \frac{1}{2}t_k z)\rangle \\
& + \frac{1}{6}t_k^3\psi^{(3)}(x)(w + \frac{1}{2}t_k z, w + \frac{1}{2}t_k z, w + \frac{1}{2}t_k z) + o(t_k^3) \\
= \;& \frac{1}{2}t_k^2\langle \nabla\psi(x), z\rangle + \frac{1}{2}t_k^2\langle w + \frac{1}{2}t_k z, \nabla^2\psi(x)(w + \frac{1}{2}t_k z)\rangle \\
& + \frac{1}{6}t_k^3\psi^{(3)}(x)(w + \frac{1}{2}t_k z, w + \frac{1}{2}t_k z, w + \frac{1}{2}t_k z) + o(t_k^3) \\
= \;& \frac{1}{2}t_k^2[\langle \nabla\psi(x), z\rangle + \langle w, \nabla^2\psi(x)w\rangle] + \frac{1}{2}t_k^3\langle w, \nabla^2\psi(x)z\rangle + \frac{1}{8}t_k^4\langle z, \nabla^2\psi(x)z\rangle \\
& + \frac{1}{6}t_k^3\psi^{(3)}(x)(w + \frac{1}{2}t_k z, w + \frac{1}{2}t_k z, w + \frac{1}{2}t_k z) + o(t_k^3).
\end{aligned}
\tag{4.3.52}
$$

Proposition 4.3.1 *Let S be given by (4.1.2). Assume that all the functions g_i with $i \in I$ and h_j with $j \in J$ are three times continuously differentiable. For each $w \in L_C(\bar{x})$, the following statements are true:*

(*i*) $\mathrm{ker}d^2S^p(\bar{x})(w \mid \cdot) = L_C^2(\bar{x} \mid w) \quad \forall p \in (\frac{2}{3}, 1]$.

(*ii*) $\mathrm{ker}d^2S^{\frac{2}{3}}(\bar{x})(w \mid \cdot) = L_C^2(\bar{x} \mid w)$ *if for every $z \in L_C^2(\bar{x} \mid w)$, it follows that*

$$
\begin{cases}
\langle w, \nabla^2 g_i(\bar{x})z\rangle + \frac{1}{3}g_i^{(3)}(\bar{x})(w, w, w) \leq 0 & \forall\, i \in I(\bar{x}, w, z), \\
\langle w, \nabla^2 h_j(\bar{x})z\rangle + \frac{1}{3}h_j^{(3)}(\bar{x})(w, w, w) = 0 & \forall\, j \in J.
\end{cases}
\tag{4.3.53}
$$

(*iii*) *If $q = 0$ (i.e., there is no equality constraint), and for every $z \in L_C^2(\bar{x} \mid w)$ with $(w, z) \neq 0$, it follows that*

$$
\langle w, \nabla^2 g_i(\bar{x})z\rangle + \frac{1}{3}g_i^{(3)}(\bar{x})(w, w, w) < 0 \qquad \forall\, i \in I(\bar{x}, w, z), \tag{4.3.54}
$$

then $\mathrm{ker}d^2S^p(\bar{x})(w \mid \cdot) = L_C^2(\bar{x} \mid w)$ *for all $p \in [0, 1]$.*

Proof. If $L_C^2(\bar{x} \mid w) = \emptyset$, we have by Lemma 3.2.2 (v)

$$
\mathrm{ker}d^2S^p(\bar{x})(w \mid \cdot) = L_C^2(\bar{x} \mid w) = \emptyset \qquad \forall p \in [0, 1].
$$

In what follows, we assume that $L^2_C(\bar{x} \mid w) \neq \emptyset$. Let $z \in L^2_C(\bar{x} \mid w)$ and let $t_k \to 0+$. For any $i \in I \backslash I(\bar{x}, w)$, we have either $g_i(\bar{x}) < 0$ or $g_i(\bar{x}) = 0$ with $\langle \nabla g_i(\bar{x}), w \rangle < 0$. In either case, it is easy to verify that for all sufficiently large k,

$$g_i(\bar{x} + t_k u + \frac{1}{2} t_k^2 w) \leq 0 \qquad \forall i \in I \backslash I(\bar{x}, w). \tag{4.3.55}$$

By (4.3.52) and the definition of $L^2_C(\bar{x} \mid w)$, we have for every $\alpha \in (0, 3)$

$$\max \left\{ \frac{g_i(\bar{x} + t_k w + \frac{1}{2} t_k^2 z)}{t_k^\alpha}, 0 \right\} \to 0 \qquad \forall i \in I(\bar{x} \mid w), \tag{4.3.56}$$

and

$$\frac{h_j(\bar{x} + t_k w + \frac{1}{2} t_k^2 z)}{t_k^\alpha} \to 0 \qquad \forall j \in J. \tag{4.3.57}$$

Combining (4.3.55), (4.3.56), and (4.3.57), we have for every $p \in (\frac{2}{3}, 1]$,

$$\begin{aligned} &\frac{S^p(\bar{x} + t_k w + \frac{1}{2} t_k^2 z)}{t_k^2} \\ = &\left(\sum_{i \in I} \max \left\{ \frac{g_i(\bar{x} + t_k w + \frac{1}{2} t_k^2 z)}{t_k^{2/p}}, 0 \right\} + \sum_{j \in J} \left| \frac{h_j(\bar{x} + t_k w + \frac{1}{2} t_k^2 z)}{t_k^{2/p}} \right| \right)^p \\ \to &\ 0. \end{aligned}$$

This implies by Lemma 3.2.2 (i) that $L^2_C(\bar{x} \mid w) \subset \ker d^2 S^p(\bar{x})(w \mid \cdot)$ for any $p \in (\frac{2}{3}, 1]$. In view of Lemma 3.2.2 (v), statement (i) is true.

Now we show that statement (ii) is true. It follows from (4.3.52), condition (4.3.54), and the definition of $L^2_C(\bar{x} \mid w)$ that (4.3.56) and (4.3.57) hold with $\alpha = 3$. Thus, together with (4.3.55), we have

$$\frac{S^{\frac{2}{3}}(\bar{x} + t_k w + \frac{1}{2} t_k^2 z)}{t_k^2} \to 0.$$

This implies by Lemma 3.2.2 (i) that $L^2_C(\bar{x}, u) \subset \ker d^2 S^{\frac{2}{3}}(\bar{x})(w \mid \cdot)$. In view of Lemma 3.2.2 (v), statement (ii) is true.

Finally, we show that statement (iii) is true. It follows from (4.3.52) and condition (4.3.54) that, for all sufficiently large k,

$$g_i(\bar{x} + t_k w + \frac{1}{2} t_k^2 z) \leq 0 \qquad \forall i \in I(\bar{x}, w). \tag{4.3.58}$$

Combining (4.3.55) and (4.3.58), we have for all sufficiently large k,

$$\bar{x} + t_k w + \frac{1}{2} t_k^2 z \in C.$$

This implies by definition that $z \in T_C^2(\bar{x} \mid w)$. Thus, we have $L_C^2(\bar{x} \mid w) \subset T_C^2(\bar{x} \mid w)$. In view of Lemma 3.2.2 (v), statement (iii) is true. This completes the proof. $\square$

In the following example, we demonstrate that condition (4.3.53) may not hold even if the LICQ holds.

Example 4.3.1 *In* (NLP), *let* $n = 2$, $m = 1$, $q = 0$, *and* $g_1(x) = x_1^3 - x_2$. *Consider a feasible point* $\bar{x} = (0,0)^T$. *Since* $\nabla g(\bar{x}) = (0,-1)^T$, *the* LICQ *holds at* $\bar{x}$. *This implies that* S *is a local error bound for* C *at* $\bar{x}$. *By Lemma 3.2.2, we have for any* $p \in [0,1]$ *and any* $w \in L_C(\bar{x}) = R \times R_+$,

$$T_C^2(\bar{x} \mid w) = \mathrm{ker} d^2 S^p(\bar{x})(w \mid \cdot) = L_C^2(\bar{x} \mid w) = \begin{cases} R \times R_+ & \text{if } w_2 = 0, \\ R^2 & \text{otherwise.} \end{cases}$$

Let $w \in L_C(\bar{x})$ *with* $w_2 = 0$ *and let* $z \in L_C^2(\bar{x} \mid w)$ *with* $z_2 = 0$. *By definition, we have* $I(\bar{x}, w, z) = \{1\}$. *Now it is easy to check that condition (4.3.53) is invalid when* $w_1 > 0$ *because*

$$\langle w, \nabla^2 g(\bar{x}) z \rangle + \frac{1}{3} g^{(3)}(\bar{x})(w,w,w) = 2w_1^3 > 0.$$

In the following example, we illustrate that even if neither the GCQ nor the SGCQ holds, Theorem 4.3.1 can be applied to derive the second-order necessary condition.

Example 4.3.2 *In* (NLP), *let* $n = 2$, $m = 3$, $q = 0$, $f(x) = -x_1^4 + x_2$, $g_1(x) = -x_2$, $g_2(x) = x_1^6 + x_2^3$, $g_3(x) = -x_1^2 + x_2^2$, *and let* $\bar{x} = 0$. *By a direct calculation, we have* $T_C(\bar{x}) = \{0\}$, $L_C(\bar{x}) = \{w \in R^2 \mid w_2 \geq 0\}$, *and* $\mathcal{V}(\bar{x}) = \{w \in R^2 \mid w_2 = 0\}$. *Moreover, we have*

$$T_C^2(\bar{x} \mid w) = \begin{cases} \{0\} & \text{if } w = 0, \\ \emptyset & \text{otherwise,} \end{cases}$$

and

$$L_C^2(\bar{x} \mid w) = \begin{cases} R \times R_+ & \text{if } w_2 = 0, \\ R^2 & \text{if } w_2 > 0, \\ \emptyset & \text{otherwise.} \end{cases}$$

Thus, for any $w \in \mathcal{V}(\bar{x})$, *we have*

$$L_C^2(\bar{x} \mid w) \neq \mathrm{clconv}[T_C^2(\bar{x} \mid w)].$$

By definition or by checking condition (4.3.53), we have for any $w \in \mathcal{V}(\bar{x})$,

$$L_C^2(\bar{x} \mid w) = \ker d^2 S^{\frac{2}{3}}(\bar{x})(w \mid \cdot).$$

In what follows, we will show that the l_p penalty function is exact at $\bar{x}$ for $p = \frac{2}{3}$ but not for $p > \frac{2}{3}$. Let $\delta \in (0,1)$ and $\tilde{\mu} = \dfrac{2}{(1-\delta^2)^{\frac{2}{3}}}$. Clearly, $\tilde{\mu} > 2$. Let $\mu \geq \tilde{\mu}$ and let $x \in R^2$ be such that $|x_1| \leq \delta$ and $|x_2| \leq \delta$. We consider two cases for x:

Case 1: $x_2 \geq 0$. *We have*

$$\begin{aligned} \mathcal{F}_{\frac{2}{3}}(x) &= -x_1^4 + x_2 + \mu\left[(-x_2)_+ + (x_1^6 + x_2^3)_+ + (-x_1^2 + x_2^2)_+\right]^{\frac{2}{3}} \qquad (4.3.59) \\ &\geq -x_1^4 + \mu(x_1^6)^{\frac{2}{3}} \\ &= (\mu - 1)x_1^4 \\ &\geq 0. \end{aligned}$$

Case 2: $x_2 < 0$. *We have from (4.3.59)*

$$\begin{aligned} \mathcal{F}_{\frac{2}{3}}(x) &\geq -x_1^4 + x_2 + \mu[(-x_2 + x_1^6 + x_2^3)_+]^{\frac{2}{3}} \\ &= -x_1^4 + x_2 + \mu[-x_2(1 - x_2^2) + x_1^6]^{\frac{2}{3}} \\ &\geq -x_1^4 + x_2 + \frac{1}{2}\mu[-x_2(1-x_2^2)]^{\frac{2}{3}} + \frac{1}{2}\mu(x_1^6)^{\frac{2}{3}} \\ &\geq (\frac{1}{2}\mu - 1)x_1^4 + (-x_2)[\frac{\mu}{2}(1-\delta^2)^{\frac{2}{3}} - 1] \\ &\geq 0, \end{aligned}$$

where the second inequality follows from Lemma 4.1 in [80].

