Variational Analysis on Local Sharp Minima via Exact Penalization

Kaiwen Meng
School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China
e-mail: mkwfly@126.com

Xiaoqi Yang
Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong
e-mail: mayangxq@polyu.edu.hk

April 15, 2022

Abstract: In this paper we study local sharp minima of the nonlinear programming problem via exact penalization. Utilizing generalized differentiation tools in variational analysis such as subderivatives and regular subdifferentials, we obtain some primal and dual characterizations for a penalty function associated with the nonlinear programming problem to have a local sharp minimum. These general results are then applied to the ℓ_p penalty function with $0 \leq p \leq 1$. In particular, we present primal and dual equivalent conditions in terms of the original data of the nonlinear programming problem, which guarantee that the ℓ_p penalty function has a local sharp minimum with a finite penalty parameter in the case of $p \in (\frac{1}{2}, 1]$ and $p = \frac{1}{2}$ respectively. By assuming the Guignard constraint qualification (resp. the generalized Guignard constraint qualification), we also show that a local sharp minimum of the nonlinear programming problem can be an exact local sharp minimum of the ℓ_p penalty function with $p \in [0, 1]$ (resp. $p \in [0, \frac{1}{2}]$). Finally, we give some formulas for calculating the smallest penalty parameter for a penalty function to have a local sharp minimum.

Keywords: local sharp minimum · subderivative · regular subdifferential · exact penalization · smallest penalty parameter

AMS Classification 49J53 · 49K99 · 65K10

1 Introduction

The notion of a local sharp minimum, due to Poljak [19], plays an important role in the convergence analysis of many iterative procedures [5, 20, 1, 10]. Note that other terminologies were often used for local sharp minima in the literature, such as strongly unique local minima [5], strict local minimizers of order 1 [28, 26, 27], and isolated local minima with order 1 [2, 25]. Characterizations of local sharp minima in terms of generalized derivatives and tangent cones were extensively investigated in [2, 25, 28, 26, 27]. Recall that for an extended-real-valued function ψ and a point $\bar{x}$ with $\psi(\bar{x})$ finite, ψ is said to have a local sharp minimum at $\bar{x}$ if there exist some $\alpha > 0$ and $\delta > 0$ such that

$$\psi(x) \geq \psi(\bar{x}) + \alpha\|x - \bar{x}\| \quad \forall x \in B_\delta(\bar{x}),$$

where $B_\delta(\bar{x})$ is a closed ball of radius δ centered at $\bar{x}$. From [28, Proposition 2.1], it follows that ψ has an unconstrained local sharp minimum at $\bar{x}$ if and only if $d\psi(\bar{x})(w) > 0$ for all $w \neq 0$, where $d\psi(\bar{x})$ denotes the subderivative (also known as the Hadamard directional derivative; see the definition at the end of the section) of ψ at $\bar{x}$.

Concerning local sharp minima in constrained cases, we consider in this paper the nonlinear programming problem

$$\begin{aligned} &\text{(NLP)} \quad \min \quad f(x) \\ &\text{s.t.} \quad x \in C := \left\{ x \in \Re^n \;\middle|\; \begin{array}{ll} g_i(x) \leq 0, & i \in I \\ h_j(x) = 0, & j \in J \end{array} \right\}, \end{aligned}$$

where $I := \{1, 2, \ldots, m\}$, $J := \{m+1, m+2, \ldots, m+q\}$, the functions $f, g_i, h_j : \Re^n \to \Re$ are all assumed to be at least twice continuously differentiable. Throughout the paper, let ϕ be a nonnegative extended-real-valued function having the property that $\phi(x) = 0$ if and only if $x \in C$. The function ϕ can be considered as a measure of violation of the constraints of (NLP). With the help of ϕ, we can define a penalty function for (NLP) as follows:

$$f(x) + \mu\phi(x), \tag{1}$$

where $\mu > 0$ is the penalty parameter. Penalty functions of (NLP) in the form of (1) include the ℓ_p $(0 \leq p \leq 1)$ penalty functions