To show that the l_p penalty function is not exact at $\bar{x}$ when $p > \frac{2}{3}$, we consider a sequence $x_k := (x_{1k}, 0) \in R^2$ with $x_{1k} \to 0+$. It is easy to check that for any $\mu > 0$ and $p > \frac{2}{3}$, the following condition holds for all sufficiently large k:

$$\mathcal{F}_p(x_k) = -x_{1k}^4 + \mu x_{1k}^{6p} < \mathcal{F}_p(\bar{x}) = 0.$$

Now Theorem 4.3.1 can be applied to derive the second-order necessary condition at $\bar{x}$. In fact, by a direct calculation, we have

$$\mathrm{KKT}(\bar{x}) = \{\lambda \in R^3 \mid \lambda_1 = 1, \lambda_2 \geq 0, \lambda_3 \geq 0\}.$$

Thus, for each $w \in \mathcal{V}(\bar{x})$, we have

$$\sup_{\lambda \in \mathrm{KKT}(\bar{x})} \langle w, \nabla^2_{xx} L(\bar{x}, \lambda) w \rangle = \sup_{\lambda_3 \geq 0} (-2\lambda_3 w_1^2) = 0.$$

Chapter 5

First-order Necessary Conditions for MPCC via Exact Penalty Functions

5.1 Introduction

In this chapter, we study first-order necessary conditions for a mathematical program with complementarity constraints as follows:

$$
\begin{array}{llll}
\text{(MPCC)} & \min & f(x) & \\
& \text{s.t.} & g_i(x) \le 0, & i \in I, \\
& & h_j(x) = 0, & j \in J, \\
& & G_k(x) \ge 0, H_k(x) \ge 0, G_k(x)H_k(x) = 0, & k \in K,
\end{array}
$$

where $f, g_i, i \in I, h_j, j \in J$ are given as in (NLP), and $G_k, H_k : R^n \to R, k \in K := \{m+q+1, m+q+2, \cdots, m+q+l\}$ are assumed to be continuously differentiable. Throughout this chapter, we denote by E the feasible set of (MPCC) and by $\bar{x}$ a fixed point in E.

Stationarity (or first-order necessary) conditions for (MPCC) have been the subject of many recent papers and books, see [146, 147, 105, 106, 165, 125, 54, 55, 57]. Since there are several different approaches for deriving optimality conditions, various stationarity conditions arise, see a recent PhD thesis by Flegel [52] for their definitions and connections.

In this chapter, we focus on strong stationarity and Mordukhovich stationarity only. Specifically, we say that $\bar{x}$ is a strongly (resp. an Mordukhovich) stationary point of (MPCC) if, there is $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in R^{m+q+2l}$ such that

$$\begin{gathered}
\nabla f(\bar{x}) + \sum_{i\in I} \lambda_i^g \nabla g_i(\bar{x}) + \sum_{j\in J} \lambda_j^h \nabla h_j(\bar{x}) - \sum_{k\in K} [\lambda_k^G \nabla G_k(\bar{x}) + \lambda_k^H \nabla H_k(\bar{x})] = 0, \\
\forall i \in I(\bar{x}), \lambda_i^g \geq 0, \quad \forall i \in I \backslash I(\bar{x}), \lambda_i^g = 0, \\
\forall k \in \gamma, \lambda_k^G = 0, \quad \forall k \in \alpha, \lambda_k^H = 0, \\
\forall k \in \beta, \lambda_k^G \geq 0, \lambda_k^H \geq 0 \\
(\text{resp. } \forall k \in \beta, \text{ either } \lambda_k^G > 0, \lambda_k^H > 0 \text{ or } \lambda_k^G \lambda_k^H = 0),
\end{gathered}$$

where $I(\bar{x}), \alpha, \beta, \gamma$ are useful index sets in the sequel:

$$\begin{aligned}
I(\bar{x}) &= \{i \in I \,|\, g_i(\bar{x}) = 0\}, \\
\alpha := \alpha(\bar{x}) &= \{k \in K \,|\, 0 = G_k(\bar{x}) < H_k(\bar{x})\}, \\
\beta := \beta(\bar{x}) &= \{k \in K \,|\, G_k(\bar{x}) = H_k(\bar{x}) = 0\}, \\
\gamma := \gamma(\bar{x}) &= \{k \in K \,|\, G_k(\bar{x}) > H_k(\bar{x}) = 0\}.
\end{aligned}$$

Clearly, strong stationarity implies Mordukhovich stationarity. Note that $\bar{x}$ is a strongly stationary point if and only if the KKT condition holds at $\bar{x}$, see [56] for details. As for various standard CQs and MPCC tailored CQs for strong stationarity and Mordukhovich stationarity, we refer to the literature review in Chapter 1. Of particular note is that, as in the context of nonlinear programming, the GCQ is the weakest CQ for strong stationarity in the sense that it is both sufficient and necessary for local minima of (MPCC) to possess strong stationarity. There can be found in Flegel and Kanzow [57] the weakest CQ for Mordukhovich stationarity, though the authors did not formulate it explicitly. Its detailed formulation is presented at the very beginning of next section.

Strong stationarity and Mordukhovich stationarity can also be derived from exact penalty functions of (MPCC). Results on exact penalization for (MPCC) are known in the literature, see [105, 106, 106, 147, 168, 55, 52, 92, 102] for example.

Let $0 \leq p \leq 1$ and let the function S be given by (4.1.2). When treated as an ordinary nonlinear programming problem, (MPCC) can be associated with an l_p penalty function defined in the spirit of the penalty function $\mathcal{F}_p$ in Chapter 4 as follows:

$$\mathcal{G}_p(x) = f(x) + \mu U^p(x),$$

where $\mu \geq 0$ is the penalty parameter, and

$$U(x) := S(x) + \sum_{k \in K} ((-G_k(x))_+ + (-H_k(x))_+ + |G_k(x)H_k(x)|) \quad \forall x \in R^n.$$

With the aid of the NCP function $\phi_{\min}(a,b) := \min\{a,b\}$, (MPCC) can be associated with another l_p penalty function as follows:

$$\mathcal{H}_p(x) = f(x) + \mu V^p(x),$$

where $\mu \geq 0$ is the penalty parameter, and

$$V(x) = S(x) + \sum_{k \in K} |\phi_{\min}(G_k(x), H_k(x))| \quad \forall x \in R^n.$$

Note that in the case of $p = 0$, we interpret both U^p and V^p as the indicator function δ_E of the feasible set E of (MPCC). When $p = 1$, the penalty function $\mathcal{H}_p(x)$ reduces to the l_1 penalty function considered in [55, 165, 92]. It is well-known that the strong stationarity holds at $\bar{x}$ or equivalently the KKT condition holds at $\bar{x}$ if the l_1 penalty function $\mathcal{G}(x)$ has a local minimum at $\bar{x}$ for some $\mu \geq 0$, see in particular Han and Mangasarian ([66], Theorem 4.8). In contrast with this result, the Mordukhovich stationarity holds at $\bar{x}$ if the l_1 penalty function $\mathcal{H}(x)$ has a local minimum at $\bar{x}$ for some $\mu \geq 0$, see [55, 165, 92, 102]. But for $0 < p < 1$, the question as to whether and how strong stationarity or Mordukhovich stationarity can be derived from the l_p exact penalty function $\mathcal{G}_p$ or $\mathcal{H}_p$ has not been addressed in the literature.

In what follows, let $\phi : R^n \to R_+ \cup \{+\infty\}$ be a lower semicontinuous function with the property that $\phi(x) = 0$ if and only if $x \in E$. Note that ϕ plays the role as a general penalty term for the feasible set E of (MPCC), and that the functions U^p and V^p with $0 \leq p \leq 1$ are particular instances of ϕ. We say that ϕ is an S-type (resp., M-type) penalty term if it has the ability to detect strong stationarity (resp., Mordukhovich stationarity) for (MPCC) in the sense as described in the following definition.

Definition 5.1.1 *We say that ϕ is an* S-type *(resp.,* M-type*) penalty term at $\bar{x}$ if the strong stationarity (resp., the Mordukhovich stationarity) holds at $\bar{x}$ whenever there is* *a continuously differentiable function $f : R^n \to R$ such that the penalty function $f + \mu\phi$ is exact at $\bar{x}$ (i.e., for some finite penalty parameter $\mu \geq 0$, this penalty function admits an unconstrained local minimum at $\bar{x}$).*

By definition, S-type penalty terms are necessarily M-type. According to the previous literature review, U is of S-type and V is of M-type.

The outline of this chapter is as follows. In Section 5.2, we establish equivalent conditions for penalty terms of (MPCC) to be of S-type or M-type. In Section 5.3, we consider two l_p penalty functions: $\mathcal{G}_p$ and $\mathcal{H}_p$, and in terms of the original data of (MPCC), we give sufficient conditions for U^p and V^p to be of S-type and/or M-type. We end this chapter by establishing some relationships between these two penalty functions.

5.2 S-type and M-type Penalty Terms for MPCC

In this section, by applying the same technique as in Section 4.2.1, we establish equivalent conditions for penalty terms of (MPCC) to be of S-type or M-type.

To begin with, we introduce two cones first appeared in Flegel and Kanzow [57] as follows:

$$\Omega_1 = \left\{(u, \xi_\beta, \eta_\beta) \in R^{n+2|\beta|} \mid \xi_k \geq 0, \eta_k \geq 0, \xi_k\eta_k = 0,\ \forall k \in \beta\right\}$$

and

$$\Omega_2 = \left\{ (u, \xi_\beta, \eta_\beta) \in R^{n+2|\beta|} \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^T u \leq 0, & i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T u = 0, & j \in J \\ \nabla G_k(\bar{x})^T u = 0, & k \in \alpha \\ \nabla H_k(\bar{x})^T u = 0, & k \in \gamma \\ \nabla G_k(\bar{x})^T u - \xi_k = 0, & k \in \beta \\ \nabla H_k(\bar{x})^T u - \eta_k = 0, & k \in \beta \end{array} \right\}.$$

The strong stationarity and Mordukhovich stationarity can be characterized in a unified way in terms of (regular) normal cones to Ω_1 and Ω_2 as shown by Proposition 5.2.1 below. This idea is borrowed from [57], where a direct proof was given to show that local minima of (MPCC) posses Mordukhovich stationarity under the MPCC-GCQ. By definition, the MPCC-GCQ holds at $\bar{x}$ if $T_E(\bar{x})^* = T^{\text{lin}}_{\text{MPCC}}(\bar{x})^*$, where $T^{\text{lin}}_{\text{MPCC}}(\bar{x})$ is the

MPCC-linearized cone of (MPCC) at $\bar{x}$ defined by

$$T^{\text{lin}}_{\text{MPCC}}(\bar{x}) := \left\{ u \in R^n \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^T u \le 0, & i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T u = 0, & j \in J \\ \nabla G_k(\bar{x})^T u = 0, & k \in \alpha \\ \nabla H_k(\bar{x})^T u = 0, & k \in \gamma \\ \nabla G_k(\bar{x})^T u \ge 0, & k \in \beta \\ \nabla H_k(\bar{x})^T u \ge 0, & k \in \beta \\ (\nabla G_k(\bar{x})^T u)(\nabla H_k(\bar{x})^T u) = 0, & k \in \beta \end{array} \right\}.$$

This linearized cone, though not necessarily convex as the terminology suggests, was first introduced in [146, 125] and later studied extensively in [54, 165, 57]. In contrast with $T^{\text{lin}}_{\text{MPCC}}(\bar{x})$, the first-order linearized cone $T^{\text{lin}}(\bar{x})$ of (MPCC) at $\bar{x}$ defined by

$$T^{\text{lin}}(\bar{x}) := \left\{ u \in R^n \;\middle|\; \begin{array}{ll} \nabla g_i(\bar{x})^T u \le 0, & i \in I(\bar{x}) \\ \nabla h_j(\bar{x})^T u = 0, & j \in J \\ \nabla G_k(\bar{x})^T u = 0, & k \in \alpha \\ \nabla H_k(\bar{x})^T u = 0, & k \in \gamma \\ \nabla G_k(\bar{x})^T u \ge 0, & k \in \beta \\ \nabla H_k(\bar{x})^T u \ge 0, & k \in \beta \end{array} \right\}$$

is always convex and in particular polyhedral.

Proposition 5.2.1 *The following statements are true:*

(*i*) $\bar{x}$ *is a strongly stationary point of* (MPCC) *if and only if*

$$-\nabla f(\bar{x}) \in \{v \in R^n \mid (v,0,0) \in \widehat{N}_{\Omega_1}(0,0,0) + \widehat{N}_{\Omega_2}(0,0,0)\}.$$

(*ii*) $\bar{x}$ *is an Mordukhovich stationary point of* (MPCC) *if and only if*

$$-\nabla f(\bar{x}) \in \{v \in R^n \mid (v,0,0) \in N_{\Omega_1}(0,0,0) + N_{\Omega_2}(0,0,0)\}.$$

(*iii*) $T^{\text{lin}}(\bar{x})^* = \{v \in R^n \mid (v,0,0) \in \widehat{N}_{\Omega_1}(0,0,0) + \widehat{N}_{\Omega_2}(0,0,0)\}$.

(*iv*) $T^{\text{lin}}_{\text{MPCC}}(\bar{x})^* = \{v \in R^n \mid (v,0,0) \in \widehat{N}_{\Omega_1 \cap \Omega_2}(0,0,0)\}$.

(*v*) $N_{\Omega_1 \cap \Omega_2}(0,0,0) \subset N_{\Omega_1}(0,0,0) + N_{\Omega_2}(0,0,0)$.

Proof. By formulas of $\widehat{N}_{\Omega_1}(0,0,0)$, $\widehat{N}_{\Omega_2}(0,0,0)$, $N_{\Omega_1}(0,0,0)$, $N_{\Omega_2}(0,0,0)$ given in Proposition 2.2 of [102], statements (i) to (iii) follow directly from Farkas' lemma and the definitions of strong stationarity and Mordukhovich stationarity. Note that $(u, \xi_\beta, \eta_\beta) \in \Omega_1 \cap \Omega_2$ if and only if $u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x})$ and for every $k \in \beta$, $\xi_k = \nabla G_k(\bar{x})^T u$, $\eta_k = \nabla H_k(\bar{x})^T u$, and that $\widehat{N}_{\Omega_1 \cap \Omega_2}(0,0,0) = T_{\Omega_1 \cap \Omega_2}(0,0,0)^* = (\Omega_1 \cap \Omega_2)^*$. Thus, $(v,0,0) \in \widehat{N}_{\Omega_1 \cap \Omega_2}(0,0,0)$ if and only if

$$\langle (v,0,0), (u, \nabla G_\beta(\bar{x})^T u, H_\beta(\bar{x})^T u) \rangle = \langle v, u \rangle \le 0, \ \forall u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x}),$$

which amounts to $v \in T^{\text{lin}}_{\text{MPCC}}(\bar{x})^*$. Thus, statement (iv) is true. Statement (v) has been shown in [57]. This completes the proof. □

Remark 5.2.1 *According to [64], any $v \in T_E(\bar{x})^*$ corresponds to a continuously differentiable objective function f such that $\bar{x}$ is a local minimum of* (MPCC) *and $v = -\nabla f(\bar{x})$. Therefore, by statement (ii), the constraint qualification*

$$T_E(\bar{x})^* \times \{0\} \times \{0\} \subset N_{\Omega_1}(0,0,0) + N_{\Omega_2}(0,0,0) \tag{5.2.1}$$

is the weakest one for Mordukhovich stationarity. It follows from statements (iv) and (v), and $\widehat{N}_{\Omega_1 \cap \Omega_2}(0,0,0) \subset N_{\Omega_1 \cap \Omega_2}(0,0,0)$ that,

$$T^{\text{lin}}_{\text{MPCC}}(\bar{x})^* \times \{0\} \times \{0\} \subset N_{\Omega_1}(0,0,0) + N_{\Omega_2}(0,0,0). \tag{5.2.2}$$

It follows from (5.2.1) and (5.2.2) that $\bar{x}$ is an Mordukhovich stationary point if the MPCC-GCQ *holds at $\bar{x}$. This result has been shown by [165] and [52]). Example 5.2.1 below illustrates that* MPCC-GCQ *can be strictly stronger than the* CQ *defined by (5.2.1).*

Example 5.2.1 *In* (MPCC), *let $n = 1$, $m = q = 0$, $l = 1$, $G(x) = x$, $H(x) = x^2$, and $\bar{x} = 0$. Then $T_E(\bar{x}) = \{0\}$, $T^{\text{lin}}_{\text{MPCC}}(\bar{x}) = T^{\text{lin}}(\bar{x}) = R_+$, and*

$$\{v \in R \mid (v,0,0) \in N_{\Omega_1}(0,0,0) + N_{\Omega_2}(0,0,0)\} = R.$$

Thus, MPCC-GCQ *does not hold at $\bar{x}$ but (5.2.1) holds.*

Now we give equivalent conditions for ϕ to be of S-type or M-type at $\bar{x}$.

Theorem 5.2.1 *Concerning strong stationarity, we have the following equivalent conditions:*

(i) $[\mathrm{ker}d\phi(\bar{x})]^* \subset \{v \in R^n \mid (v,0,0) \in \widehat{N}_{\Omega_1}(0,0,0) + \widehat{N}_{\Omega_2}(0,0,0)\}$.

(ii) $\mathrm{pos}(\widehat{\partial}\phi(\bar{x})) \subset \{v \in R^n \mid (v,0,0) \in \widehat{N}_{\Omega_1}(0,0,0) + \widehat{N}_{\Omega_2}(0,0,0)\}$.

(iii) ϕ *is an* S-type *penalty term at* $\bar{x}$.

Concerning Mordukhovich stationarity, we have the following equivalent conditions:

(i') $[\mathrm{ker}d\phi(\bar{x})]^* \subset \{v \in R^n \mid (v,0,0) \in N_{\Omega_1}(0,0,0) + N_{\Omega_2}(0,0,0)\}$.

(ii') $\mathrm{pos}(\widehat{\partial}\phi(\bar{x})) \subset \{v \in R^n \mid (v,0,0) \in N_{\Omega_1}(0,0,0) + N_{\Omega_2}(0,0,0)\}$.

(iii') ϕ *is an* M-type *penalty term at* $\bar{x}$.

Proof. In view of Proposition 5.2.1, we can get all results by applying the same technique as in the proof of Theorem 4.2.1. □

Remark 5.2.2 *Due to Proposition 5.2.1 (iii), the right-hand sides in conditions* (i) *and* (ii) *are exactly the set* $T^{\mathrm{lin}}(\bar{x})^*$. *But if we replace the right-hand sides in conditions* (i') *and* (ii') *by* $T^{\mathrm{lin}}_{\mathrm{MPCC}}(\bar{x})^*$, *we can get only the sufficient conditions for* ϕ *being an* M-type *penalty term at* $\bar{x}$. *This is because the set* $T^{\mathrm{lin}}_{\mathrm{MPCC}}(\bar{x})^*$ *may be merely a proper subset of* $\{v \in R^n \mid (v,0,0) \in N_{\Omega_1}(0,0,0) + N_{\Omega_2}(0,0,0)\}$, *as can be seen from Example 5.2.1.*

5.3 On Two Particular Penalty Functions for MPCC

In this section, we consider the l_p penalty functions $\mathcal{G}_p$ and $\mathcal{H}_p$, which have been defined in the Introduction section. By applying Theorem 5.2.1 obtained in last section, we give sufficient conditions in terms of the original data of (MPCC) for U^p and V^p to be S-type and/or M-type penalty terms. We end this chapter by establishing the relationships between exactness of $\mathcal{G}_p$ and $\mathcal{H}_p$, and between $\mathrm{ker}dU^p(\bar{x})$ and $\mathrm{ker}dV^p(\bar{x})$.

Recall that in defining $\mathcal{G}_p$, we use the function

$$U(x) = S(x) + \sum_{k\in K} \{(-G_k(x))_+ + (-H_k(x))_+ + |G_k(x)H_k(x)|\}, \tag{5.3.3}$$

where S is given by (4.1.2).