$$f(x) + \mu S^p(x), \tag{2}$$

as special cases, where

$$S^p(x) := \left(\sum_{i \in I} \max\{g_i(x), 0\} + \sum_{j \in J} |h_j(x)| \right)^p \tag{3}$$

with the convention $0^0 := 0$ being used when $p = 0$. We refer the reader to the excellent survey paper [4] and the references therein for a comprehensive investigation on the central roles that the ℓ_1 penalty function, dating back to [6] and [30], plays in constrained optimization. When $p < 1$, the ℓ_p penalty function is often referred to as a lower order penalty function, which was introduced in [14] for the study of mathematical programs with equilibrium constraints, and rediscovered from nonlinear Lagrangian and unified augmented Lagrangian schemes in [23] and [12] respectively. Recently, first- and second-order necessary optimality conditions have been derived for local minima of (NLP) by assuming exactness of lower order penalty functions and imposing some regularity conditions on the constraints; see [29, 17, 18].

In this paper, we shall study local sharp minima of (NLP) in connection with penalty functions associated with (NLP). Let $\bar{x} \in C$. For the sake of simplicity and unambiguity, we say that (NLP) has a local sharp minimum at $\bar{x}$ if there exist some $\alpha > 0$ and $\delta > 0$ such that

$$f(x) \geq f(\bar{x}) + \alpha \|x - \bar{x}\| \quad \forall x \in C \cap B_\delta(\bar{x}), \tag{4}$$

and that the penalty function (1) has an exact local sharp minimum at $\bar{x}$ if $f + \mu\phi$ has a local sharp minimum at $\bar{x}$ with some finite penalty parameter $\mu > 0$, or explicitly there exist some $\mu > 0$, $\alpha > 0$ and $\delta > 0$ such that

$$f(x) + \mu\phi(x) \geq f(\bar{x}) + \mu\phi(\bar{x}) + \alpha \|x - \bar{x}\| \quad \forall x \in B_\delta(\bar{x}). \tag{5}$$

It is clear that (5) implies (4). Given a local sharp minimum $\bar{x}$ of (NLP) and a measure ϕ of violation of the constraints of (NLP), we can define the smallest penalty parameter (for short, SPP) for the penalty function (1) to have a local sharp minimum at $\bar{x}$ as follows:

$$\mathrm{SPP}(f, \phi, \bar{x}) := \inf\{\mu > 0 \mid \text{the function } f + \mu\phi \text{ has a local sharp minimum at } \bar{x}\}, \tag{6}$$

where the convention $\inf \emptyset := +\infty$ is used. By definition, the penalty function (1) has an exact local sharp minimum at $\bar{x}$ if and only if $\mathrm{SPP}(f, \phi, \bar{x}) < +\infty$, and for each $\mu > \mathrm{SPP}(f, \phi, \bar{x})$, the function $f + \mu\phi$ must have a local sharp minimum at $\bar{x}$.

From [28, Theorem 2.2], it follows that $\bar{x}$ is a local sharp minimum of (NLP) if and only if

$$\langle \nabla f(\bar{x}), w \rangle > 0 \quad \forall w \in T_C(\bar{x}) \backslash \{0\}, \tag{7}$$

where $T_C(\bar{x})$ denotes the tangent cone to C at $\bar{x}$. As it has been pointed out in [18] that the kernel of the subderivative of ϕ at $\bar{x}$, denoted by $\ker d\phi(\bar{x})$, is a closed cone satisfying $T_C(\bar{x}) \subset \ker d\phi(\bar{x})$, the following condition is clearly sufficient for $\bar{x}$ to be a local sharp minimum of (NLP):

$$\langle \nabla f(\bar{x}), w \rangle > 0 \quad \forall w \in \ker d\phi(\bar{x}) \backslash \{0\}. \tag{8}$$

Observing that the measure ϕ of violation of the constraints of (NLP) is involved in (8), the question then arises, "Does the property (8) have some connections with exact local sharp minima of the penalty function (1)?" The answer is yes as will be seen in the sequel.