Lemma 5.3.1 $\ker dU^{0.5}(\bar{x}) \subset T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \subset \ker dU(\bar{x}) = T^{\text{lin}}(\bar{x})$.

Proof. By Proposition 3.2.1 (i), it is easy to verify that

$$\ker dU^1(\bar{x}) = T^{\text{lin}}(\bar{x}).$$

Let $u \in \ker dU^{0.5}(\bar{x})$ and let $k \in \beta$. To show $\ker dU^{0.5}(\bar{x}) \subset T^{\text{lin}}_{\text{MPCC}}(\bar{x})$, it suffices to show

$$(\nabla G_k(\bar{x})^T u)(\nabla H_k(\bar{x})^T u) = 0. \tag{5.3.4}$$

By Proposition 3.2.1 (i) again, there exist $t_\nu \to 0+$ and $u_\nu \to u$ such that

$$\frac{G_k(\bar{x} + t_\nu u_\nu)H_k(\bar{x} + t_\nu u_\nu)}{(t_\nu)^2} \to 0.$$

Since G_k and H_k are continuously differentiable, we have by Taylor expansion rule

$$\frac{G_k(\bar{x} + t_\nu u_\nu)H_k(\bar{x} + t_\nu u_\nu)}{(t_\nu)^2} = (\nabla G_k(\bar{x})^T u_\nu + \frac{o(\|t_\nu u_\nu\|)}{t_\nu})(\nabla H_k(\bar{x})^T u_\nu + \frac{o(\|t_\nu u_\nu\|)}{t_\nu}).$$

Thus, (5.3.4) holds. This completes the proof. $\square$

Theorem 5.3.1 *Let U be given by (5.3.3). Then, U is an* S-type *penalty term at $\bar{x}$, and for any $0 \le p < 1$, the following condition are equivalent:*

(i) $[\ker dU^p(\bar{x})]^* = T^{\text{lin}}(\bar{x})^*$.

(ii) $\widehat{\partial} U^p(\bar{x}) = T^{\text{lin}}(\bar{x})^*$.

(iii) *U^p is an* S-type *penalty term at $\bar{x}$.*

Moreover, U is an M-type *penalty term at $\bar{x}$, and U^p with $0 \le p < 1$ is an* M-type *penalty term at $\bar{x}$ if one of the following equivalent conditions is satisfied:*

(i') $[\ker dU^p(\bar{x})]^* = T^{\text{lin}}_{\text{MPCC}}(\bar{x})^*$.

(ii′) $\operatorname{pos}(\widehat{\partial} U^p(\bar{x})) = T^{\text{lin}}_{\text{MPCC}}(\bar{x})^*$.

In particular, condition (ii′) can be refined as $\widehat{\partial} U^p(\bar{x}) = T^{\text{lin}}_{\text{MPCC}}(\bar{x})^*$ *when* $0 \leq p \leq 0.5$.

Proof. In view of Theorem 5.2.1, Remark 5.2.2, and Lemma 5.3.1, we can show the results in a similar way as in the proof of Theorem 4.2.2. □

In terms of the original data of (MPCC), we give sufficient conditions for U^p to be an S-type (M-type) penalty term at $\bar{x}$.

Proposition 5.3.1 *The following statements are true:*

(*a*) *If the functions* $g_i, i \in I(\bar{x})$, $h_j, j \in J$, G_k *and* H_k, $k \in K$ *are* $\mathcal{C}^{1,1}$, *then*

$$\ker dU^p(\bar{x}) = T^{\text{lin}}(\bar{x}) \qquad \forall p \in (0.5, 1],$$

which implies that U^p *with* $0.5 < p \leq 1$ *is an* S-type *penalty term at* $\bar{x}$.

(*b*) *If the functions* g_i, $i \in I(\bar{x})$, h_j, $j \in J$, G_k *and* H_k, $k \in K$ *are twice continuously differentiable, and the following conditions are satisfied:*

(i) $u^T \nabla^2 g_i(\bar{x}) u \leq 0 \quad \forall i \in I(\bar{x}), \forall u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \cap \nabla g_i(\bar{x})^\perp$,

(ii) $u^T \nabla^2 h_j(\bar{x}) u = 0 \quad \forall j \in J, \forall u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x})$,

(iii) $u^T \nabla^2 G_k(\bar{x}) u = 0 \quad \forall k \in \alpha, \forall u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x})$,

(vi) $u^T \nabla^2 H_k(\bar{x}) u = 0 \quad \forall k \in \gamma, \forall u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x})$,

(v) $u^T \nabla^2 G_k(\bar{x}) u \geq 0 \quad \forall k \in \beta, \forall u \in (T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \cap \nabla G_k(\bar{x})^\perp) \backslash \nabla H_k(\bar{x})^\perp$,

(vi) $u^T \nabla^2 H_k(\bar{x}) u \geq 0 \quad \forall k \in \beta, \forall u \in (T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \cap \nabla H_k(\bar{x})^\perp) \backslash \nabla G_k(\bar{x})^\perp$,

(vii) $u^T \nabla^2 G_k(\bar{x}) u \geq 0$ *and* $u^T \nabla^2 H_k(\bar{x}) u \geq 0 \quad \forall k \in \beta$,
$\forall u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \bigcap \nabla G_k(\bar{x})^\perp \cap \nabla H_k(\bar{x})^\perp$,

then

$$\ker dU^{0.5}(\bar{x}) = T^{\text{lin}}_{\text{MPCC}}(\bar{x}),$$

which implies that U^p *with* $p = 0.5$ *is an* M-type *penalty term at* $\bar{x}$.

Proof. Since U is defined in the same way as S in the sense that the complementarity constraints $G_k(x) \geq 0$, $H_k(x) \geq 0$, and $G_k(x)H_k(x) = 0$ are treated as general inequality and equality constraints, statement (a) follows immediately from Proposition 4.2.2 (i).

Now we show that statement (b) is true. In view of Lemma 5.3.1, it suffices to show

$$T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \subset \ker dU^{0.5}(\bar{x}).$$

To start with, it is easy to check that, for each $0 < p \leq 1$,

$$\left\{ u \in R^n \;\middle|\; \begin{array}{ll} \limsup\limits_{t\to 0+} \dfrac{g_i(\bar{x}+tu)}{t^{1/p}} \leq 0, & i \in I(\bar{x}), \\ \lim\limits_{t\to 0+} \dfrac{h_j(\bar{x}+tu)}{t^{1/p}} = 0, & j \in J, \\ \lim\limits_{t\to 0+} \dfrac{G_k(\bar{x}+tu)}{t^{1/p}} = 0, & k \in \alpha, \\ \lim\limits_{t\to 0+} \dfrac{H_k(\bar{x}+tu)}{t^{1/p}} = 0, & k \in \gamma, \\ \liminf\limits_{t\to 0+} \dfrac{G_k(\bar{x}+tu)}{t^{1/p}} \geq 0, & k \in \beta, \\ \liminf\limits_{t\to 0+} \dfrac{H_k(\bar{x}+tu)}{t^{1/p}} \geq 0, & k \in \beta, \\ \lim\limits_{t\to 0+} \dfrac{G_k(\bar{x}+tu)H_k(\bar{x}+tu)}{t^{1/p}} = 0, & k \in \beta. \end{array} \right\} \subset \ker dU^{0.5}(\bar{x}). \tag{5.3.5}$$

Let $\varphi, \psi : R^n \to R$ be twice continuously differentiable functions. If $\varphi(x) = 0$ and $\nabla\varphi(x)^T u < 0$, we have

$$\frac{\varphi(x+tu)}{t^2} = \frac{\nabla\varphi(x)^T u}{t} + \frac{1}{2}u^T\nabla^2\varphi(x)u + \frac{o(t^2)}{t^2} \to -\infty \text{ as } t \to 0+\,. \tag{5.3.6}$$

If $\varphi(x) = 0$ and $\nabla\varphi(x)^T u = 0$, we have

$$\frac{\varphi(x+tu)}{t^2} = \frac{1}{2}u^T\nabla^2\varphi(x)u + \frac{o(t^2)}{t^2} \to \frac{1}{2}u^T\nabla^2\varphi(x)u \;\text{ as } t \to 0+\,. \tag{5.3.7}$$

If $\varphi(x) = \psi(x) = 0$ and $\nabla\varphi(x)^T u = 0$, we have

$$\frac{\varphi(x+tu)\psi(x+tu)}{t^2} = (\frac{1}{2}u^T\nabla^2\varphi(x)u + \frac{o(t^2)}{t^2})\psi(x+tu) \to 0 \text{ as } t \to 0+\,. \tag{5.3.8}$$

In view of the limits in (5.3.6), (5.3.7) and (5.3.8), we can easily show that the set $T^{\text{lin}}_{\text{MPCC}}(\bar{x})$ is a subset of the left-hand said of (5.3.5) with $p = 0.5$. This completes the proof. □

Recall that in defining $\mathcal{H}_p$, we use the function

$$V(x) := S(x) + \sum_{k\in K} |\phi_{\min}(G_k(x), H_k(x))| \qquad \forall x \in R^n. \tag{5.3.9}$$

where $\phi_{\min}(a,b) = \min\{a,b\}$ and S is given by (4.1.2).

Lemma 5.3.2 $\ker dV(\bar{x}) = T^{\rm lin}_{\rm MPCC}(\bar{x})$.

Proof. Let $u \in \ker dV(\bar{x})$. By Proposition 3.2.1 (i), there exist $t_\nu \to 0+$ and $u_\nu \to u$ such that

$$\frac{\max\{g_i(\bar{x}+t_\nu u_\nu), 0\}}{t_\nu} \to 0 \qquad \forall i \in I(\bar{x}), \tag{5.3.10}$$

$$\frac{h_j(\bar{x}+t_\nu u_\nu)}{t_\nu} \to 0 \qquad \forall j \in J, \tag{5.3.11}$$

$$\frac{G_k(\bar{x}+t_\nu u_\nu)}{t_\nu} \to 0 \qquad \forall k \in \alpha, \tag{5.3.12}$$

$$\frac{H_k(\bar{x}+t_\nu u_\nu)}{t_\nu} \to 0 \qquad \forall k \in \gamma, \tag{5.3.13}$$

and

$$\frac{\min\{G_k(\bar{x}+t_\nu u_\nu), H_k(\bar{x}+t_\nu u_\nu)\}}{t_\nu} \to 0 \qquad \forall k \in \beta. \tag{5.3.14}$$

By applying the Taylor expansion rule to (5.3.10), (5.3.11), (5.3.12), (5.3.13) and (5.3.14) and noticing that $\min\{\nabla G_k(\bar{x})^T u, \nabla H_k(\bar{x})^T u\} = 0$ if and only if

$$\nabla G_k(\bar{x})^T u \geq 0, \nabla H_k(\bar{x})^T u \geq 0, (G_k(\bar{x})^T u)(H_k(\bar{x})^T u) = 0,$$

we can easily get $\ker dV(\bar{x}) \subset T^{\rm lin}_{\rm MPCC}(\bar{x})$. Let $t_\nu \to 0+$ and let $u_\nu \equiv u$ with $u \in T^{\rm lin}_{\rm MPCC}(\bar{x})$. By applying again the Taylor expansion rule, we get (5.3.10), (5.3.11), (5.3.12), (5.3.13) and (5.3.14). This implies that $T^{\rm lin}_{\rm MPCC}(\bar{x}) \subset \ker dV(\bar{x})$. This completes the proof. $\square$

Theorem 5.3.2 *Let V be given by (5.3.9). Then V is an* M-type *penalty term at $\bar{x}$, and V^p with $0 \leq p < 1$ is an* M-type *penalty term at $\bar{x}$ if one of the following equivalent conditions is satisfied:*

(i) $[\ker dV^p(\bar{x})]^* = T^{\rm lin}_{\rm MPCC}(\bar{x})^*$.