This paper aims at characterizing exact local sharp minima of penalty functions associated with (NLP) by using conditions (8) and their dual counterparts, and at giving some formulas for calculating the smallest penalty parameters defined by (6). We will show that the penalty function (1) has an exact local sharp minimum at $\bar{x}$ if and only if (8) holds, or equivalently $-\nabla f(\bar{x})$ belongs to $\operatorname{int}(\operatorname{pos}(\widehat{\partial}\phi(\bar{x}))$, the interior of the positive hull of the regular subdifferential $\widehat{\partial}\phi(\bar{x})$ of ϕ at $\bar{x}$. Moreover, in the case of $\widehat{\partial}\phi(\bar{x}) \neq \emptyset$, we will show that

$$\mathrm{SPP}(f, \phi, \bar{x}) = \gamma_{\widehat{\partial}\phi(\bar{x})}(-\nabla f(\bar{x})),$$

where for a closed and convex set M with $0 \in M$, γ_M is the gauge of M to be defined below.

The general results obtained for the penalty function (1) are then applied to the ℓ_p ($0 \leq p \leq 1$) penalty functions (2). We will answer the question as to when the ℓ_p penalty function has an exact local sharp minimum at $\bar{x}$ without or with the help of some constraint qualifications. In particular, we show in Theorem 3.1 that the ℓ_p penalty function with $\frac{1}{2} < p \leq 1$ has an exact local sharp minimum at $\bar{x}$, if and only if, the linear system

$$\langle \nabla f(\bar{x}), w \rangle \leq 0, \quad \langle \nabla g_i(\bar{x}), w \rangle \leq 0 \ \ \forall i \in I(\bar{x}), \quad \langle \nabla h_j(\bar{x}), w \rangle = 0 \ \ \forall j \in J \tag{9}$$

has a unique solution $w = 0$, where $I(\bar{x}) := \{i \in I \mid g_i(\bar{x}) = 0\}$ denotes the active inequality index set of (NLP) at $\bar{x}$. The equivalent dual counterpart of (9) is also obtained as follows: The KKT condition holds at $\bar{x}$ with some multiplier satisfying the strict complementarity condition, and the linear space spanned by vectors $\nabla g_i(\bar{x})$ and $\nabla h_j(\bar{x})$ with $i \in I(\bar{x})$ and $j \in J$ respectively is $\Re^n$. In Theorem 3.2, we show that the $\ell_{\frac{1}{2}}$ penalty function has an exact local sharp minimum at $\bar{x}$ if and only if the system (27) of finitely many quadratic forms mixed with linear equations and inequalities has a unique solution, or equivalently the second-order sufficient condition (28) holds. With the help of some constraint qualification, we further show that a local sharp minimum of (NLP) is an exact local minimum of the ℓ_p penalty function with p ranging in an interval depending on the constraint qualification assumed. For instance, under the Guignard constraint qualification originating with [9], we show that $\bar{x}$ is a local sharp minimum of (NLP) if and only if the ℓ_p penalty function with $p \in [0,1]$ has an exact local sharp minimum at $\bar{x}$, while under the so-called generalized Guignard constraint qualification newly introduced in Definition 3.1, we show that $\bar{x}$ is a local sharp minimum of (NLP) if and only if the ℓ_p penalty function with $p \in [0, \frac{1}{2}]$ has an exact local sharp minimum at $\bar{x}$. Moreover, we give formulas for calculating $\mathrm{SPP}(f, S^p, \bar{x})$ in various settings.

The outline of the paper is as follows. In Section 2, we present primal and dual characterizations for exact local minima of the penalty function (1). Applications of these characterizations to the ℓ_p $(0 \le p \le 1)$ penalty functions (2) can be found in Section 3. The paper is ended with some conclusions in Section 4.