(ii) $\widehat{\partial} V^p(\bar{x}) = T^{\rm lin}_{\rm MPCC}(\bar{x})^*$.

Proof. In view of Theorem 5.2.1, Remark 5.2.2, and Lemma 5.3.2, we can show the results in a similar way as in the proof of Theorem 4.2.2. $\square$

In terms of the original data of (MPCC), we give sufficient conditions for V^p to be of M-type at $\bar{x}$.

Proposition 5.3.2 *The following statements are true:*

(a) *If the functions* $g_i, i \in I(\bar{x})$, $h_j, j \in J$, G_k *and* H_k, $k \in K$ *are* $\mathcal{C}^{1,1}$, *then*

$$\ker dV^p(\bar{x}) = T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \qquad \forall p \in (0.5, 1],$$

which implies that V^p *with* $0.5 < p \leq 1$ *is an* M-type *penalty term at* $\bar{x}$.

(b) *If the functions* g_i, $i \in I(\bar{x})$, h_j, $j \in J$, G_k *and* H_k, $k \in K$ *are twice continuously differentiable, and conditions* $(i) - (iv)$ *in Proposition 5.3.1 and the following conditions are satisfied:*

(v') $u^T\nabla^2 G_k(\bar{x})u = 0 \quad \forall k \in \beta, \forall u \in (T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \cap \nabla G_k(\bar{x})^\perp) \backslash \nabla H_k(\bar{x})^\perp$,

(vi') $u^T\nabla^2 H_k(\bar{x})u = 0 \quad \forall k \in \beta, \forall u \in (T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \cap \nabla H_k(\bar{x})^\perp) \backslash \nabla G_k(\bar{x})^\perp$,

(vii') $\phi_{\min}(u^T\nabla^2 G_k(\bar{x})u, u^T\nabla^2 H_k(\bar{x})u) = 0 \quad \forall u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \cap \nabla G_k(\bar{x})^\perp \cap \nabla H_k(\bar{x})^\perp$, $\forall k \in \beta$,

then

$$\ker dV^{0.5}(\bar{x}) = T^{\text{lin}}_{\text{MPCC}}(\bar{x}),$$

which implies that $V^{0.5}$ *is an* M-type *penalty term at* $\bar{x}$.

Proof. To start with, it is easy to check that for every $0 < p \leq 1$,

$$\left\{ u \in R^n \;\middle|\; \begin{array}{ll} \limsup\limits_{t \to 0+} \dfrac{g_i(\bar{x}+tu)}{t^{1/p}} \leq 0, & i \in I(\bar{x}) \\ \lim\limits_{t \to 0+} \dfrac{h_j(\bar{x}+tu)}{t^{1/p}} = 0, & j \in J \\ \lim\limits_{t \to 0+} \dfrac{G_k(\bar{x}+tu)}{t^{1/p}} = 0, & k \in \alpha \\ \lim\limits_{t \to 0+} \dfrac{H_k(\bar{x}+tu)}{t^{1/p}} = 0, & k \in \gamma \\ \lim\limits_{t \to 0+} \dfrac{\min\{G_k(\bar{x}+tu), H_k(\bar{x}+tu)\}}{t^{1/p}} = 0, & k \in \beta \end{array} \right\} \subset \ker dV^p(\bar{x}). \tag{5.3.15}$$

First we show that statement (a) is true. Let $0.5 < p \leq 1$ and let $k \in \beta$. In view of Proposition 3.2.1 (iii) and Lemma 5.3.2, it suffices to show $T^{\text{lin}}_{\text{MPCC}}(\bar{x}) \subset \ker dV^p(\bar{x})$. This can be done by showing that $T^{\text{lin}}_{\text{MPCC}}(\bar{x})$ is a subset of the left-hand said of (5.3.15). Let $u \in T^{\text{lin}}_{\text{MPCC}}(\bar{x})$. Then $\min\{\nabla G_k(\bar{x})^T u, \nabla H_k(\bar{x})^T u\} = 0$. Following the proof of Proposition 4.2.2 (i), it now remains to show

$$\lim_{t \to 0+} \frac{\min\{G_k(\bar{x}+tu), H_k(\bar{x}+tu)\}}{t^{1/p}} = 0. \tag{5.3.16}$$

By a generalized Taylor expansion rule, see [38], we have

$$\begin{aligned} &\frac{\min\{G_k(\bar{x}+tu), H_k(\bar{x}+tu)\}}{t^{\frac{1}{p}}} \\ \leq\ & t^{2-\frac{1}{p}} \min\left\{ \frac{\nabla G_k(\bar{x})^T u}{t} + \frac{1}{2} G_k^{oo}(\bar{x}+t\theta u; u), \frac{\nabla H_k(\bar{x})^T u}{t} + \frac{1}{2} H_k^{oo}(\bar{x}+t\omega u; u) \right\}, \end{aligned} \tag{5.3.17}$$

where $0 < \theta < 1$ and $0 < \omega < 1$. Note that the functions $x \to G_k^{oo}(x;u)$ and $x \to H_k^{oo}(x;u)$ are upper semicontinuous with $G_k^{oo}(x;u)$ and $H_k^{oo}(x;u)$ being finite when G_k and H_k are $\mathcal{C}^{1,1}$. The min term on the right hand side of (5.3.17) must be finite for sufficiently small $t > 0$. Since $t^{2-\frac{1}{p}} \to 0$ as $t \to 0+$, it then follows from (5.3.17) that (5.3.16) holds.

Now we show that statement (b) is true. As in (a), it suffices to show that $T^{\text{lin}}_{\text{MPCC}}(\bar{x})$ is a subset of the left-hand said of (5.3.15) with $p = 0.5$. To that end, we need not only limits in (5.3.6), (5.3.7) and (5.3.8), but some more limits on twice continuously differentiable functions $\varphi : R^n \to R$ and $\psi : R^n \to R$. If $\varphi(x) = \psi(x) = 0$, $\nabla\varphi(x)^T u = 0$ and $\nabla\psi(x)^T u > 0$, we have

$$\begin{aligned} &\frac{\min\{\varphi(x+tu), \psi(x+tu)\}}{t^2} \\ &= \min\{\frac{1}{2}u^T\nabla^2\varphi(x)u + \frac{o(t^2)}{t^2}, \frac{\nabla\psi(x)^T u}{t} + \frac{1}{2}u^T\nabla^2\psi(x)u + \frac{o(t^2)}{t^2}\} \\ &\to \tfrac{1}{2}u^T\nabla^2\varphi(x)u \text{ as } t \to 0+. \end{aligned} \tag{5.3.18}$$

If $\varphi(x) = \psi(x) = 0$ and $\nabla\varphi(x)^T u = \nabla\psi(x)^T u = 0$, we have

$$\begin{aligned} &\frac{\min\{\varphi(x+tu), \psi(x+tu)\}}{t^2} \\ &= \min\{\frac{1}{2}u^T\nabla^2\varphi(x)u + \frac{o(t^2)}{t^2}, \frac{1}{2}u^T\nabla^2\psi(x)u + \frac{o(t^2)}{t^2}\} \\ &\to \min\{\frac{1}{2}u^T\nabla^2\varphi(x)u, \frac{1}{2}u^T\nabla^2\psi(x)u\} \text{ as } t \to 0+. \end{aligned} \tag{5.3.19}$$

In view of (5.3.15) with $p = 0.5$, and the limits in (5.3.6), (5.3.7), (5.3.8), (5.3.18), (5.3.19), we can easily show that the set $T^{\text{lin}}_{\text{MPCC}}(\bar{x})$ is a subset of the left-hand said of (5.3.15) with $p = 0.5$. This completes the proof. □

The following lemma and proposition are helpful for establishing some relationships between exactness of $\mathcal{G}_p$ and $\mathcal{H}_p$, and between $\ker dU^p(\bar{x})$ and $\ker dV^p(\bar{x})$.

Lemma 5.3.3 *Let $a, b \in R$, $\sigma_1 \geq \max\{a+1, b+1, 2-a, 2-b\}$ and $\sigma_2 \geq \max\{\sqrt{|a|}, \sqrt{|b|}, 1\}$. Then,*

$$\frac{(-a)_+ + (-b)_+ + |ab|}{\sigma_1} \leq |\min\{a,b\}| \leq \sigma_2 \sqrt{(-a)_+ + (-b)_+ + |ab|}. \tag{5.3.20}$$

Proof. Noting that a and b are symmetrical, we only need to consider three cases: (i) $0 \leq b \leq a$, (ii) $b < 0 < a$, (iii) $b \leq a \leq 0$. For case (i), we have

$$(-a)_+ + (-b)_+ + |ab| = ab \leq (a+1)b \leq \sigma_1 b = \sigma_1 |\min\{a,b\}|,$$

and

$$|\min\{a,b\}| = b = \sqrt{b^2} \leq \sqrt{ab} \leq \sigma_2 \sqrt{(-a)_+ + (-b)_+ + |ab|}.$$

For case (ii), we have

$$(-a)_+ + (-b)_+ + |ab| = -b(a+1) = (a+1)|\min\{a,b\}| \leq \sigma_1 |\min\{a,b\}|,$$

and

$$|\min\{a,b\}| = -b \leq -b\sqrt{a+1} = \sqrt{-b}\sqrt{-b(a+1)} \leq \sigma_2 \sqrt{(-a)_+ + (-b)_+ + |ab|}.$$

For case (iii), we have

$$(-a)_+ + (-b)_+ + |ab| = -a - b + ab \leq -2b + ab = (2-a)|\min\{a,b\}| \leq \sigma_1 |\min\{a,b\}|,$$

and

$$|\min\{a,b\}| = -b \leq -b\sqrt{-a+1} \leq \sqrt{-b}\sqrt{-a-b+ab} \leq \sigma_2 \sqrt{(-a)_+ + (-b)_+ + |ab|}.$$

This completes the proof. $\square$

Proposition 5.3.3 *Let $\delta > 0$ and let $y \in R^n$. Then, there exist $\theta > 0$ and $\eta > 0$ such that*

$$\frac{1}{\theta} U(x) \leq V(x) \leq \eta \sqrt{U(x)} \qquad \forall\, x \in B_\delta(y), \tag{5.3.21}$$

which implies that for any $0 \leq p \leq 1$,

$$\ker dU^{\frac{p}{2}}(\bar{x}) \subset \ker dV^p(\bar{x}) \subset \ker dU^p(\bar{x}). \tag{5.3.22}$$

Proof. We only need to show (5.3.21) since (5.3.22) follows from (5.3.21) readily. Let $\theta = \max_{k \in K} \theta_k$ where

$$\theta_k = \max_{x \in B_\delta(y)} \max \left\{ G_k(x) + 1, H_k(x) + 1, 2 - G_k(x), 2 - H_k(x) \right\}.$$

Clearly, $\theta \geq \frac{3}{2}$. By Lemma 5.3.3 and the definitions of U and V, we have

$$U(x) \leq \theta V(x), \quad \forall\, x \in B_\delta(y),$$

which gives the first inequality in (5.3.21). Now, let $\eta = (|K| + 1)\tilde{\eta}$ where $\tilde{\eta} = \max\{\max_{k \in K} \eta_k, S_{\max}\}$, $\eta_k = \max_{x \in B_\delta(y)} \max\{\sqrt{|G_k(x)|}, \sqrt{|H_k(x)|}, 1\}$, and $S_{\max} = \max_{x \in B_\delta(y)} \sqrt{S(x)}$. By Lemma 5.3.3 and the definition of η_k, we have, for each $x \in B_\delta(y)$ and each $k \in K$,

$$|\min\{G_k(x), H_k(x)\}| \leq \eta_k \sqrt{(-G_k(x))_+ + (-H_k(x))_+ + |G_k(x) H_k(x)|}. \tag{5.3.23}$$

By the definition of $S_{\max}$, we have, for each $x \in B_\delta(y)$

$$S(x) \leq S_{\max} \sqrt{S(x)}. \tag{5.3.24}$$

Then, it follows from (5.3.23), (5.3.24) and the definitions of U and V that, for each $x \in B_\delta(y)$,

$$\begin{aligned} V(x) &\leq \tilde{\eta} \left\{ \sqrt{S(x)} + \sum_{k \in K} \sqrt{(-G_k(x))_+ + (-H_k(x))_+ + |G_k(x) H_k(x)|} \right\} \\ &\leq (|K| + 1)\tilde{\eta} \sqrt{S(x) + \sum_{k \in K} [(-G_k(x))_+ + (-H_k(x))_+ + |G_k(x) H_k(x)|]} \\ &= \eta \sqrt{U(x)}, \end{aligned}$$

where the second inequality follows from Lemma 4.1 in [80]. Therefore, we have shown that (5.3.21) holds. This completes the proof. □

Remark 5.3.1 *Let $0 \leq p \leq 1$. It follows from (5.3.21) that, $\mathcal{H}_p$ is exact at $\bar{x}$ if $\mathcal{G}_p$ is exact at $\bar{x}$, and $\mathcal{G}_{\frac{p}{2}}$ is exact at $\bar{x}$ if $\mathcal{H}_p$ is exact at $\bar{x}$.*

Chapter 6

Conclusion and Future Work

In this thesis, by means of modern variational analysis, we developed a unified framework and provided a detailed exposition of optimality conditions from the viewpoint of exact penalty functions. We studied sufficient conditions for penalty terms to possess local error bounds which guarantee exactness of penalty functions. By using subderivatives, second-order subderivatives, and parabolic subderivatives, we also studied first- and second-order necessary and sufficient conditions for penalty functions to be exact. The kernels of these derivatives, representing directions at which derivatives vanish, played an key role in our investigation. In particular, we showed an interesting auxiliary result which asserts that, the polar cone of the subderivative kernel of an extended real-valued function at a local minimum is the same as the positive hull of its regular subgradients at the same point. We showed how KKT conditions and second-order necessary conditions in nonlinear programming, and strong and Mordukhovich stationarities in mathematical programs with complementarity constraints, can be derived from exactness of penalty functions under some additional conditions on constraint functions. In presenting these additional conditions, it turned out that the kernels of (parabolic) subderivatives of penalty terms are very crucial. By virtue of these kernels and a variational description of regular subgradients, we showed necessity and sufficiency of these additional conditions. We also presented conditions in terms of the original data by applying (generalized) Taylor expansions to calculate these kernels.

Beyond these positive results and contributions, there are many other issues that are needed to deal with in the future work. We summarize three directions for future work as follows. (*i*) Avakov et al. [9] obtained new first-order necessary conditions (unlike KKT conditions) for (NLP) by introducing a generalized Lagrangian function and using the theory of 2-Regularity. It is still unknown at the current stage whether these new first-order necessary conditions can be derived from the viewpoint of the $l_{\frac{1}{2}}$ exact penalty function. (*ii*) It is an open question as to whether our results are able to find applications in the design of numerical methods, and this is definitely a challenging research topic in the future. (*iii*) Most of our results do not rely on smooth data, but some do. For those results that do depend on smooth data, is it possible to relax the differentiability assumptions? This could be an interesting research topic in the future.

Bibliography

[1] J. Abadie. On the Kuhn-Tucker theorem. In J. Abadie, editor, *Nonlinear programming*, pages 21–36. Amsterdam: North-Holland Publishing Co., 1967.

[2] C. M. Ablow and G. Brigham. An analog solution of programming problems. *Journal of the Operations Research Society of America*, 3(4):pp. 388–394, 1955.

[3] R. Andreani, C. Echagüe, and M. Schuverdt. Constant-rank condition and second-order constraint qualification. *Journal of Optimization Theory and Applications*, 146:255–266, 2010.

[4] R. Andreani, J. M. Martínez, and M. L. Schuverdt. On second-order optimality conditions for nonlinear programming. *Optimization*, 56(5-6):529–542, 2007.

[5] M. Anitescu. Degenerate nonlinear programming with a quadratic growth condition. *SIAM Journal on Optimization*, 10:1116–1135, 1999.

[6] A. V. Arutyunov. Perturbations of extremal problems with constraints and necessary optimality conditions. *Journal of Mathematical Sciences*, 54:1342–1400, 1991.

[7] J.-P. Aubin. Lipschitz behavior of solutions to convex minimization problems. *Mathematics of Operations Research*, 9(1):87–111, 1984.

[8] J.-P. Aubin and L. Ekeland. *Applied Nonlinear Analysis*. John Wieley, New York, 1984.

[9] E. Avakov, A. Arutyunov, and A. Izmailov. Necessary conditions for an extremum in a mathematical programming problem. *Proceedings of the Steklov Institute of Mathematics*, 256:2–25, 2007.

[10] D. Azé. A survey on error bounds for lower semicontinuous functions. *ESAIM Proceedings.*, 13:1–17, 2003.

[11] D. Azé and J.-N. Corvellec. Characterizations of error bounds for lower semicontinuous functions on metric spaces. *ESAIM: Control, Optimisation and Calculus of Variations*, 10:409–425, 2004.

[12] A. Baccari. On the classical necessary second-order optimality conditions. *Journal of Optimization Theory and Applications*, 123:213–221, 2004.

[13] A. Baccari and A. Trad. On the classical necessary second-order optimality conditions in the presence of equality and inequality constraints. *SIAM Journal on Optimization*, 15:394–408, 2005.

[14] M. S. Bazaraa and J. J. Goode. Sufficient conditions for a globally exact penalty function without convexity. In *Optimality and Stability in Mathematical Programming*, volume 19 of *Mathematical Programming Studies*, pages 1–15. Springer Berlin Heidelberg, 1982.

[15] M. S. Bazaraa, J. J. Goode, and C. M. Shetty. Constraint qualifications revisited. *Management Science*, 18(9):567–573, 1972.

[16] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming*. Wiley, New York, 1993.

[17] M. S. Bazaraa and C.M. Shetty. *Foundations of Optimization*. Springer-Verlag, Berlin, Heidelberg, New York, 1976.

[18] A. Ben-Israel, A. Ben-Tal, and S. Zlobec. *Optimality in Nonlinear Programming : A Feasible Directions Approach*. Wiley, New York, 1981.

[19] A. Ben-Tal. Second-order and related extremality conditions in nonlinear programming. *Journal of Optimization Theory and Applications*, 31(2):143–165, 1980.

[20] A. Ben-Tal and J. Zowe. A unified theory of first and second order conditions for extremum problems in topological vector spaces. *Mathematical Programming Study*, 19:39–76, 1982.

[21] D.P. Bertsekas. *Constrained Optimization and Lagrangian Multiplier Methods.* Academic Press, New York, 1982.

[22] D.P. Bertsekas and A.E. Ozdaglar. Pseudonormality and a Lagrange multiplier theory for constrained optimization. *Journal of Optimization Theory and Applications*, 114:287–343(57), 2002.

[23] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems.* Springer-Verlag, New York, 2000.

[24] J. Burke. Second order necessary and sufficient conditions for convex composite NDO. *Mathematical Programming*, 38:287–302, 1987.

[25] J. V. Burke. Calmness and exact penalization. *SIAM Journal on Control and Optimization*, 29(2):493–497, 1991.

[26] J. V. Burke. An exact penalization viewpoint of constrained optimization. *SIAM Journal on Control and Optimization*, 29(4):968–998, 1991.

[27] J. V. Burke and S. Deng. Weak sharp minima revisited Part I: basic theory. *Control and Cybernetics*, 31(3):439–469, 2002.

[28] J. V. Burke and S. Deng. Weak sharp minima revisited, Part II: application to linear regularity and error bounds. *Mathematical Programming*, 104(2-3):235–261, 2005.

[29] J. V. Burke and S. Deng. Weak sharp minima revisited, Part III: error bounds for differentiable convex inclusions. *Mathematical Programming*, 116(1-2):37–56, 2009.

[30] J. V. Burke and M. C. Ferris. Weak sharp minima in mathematical programming. *SIAM Journal on Control and Optimization*, 31(5):1340–1359, 1993.

[31] J. V. Burke and R. A. Poliquin. Optimality conditions for non-finite valued convex composite functions. *Mathematical Programming*, 57:103–120, 1992.

[32] C. W. Carroll. The created response surface technique for optimizing nonlinear, restrained systems. *Operations Research*, 9(2):169–184, 1961.

[33] C. Charalambous. A lower bound for the controlling parameters of the exact penalty functions. *Mathematical Programming*, 15:278–290, 1978.