We conclude this section by reviewing some notions and notation that are needed in this paper. The notation that we employ in this paper is for the most part borrowed from the book [22]. Let $\overline{\Re} := \Re \cup \{\pm\infty\}$ and let $\Re_+ := \{t \in \Re \mid t \ge 0\}$. For vectors x, y in $\Re^n$, we denote by x^T the transpose of x, by $\langle x, y \rangle$ 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 any function $f : \Re^n \to \Re_+ \cup \{+\infty\}$ and any $p > 0$, let $f^p(x) = (f(x))^p$ for all $x \in \Re^n$ with the convention that $(+\infty)^p = +\infty$. For a closed and convex set $M \subset \Re^n$ with $0 \in M$, the gauge of M is the function $\gamma_M : \Re^n \to \overline{\Re}$ defined by

$$\gamma_M(x) := \inf\{\lambda \ge 0 \mid x \in \lambda M\},$$

where the convention $\inf \emptyset = +\infty$ is used.

For a given subset A of $\Re^n$, we denote the closure of A, the interior of A, the boundary of A, the linear space spanned by A and the convex hull of A respectively by $\mathrm{cl}\, A$, $\mathrm{int}\, A$, $\mathrm{bd}\, A$, $\mathrm{span}\, A$ and $\mathrm{con}\, A$. The polar cone of A is defined by

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

The positive hull of A is defined by

$$\mathrm{pos}\, A := \{\lambda x \mid x \in A, \lambda \ge 0\}.$$

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

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

The support function $\sigma_A : \Re^n \to \overline{\Re}$ of A is defined by

$$\sigma_A(w) := \sup_{v \in A} \langle v, w \rangle.$$

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

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

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

$$A^* := \Re^n, \quad \operatorname{pos} A := \{0\}, \quad A^\infty := \{0\}, \quad \sigma_A(\cdot) = -\infty, \quad d_A(\cdot) := +\infty, \quad \text{and} \quad \delta_A(\cdot) := +\infty.$$

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

Let $f : \Re^n \to \overline{\Re}$ be an extended-real-valued function. The effective domain of f is the set

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

the kernel of f is the set

$$\ker f := \{x \in \Re^n \mid f(x) = 0\},$$

and the epigraph of f is the set

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

The function f is said to be lower semicontinuous iff $\operatorname{epi} f$ is closed in $\Re^n \times \Re$. Moreover, f is said to be positively homogeneous, iff $0 \in \operatorname{dom} f$ and $f(\lambda x) = \lambda f(x)$ for all x and all $\lambda > 0$, and it is sublinear, iff in addition

$$f(x + x') \leq f(x) + f(x') \text{ for all } x \text{ and } x'.$$

Let $\bar{x}$ be a point with $f(\bar{x})$ finite. The notions of subgradients and subderivatives that we need throughout the paper are summarized below; see [22, Chapters 8 and 13] for more details.

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

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

(ii) For any $w \in \Re^n$, the subderivative of f at $\bar{x}$ for w is defined by

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

(iii) For any vector $v \in \Re^n$, the second subderivative at $\bar{x}$ for v and w is defined by

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

2 Primal and Dual Characterizations of Exact Local Sharp Minima

To begin with, we exploit some relations between a closed and convex set M, consisting of the gradients of the affine supports of a nonnegative, lower semicontinuous and positively homogeneous function h, and the kernel $\ker h$ of h.

Lemma 2.1 *For any nonnegative, lower semicontinuous and positively homogeneous function* $h : \Re^n \to \overline{\Re}$, $\ker h$ *is a nonempty closed cone, the set defined by*

$$M := \{v \in \Re^n \mid \langle v, w\rangle \le h(w) \quad \forall w \in \Re^n\}$$

is closed and convex with $0 \in M$*, and the following statements are true:*

(a) $M^\infty = [\operatorname{dom} h]^* \subset M \subset [\ker h]^*$*, for which the first inclusion is an equality if and only if* M *is a cone, while the second inclusion is an equality if and only if* $[\operatorname{dom} h]^* = [\ker h]^*$*. If furthermore* h *is convex, then* $\operatorname{cl}(\operatorname{pos} M) = [\ker h]^*$.

(b) M *has nonempty interior if and only if* $\ker h$ *is pointed. In that case,* $\operatorname{cl}(\operatorname{pos} M) = [\ker h]^*$.