[34] Y. Chen and M. Florian. The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions. *Optimization*, 32:193–209, 1995.

[35] F. H. Clarke. A new approach to Lagrange multipliers. *Mathematics of Operations Research*, 1(2):165–174, 1976.

[36] F. H. Clarke. *Optimzation and Nonsmooth Analysis*. John Wieley, New York, 1983.

[37] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1998.

[38] R. Cominetti and R. Correa. A generalized second-order derivative in nonsmooth optimization. *SIAM Journal on Control and Optimization*, 28(4):789–809, 1990.

[39] G. Contaldi, G. Di Pillo, and S. Lucidi. A continuously differentiable exact penalty function for nonlinear programming problems with unbounded feasible set. *Operations Research Letters*, 14(3):153 – 161, 1993.

[40] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49, 1943.

[41] S. Deng. Computable error bounds for convex inequality systems in reflexive Banach spaces. *SIAM Journal on Optimization*, 7(1):274–279, 1997.

[42] S. Deng. Global error bounds for convex inequality systems in Banach spaces. *SIAM Journal on Control and Optimization*, 36(4):1240–1249, 1998.

[43] S. Deng. Perturbation analysis of a condition number for convex inequality systems and global error bounds for analytic systems. *Mathematical Programming*, 83(2):263–276, 1998.

[44] S. Deng and X. Q. Yang. Weak sharp minima in multicriteria linear programming. *SIAM Journal on Optimization*, 15(2):456–460, 2004.

[45] S. Di and R. Poliquin. Calmness in optimization problems with equality constraints. *Journal of Mathematical Analysis and Applications*, 189(2):502 – 513, 1995.

[46] G. Di Pillo and L. Grippo. Exact penalty functions in constrained optimization. *SIAM Journal on Control and Optimization*, 27:1333–1360, 1989.

[47] A. L. Dontchev and R. T. Rockafellar. Regularity and conditioning of solution mappings in variational analysis. *Set-Valued Analysis*, 12(1):79–109, 2004.

[48] I. I. Eremin. The penalty method in convex programming. *Cybernetics and Systems Analysis*, 3(4):53–56, 1967.

[49] M. C. Ferris. *Weak Sharp Minima and Penalty Functions in Mathematical Programming*. PhD thesis, University of Cambridge, Cambridge, 1988.

[50] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Wiley, New York., 1968.

[51] A. Fischer and H. Y. Jiang. Merit functions for complementarity and related problems: A survey. *Computational Optimization and Applications*, 17(2):159–182, 2000.

[52] M. L. Flegel. *Constraint Qualifications and Stationarity Concepts for Mathematical Programs with Equilibrium Constraints*. PhD thesis, University of Würzburg, 2005.

[53] M. L. Flegel and C. Kanzow. A Fritz John approach to first order optimality conditions for mathematical programs with equilibrium constraints. *Optimization*, 52(3):277–286, 2003.

[54] M. L. Flegel and C. Kanzow. Abadie-type constraint qualification for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 124(3):595–614, 2005.

[55] M. L. Flegel and C. Kanzow. On M-stationary points for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 310(1):286–302, 2005.

[56] M. L. Flegel and C. Kanzow. On the Guignard constraint qualification for mathematical programs with equilibrium constraints. *Optimization*, 54(6):517–534, 2005.

[57] M. L. Flegel and C. Kanzow. A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints. *Optimization with Multivalued Mappings*, pages 111–122, 2006.

[58] M. L. Flegel, C. Kanzow, and J. V. Outrata. Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints. *Set-Valued Analysis*, 15(2):139–162, 2007.

[59] R. Fletcher. An exact penalty function for nonlinear programming with inequalities. *Mathematical Programming*, 5:129–150, 1973.

[60] R. Fletcher. *Practical Methods of Optimization*. Wiley-Interscience, New York, NY, USA, 1987.

[61] R. Fletcher and S. Leyffer. Solving mathematical programs with complementarity constraints as nonlinear programs. *Optimization Methods and Software*, 19(1):15 – 40, 2004.

[62] J. Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, 12:136–138, 1977.

[63] E. De Giorgi, A. Marino, and M. Tosques. Problemi di evoluzione in spazi metrici e curve di massima pendenza (evolution problems in metric spaces and curves of maximal slope). *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.*, 68:180–187, 1980.

[64] F. J. Gould and Jon W. Tolle. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, 20(2):164–172, 1971.

[65] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. *SIAM Journal on Control*, 7(2):232–241, 1969.

[66] S.-P. Han and O. L. Mangasarian. Exact penalty functions in nonlinear programming. *Mathematical Programming*, 17(1):251–269, 1979.

[67] S.-P. Han and O. L. Mangasarian. A dual differentiable exact penalty function. *Mathematical Programming*, 25:293–306, 1983.

[68] Y. R. He and J. Sun. A new constraint qualification and a second-order necessary optimality condition for mathematical programming problems. *Pacific Journal of Optimization*, 7(1):9–28, 2011.

[69] R. Henrion and A. Jourani. Subdifferential conditions for calmness of convex constraints. *SIAM Journal on Optimization on Optimization*, 13(2):520–534, 2002.

[70] R. Henrion, A. Jourani, and J. Outrata. On the calmness of a class of multifunctions. *SIAM Journal on Optimization*, 13(2):603–618, 2002.

[71] R. Henrion and J. Outrata. A subdifferential condition for calmness of multifunctions. *Journal of Mathematical Analysis and Applications* , 258(1):110–130, 2001.

[72] R. Henrion and J. Outrata. A subdifferential condition for calmness of multifunctions. *Journal of Mathematical Analysis and Applications* , 258(1):110–130, 2001.

[73] R. Henrion and J. Outrata. Calmness of constraint systems with applications. *Mathematical Programming*, 104(2):437–464, 2005.

[74] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

[75] M. R. Hestenes. *Optimization Theory: The Finite-Dimensional Case*. John Wiley and Sons, New York, NY, 1975.

[76] S. Howe. New conditions for exactness of a simple penalty function. *SIAM Journal on Control*, 11(2):378–381, 1973.

[77] X. M. Hu and D. Ralph. A note on sensitivity of value functions of mathematical programs with complementarity constraints. *Mathematical Programming*, 93(2):265–279, 2002.

[78] X. M. Hu and D. Ralph. Convergence of a penalty method for mathematical programming with complementarity constraints. *Journal of Optimization Theory and Applications*, 123(2):365–390, 2004.

[79] L. R. Huang and K. F. Ng. On first- and second-order conditions for error bounds. *SIAM Journal on Optimization*, 14(4):1057–1073, 2004.

[80] X. X. Huang and X. Q. Yang. A unified augmented Lagrangian approach to duality and exact penalization. *Mathmatics of Operations Research*, 28(3):533–552, 2003.

[81] X. X. Huang, X. Q. Yang, and D. L. Zhu. A sequential smooth penalization approach to mathematical programs with complementarity constraints. *Numerical Functional Analysis and Optimization*, 27(1):71–98, 2006.

[82] A. D. Ioffe. Necessary and sufficient conditions for a local minimum. 1: A reduction theorem and first order conditions. *SIAM Journal on Control and Optimization*, 17(2):245–250, 1979.

[83] A. D. Ioffe. Necessary and sufficient conditions for a local minimum. 3: Second order conditions and augmented duality. *SIAM Journal on Control and Optimization*, 17(2):266–288, 1979.

[84] A. D. Ioffe. Regular points of Lipschitz functions. *Transactions of the American Mathematical Society*, 251:61–69, 1979.

[85] A. D. Ioffe. Metric regularity and subdifferential calculus. *Russian Mathematical Surveys*, 55(3):501–558, 2000.

[86] A. D. Ioffe and J. V. Outrata. On metric and calmness qualification conditions in subdifferential calculus. *Set-Valued Analysis*, 16(2):199–227, 2008.

[87] R. Janin. Directional derivative of the marginal function in nonlinear programming. In *Sensitivity, Stability and Parametric Analysis*, volume 21 of *Mathematical Programming Studies*, pages 110–126. Springer Berlin Heidelberg, 1984.

[88] V. Jeyakumar and X. Q. Yang. Convex composite minimization with $\mathcal{C}^{1,1}$ functions. *Journal of Optimization Theory and Applications*, 86:631–648, 1995.

[89] H. Y. Jiang and D. Ralph. Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. *SIAM Journal on Optimization*, 10(3):779–808, 1999.

[90] F. John. Extremum Problems with Inequalities as Subsidiary Conditions. In K. O. Friedrichs, O. E. Neugebauer, and J. J. Stoker, editors, *Studies and*

Essays: Courant Anniversary Volume, pages 187–204. Wiley-Interscience, New York, 1948.

[91] A. Jourani. Hoffman's error bound, local controllability, and sensitivity analysis. *SIAM Journal on Control and Optimization*, 38(3):947–970, 2000.

[92] C. Kanzow and A. Schwartz. Mathematical programs with equilibrium constraints: Enhanced Fritz John-conditions, new constraint qualifications, and improved exact penalty results. *SIAM Journal on Optimization*, 20(5):2730–2753, 2010.

[93] W. Karush. Minima of functions of several variables with inequalities as side constraints. Master's thesis, Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois, 1939.

[94] H. Kawasaki. Second-order necessary conditions of the Kuhn-Tucker type under new constraint qualifications. *Journal of Optimization Theory and Applications*, 57(2):253–264, 1988.

[95] D. Klatte and B. Kummer. Constrained minima and Lipschitzian penalties in metric spaces. *SIAM Journal on Optimization*, 13(2):619–633, 2002.

[96] H. W. Kuhn and A. W. Tucker. Nonlinear programming. *Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press*, pages 481–492, 1951.

[97] J.B. Lasserre. Exact penalty functions and Lagrange multipliers. *RAIRO Automat./Systems Anal. Control*, 14, 1980.

[98] A. S. Lewis and J. S. Pang. Error bounds for convex inequality systems. In Crouzeix, J. P. and MartinezLegaz, J. E. and Volle, M., editor, *Generalized Convexity, Generalized Monotonicity: Recent Results*, volume 27 of *Nonconvex Optimization and Its Applications*, pages 75–110, 1998.

[99] S. J. Li, K. W. Meng, and J.-P. Penot. Calculus rules for derivatives of multimaps. *Set-Valued and Variational Analysis*, 17(1):21–39, 2009.

[100] W. Li. Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities. *SIAM Journal on Optimization*, 7(4):966–978, 1997.

[101] W. Li and I. Singer. Global error bounds for convex multifunctions and applications. *Mathmatics of Operations Research*, 23(2):443–462, 1998.

[102] G. S. Liu, J. J. Ye, and J. P. Zhu. Partial exact penalty for mathematical programs with equilibrium constraints. *Set-Valued Analysis*, 16(5):785–804, 2008.

[103] S. Lu. Implications of the constant rank constraint qualification. *Mathematical Programming*, 126:365–392, 2011.

[104] Z. Q. Luo and J. S. Pang. Error bounds for analytic systems and their applications. *Mathematical Programming*, 67:1–28, 1994.

[105] Z. Q. Luo, J. S. Pang, and D. Ralph. *Mathematical Pprograms with Equilibrium Constraints*. Cambridge University Press, Cambridge, 1996.

[106] Z. Q. Luo, J. S. Pang, D. Ralph, and S. Q. Wu. Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints. *Mathematical Programming*, 75(1):19–76, 1996.

[107] O. L. Mangasarian. *Nonlinear Programming*. Mc Graw-Hill, New York, 1969.

[108] O. L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17(1):37–47, 1967.

[109] G. P. McCormick. Second order conditions for constrained minima. *SIAM Journal on Applied Mathematics*, 15(3):641–652, 1967.

[110] K. W. Meng and X. Q. Yang. Optimality conditions via penalty function. *SIAM Journal on Optimization*, 20(6):3208–3231, 2010.

[111] L. Minchenko and S. Stakhovski. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 2011(iFirst).

[112] L. Minchenko and A. Tarakanov. On error bounds for quasinormal programs. *Journal of Optimization Theory and Applications*, 148:571–579, 2011.

[113] B. S. Mordukhovich. Generalized differential calculus for nonsmooth and set-valued mappings. *Journal of Mathematical Analysis and Applications*, 183(1):250–288, 1994.

[114] K. F. Ng and W. H. Yang. Error bounds for some convex functions and distance composite functions. *SIAM Journal on Optimization*, 15(4):1042–1056, 2005.

[115] K. F. Ng and X. Y. Zheng. Global error bounds with fractional exponents. *Mathematical Programming*, 88(2):357–370, 2000.

[116] K. F. Ng and X. Y. Zheng. Error bounds for lower semicontinuous functions in normed spaces. *SIAM Journal on Optimization*, 12(1):1–17, 2001.

[117] K. F. Ng and X. Y. Zheng. Global weak sharp minima on Banach spaces. *SIAM Journal on Control and Optimization*, 41(6):1868–1885, 2003.

[118] H. V. Ngai and M. Théra. Error bounds for convex differentiable inequality systems in banach spaces. *Mathematical Programming*, 104:465–482, 2005.

[119] H. V. Ngai and M. Théra. Error bounds in metric spaces and application to the perturbation stability of metric regularity. *SIAM Journal on Optimization*, 19(1):1–20, 2008.

[120] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2006.

[121] J. V. Outrata. Optimality conditions for a class of mathematical programs with equilibrium constraints. *Mathematics of Operations Research*, 24(3):627–644, 1999.

[122] J. V. Outrata. On mathematical programs with complementarity constraints. *Optimization Methods and Software*, 14(1-2):117–137, 2000.

[123] J. V. Outrata, M. Kocvara, and J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraits*. Kluwer Academic Publishers, 1998.

[124] J. S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79(1):299–332, 1997.

[125] J. S. Pang and M. Fukushima. Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints. *Computational Optimization and Applications*, 13(1):111–136, 1999.

[126] J.-P. Penot. A new constraint qualification condition. *Journal of Optimization Theory and Applications*, 48:459–468, 1986.

[127] D. W. Peterson. A review of constraint qualifications in finite-dimensional spaces. *SIAM Review*, 15(3):639–654, 1973.

[128] T. Pietrzykowski. An exact potential method for constrained maxima. *SIAM Journal on Numerical Analysis*, 6(2):299–304, 1969.

[129] G. Di Pillo and L. Grippo. A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints. *SIAM Journal on Control and Optimization*, 23(1):72–84, 1985.

[130] M.J.D. Powell. A method for nonlinear constraints in minimization problems. In Optimization, editor, *R. Fletcher*, pages 283–298. Academic Press, New York, NY, 1969.

[131] L. Qi and Z. X. Wei. On the constant positive linear dependence condition and its application to SQP methods. *SIAM Journal on Optimization*, 10:963–981, 2000.

[132] S. M. Robinson. Regularity and stability for convex multivalued functions. *Mathematics of Operations Research*, 1(2):130–143, 1976.

[133] S. M. Robinson. Stability theory for systems of inequalities, part ii: Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13(4):pp. 497–513, 1976.

[134] S. M. Robinson. Some continuity properties of polyhedral multifunctions. *Mathematical Programming Studies*, 14:206–214, 1981.

[135] R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12(2):268–285, 1974.

[136] R. T. Rockafellar. Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming. In *Nondifferential and Variational Techniques in Optimization*, volume 17 of *Mathematical Programming Studies*, pages 28–66. Springer Berlin Heidelberg, 1982.

[137] R. T. Rockafellar. Marginal values and second-order necessary conditions for optimality. *Mathematical Programming*, 26:245–286, 1983.

[138] R. T. Rockafellar. First- and second-order epi-differentiability in nonlinear programming. *Transactions of the American Mathematical Society*, 307(1):pp. 75–108, 1988.

[139] R. T. Rockafellar. Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives. *Mathmatics of Operations Research*, 14:462–484, 1989.

[140] R. T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35:183–238, 1993.

[141] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, Berlin, 1998.

[142] E. Rosenberg. Exact penalty functions and stability in locally lipschitz programming. *Mathematical Programming*, 30:340–356, 1984.

[143] H. Rubin and P. Ungar. Motion under a strong constraining force. *Communications on Pure and Applied Mathematics*, 10(1):65–87, 1957.

[144] A. M. Rubinov and X. Q. Yang. L*agrange-Type Functions in Constrained Non-Convex Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[145] A. M. Rubinov, X. Q. Yang, A. M. Bagirov, and R. N. Gasimov. Lagrange-type functions in constrained optimization. *Journal of Mathematical Sciences*, 115:2437–2505, 2003.

[146] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[147] S. Scholtes and M. Stöhr. Exact penalization of mathematical programs with equilibrium constraints. *SIAM Journal on Control and Optimization*, 37(2):617–652, 1999.

[148] W. Song. Calmness and error bounds for convex constraint systems. *SIAM Journal on Optimization*, 17(2):353–371, 2006.

[149] G. Still and M. Streng. Optimality conditions in smooth nonlinear programming. *Journal of Optimization Theory and Applications*, 90(3):483–515, 1996.

[150] M. Studniarski. On weak sharp minima for a special class of nonsmooth functions. *German-Polish Conference on Optimization—Methods and Applications (Żagań, 1999). Discuss. Math. Differ. Incl. Control Optim.*, 20(2):195–207, 2000.

[151] M. Studniarski and D. E. Ward. Weak sharp minima: Characterizations and sufficient conditions. *SIAM Journal on Control and Optimization*, 38(1):219–236, 1999.

[152] A. Uderzo. Exact penalty functions and calmness for mathematical programming under nonlinear perturbations. *Nonlinear Analysis: Theory, Methods & Applications*, 73(6):1596 – 1609, 2010.

[153] T. Wang and J. S. Pang. Global error bounds for convex quadratic inequality systems. *Optimization*, 31(1):1–12, 1994.

[154] D. E. Ward. Differential stability in non-lipschitzian optimization. *Journal of Optimization Theory and Applications*, 73:101–120, 1992.

[155] Z. L. Wu and J. J. Ye. On error bounds for lower semicontinuous functions. *Mathematical Programming*, 92(2):301–314, 2002.

[156] Z. L. Wu and J. J. Ye. Sufficient conditions for error bounds. *SIAM Journal on Optimization*, 12(2):421–435, 2002.

[157] Z. L. Wu and J. J. Ye. First-order and second-order conditions for error bounds. *SIAM Journal on Optimization*, 14(3):621–645, 2003.

[158] Z. L. Wu and J.J. Ye. Error bounds for lower semicontinuous inequality systems. In *Decision and Control, 2002, Proceedings of the 41st IEEE Conference on*, volume 3, pages 3553 – 3554, 2002.

[159] X. Q. Yang. On relations and applications of generalized second-order directional derivatives. *Nonlinear Analysis*, 36(5):595–614, 1999.

[160] X. Q. Yang and X. X. Huang. Lower-order penalty methods for mathematical programs with complementarity constraints. *Optimization Methods and Software*, 19(6):693–720, 2004.

[161] X. Q. Yang and Z. Q. Meng. Lagrange multipliers and calmness conditions of order p. *Mathmatics of Operations Research*, 32(1):95–101, 2007.

[162] X. Q. Yang and Y. Y. Zhou. Second-order analysis of penalty function. *Journal of Optimization Theory and Applications*, 146:445–461, 2010.

[163] J. J. Ye. New uniform parametric error bounds. *Journal of Optimization Theory and Applications*, 98(1):197–219, 1998.

[164] J. J. Ye. Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints. *SIAM Journal on Optimization*, 10(4):943–962, 2000.

[165] J. J. Ye. Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 307(1):350–369, 2005.

[166] J. J. Ye and X. Y. Ye. Necessary optimality conditions for optimization problems with variational inequality constraints. *Mathematics of Operations Research*, 22(4):977–997, 1997.

[167] J. J. Ye and D. L. Zhu. Optimality conditions for bilevel programming problems. *Optimization*, 33:9–27, 1995.

[168] J. J. Ye, D. L. Zhu, and Q. J. Zhu. Exact penalization and necessary optimality conditions for generalized bilevel programming problems. *SIAM Journal on Optimization*, 7(2):481–507, 1997.

[169] W. I. Zangwill. Non-linear programming via penalty functions. *Management Science*, 13(5):344–358, 1967.

[170] X. Y. Zheng and K. F. Ng. Metric regularity and constraint qualifications for convex inequalities on Banach spaces. *SIAM Journal on Optimization*, 14(3):757–772, 2003.

[171] X. Y. Zheng and K. F. Ng. Error bound moduli for conic convex systems on banach spaces. *Mathematics of Operations Research*, 29(2):213–228, 2004.

[172] X. Y. Zheng and K. F. Ng. Perturbation analysis of error bounds for systems of conic linear inequalities in Banach spaces. *SIAM Journal on Optimization*, 15(4):1026–1041, 2005.

[173] X. Y. Zheng and K. F. Ng. Metric subregularity and constraint qualifications for convex generalized equations in banach spaces. *SIAM Journal on Optimization*, 18(2):437–460, 2007.

[174] X. Y. Zheng and X. Q. Yang. Weak sharp minima for semi-infinite optimization problems with applications. *SIAM Journal on Optimization*, 18(2):573–588, 2007.

[175] Y. Y. Zhou and X. Q. Yang. Some results about duality and exact penalization. *Journal of Global Optimization*, 29:497–509, 2004.

[176] Y. Y. Zhou and X. Q. Yang. Duality and penalization in optimization via an augmented Lagrangian function with applications. *Journal of Optimization Theory and Applications*, 140(1):171–188, 2009.