(c) $\operatorname{int}(\operatorname{pos} M) = \{\lambda w \mid w \in \operatorname{int} M, \lambda > 0\} = \{v \in \Re^n \mid \langle v, w\rangle < 0 \quad \forall w \in \ker h \backslash \{0\}\}$.

(d) *If there exists some* $\tau > 0$ *such that* $d(w, \ker h) \le \tau h(w)$ *for all* w*, then* $\operatorname{pos} M = [\ker h]^*$.

Proof. In view of the fact that M is the intersection of a collection of closed half-spaces, M is closed and convex. By the nonnegativity of h and the definition of M, we have $0 \in M$. This further implies that

$$\operatorname{int}(\operatorname{pos} M) = \operatorname{int}(\operatorname{cl}(\operatorname{pos} M)) = \operatorname{int} T_M(0) = \{\lambda w \mid w \in \operatorname{int} M, \lambda > 0\}, \tag{10}$$

where the first equality follows from the convexity of $\operatorname{pos} M$, and the last two equalities follow from the formulas for the tangent cone to a convex set ([22, Theorem 6.9]). Since h is nonnegative and positively homogeneous, we have $h(0) = 0$ and $h(\lambda x) = 0$ for all $\lambda > 0$ and $x \in \ker h$. Thus, $\ker h$ is a nonempty cone. Since h is nonnegative and lower semicontinuous, it follows that $\ker h = \{w \in \Re^n \mid h(w) \le 0\}$ is closed. That is, $\ker h$ is a nonempty closed cone. Since $\langle v, w\rangle \le 0$ for all $v \in M$ and $w \in \ker h$, we have

$$M \subset [\ker h]^*. \tag{11}$$

Let $v \in [\operatorname{dom} h]^*$. By definition, we have $\langle v, w\rangle \le 0$ for all $w \in \operatorname{dom} h$. Since h is nonnegative, it follows that $\langle v, w\rangle \le h(w)$ for all $w \in \Re^n$, i.e., $v \in M$. Thus, we have $[\operatorname{dom} h]^* \subset M$. By [22, Exercise 3.24] and the definition of M, we have $M^\infty = [\operatorname{dom} h]^*$. By the definition of the horizon cone, we have $[\operatorname{dom} h]^* = M$ if and only if M is a cone. Moreover, $M = [\ker h]^*$ (implying that M is a cone and hence $[\operatorname{dom} h]^* = M$) if and only if $[\operatorname{dom} h]^* = [\ker h]^*$. By [22, Theorem 8.24], if h is convex, we have $h = \sigma_M$, which implies that $w \in \ker h$ if and only if $\langle v, w\rangle \le 0$ for all $v \in M$, or in other words $w \in M^*$. That is, when h is convex, we have $\ker h = M^*$. Noting that $M^* = (\operatorname{cl}(\operatorname{pos} M))^*$, that $\operatorname{pos} M$ is convex as M is convex, and that $\ker h$ is closed and convex, we assert that $\ker h = M^*$ if and only if $\operatorname{cl}(\operatorname{pos} M) = [\ker h]^*$, by applying [22, Corollary 6.21], which says that for a cone K, the polar cone K^* is closed and convex, and $K^{**} = \operatorname{cl}(\operatorname{con} K)$. Thus, (a) is proved.

We now prove (b) and (c) together. Let $A := \{v \mid \langle v, w\rangle < 0 \quad \forall w \in \ker h\backslash\{0\}\}$. It follows from [22, Exercise 6.22] that $\operatorname{int}[\ker h]^* = A$, and that the following conditions are equivalent: (i) $A \neq \emptyset$; (ii) $\ker h$ is pointed; and (iii) $[\ker h]^*$ has nonempty interior.

Assume that $\ker h$ is not pointed. We claim that $\operatorname{int} M = \emptyset$, for otherwise it follows from the inclusion (11) that $[\ker h]^*$ has nonempty interior, which amounts to the pointedness of $\ker h$. By (10), we have $\operatorname{int}(\operatorname{pos} M) = \emptyset$. Therefore, whenever $\ker h$ is not pointed, M has empty interior and the property (c) holds trivially.

Assume now that $\ker h$ is pointed. Since the closed cone $\operatorname{epi} h$ is a subset of $\Re^n \times \Re_+$ and $\operatorname{epi} h \cap (\Re^n \times \{0\}) = \ker h \times \{0\}$, the pointedness of $\ker h$ amounts to the pointedness of $\operatorname{epi} h$. It follows from [22, Theorem 3.15] that both the cones $\operatorname{con}(\ker h)$ and $\operatorname{con}(\operatorname{epi} h)$ are closed and pointed, and that

$$
\begin{aligned}
\operatorname{con}(\operatorname{epi} h) \cap (\Re^n \times \{0\}) &= (\operatorname{epi} h + \cdots + \operatorname{epi} h(n \text{ terms})) \cap (\Re^n \times \{0\}) \\
&= (\operatorname{epi} h \cap (\Re^n \times \{0\})) + \cdots + (\operatorname{epi} h \cap (\Re^n \times \{0\}))(n \text{ terms}) \\
&= (\ker h \times \{0\}) + \cdots + (\ker h \times \{0\})(n \text{ terms}) \\
&= \operatorname{con}(\ker h \times \{0\}) = \operatorname{con}(\ker h) \times \{0\}.
\end{aligned}
$$

Then by [22, Corollary 6.21], we have $[\ker h]^{**} = \operatorname{con}(\ker h)$ and $[\operatorname{epi} h]^{**} = \operatorname{con}(\operatorname{epi} h)$. This entails that $w \in [\ker h]^{**}$ if and only if $(w, 0) \in [\operatorname{epi} h]^{**}$. Note that $\langle (v, -1), (w, \beta)\rangle \le 0$ for all $(w, \beta) \in \operatorname{epi} h$ if and only if $v \in M$, and that $\langle (v, 0), (w, \beta)\rangle \le 0$ for all $(w, \beta) \in \operatorname{epi} h$ if and only if $\langle v, w\rangle \le 0$ for all $w \in \operatorname{dom} h$, or equivalently $v \in [\operatorname{dom} h]^* = M^\infty$. Thus, we have

$$
[\operatorname{epi} h]^* = \{\lambda(v, -1) \mid v \in M, \lambda > 0\} \cup \{(v, 0) \mid v \in M^\infty\}. \tag{12}
$$

Let $w \in M^*$. In view of (12) and the inclusion $M^\infty \subset M$, we have $(w, 0) \in [\operatorname{epi} h]^{**}$ and hence $w \in [\ker h]^{**}$. This implies that $M^* \subset [\ker h]^{**}$, for which the reverse inclusion follows from the inclusion (11). That is, we have $M^* = [\ker h]^{**}$ or equivalently $\mathrm{cl}\,(\mathrm{pos}\, M) = [\ker h]^*$ by $M^{**} = \mathrm{cl}\,(\mathrm{pos}\, M)$. Since $\mathrm{pos}\, M$ is convex, we have $\mathrm{int}\,(\mathrm{pos}\, M) = \mathrm{int}\,(\mathrm{cl}\,(\mathrm{pos}\, M)) = \mathrm{int}\,[\ker h]^* = A$. So, if $\ker h$ is pointed, then the property (c) holds. Since the pointedness of $\ker h$ amounts to $A \neq \emptyset$, we thus have $\mathrm{int}\,(\mathrm{pos}\, M) \neq \emptyset$. In view of (10), we have $\mathrm{int}\, M \neq \emptyset$. This completes the proof of (b).

Now (d) follows from [16, Theorem 5.1]. This completes the proof. $\square$

For a function $\psi : \Re^n \to \overline{\Re}$ and a point $\bar{x}$ with $\psi(\bar{x})$ finite, we have by [22, Exercise 8.4]

$$\widehat{\partial}\psi(\bar{x}) = \{v \in \Re^n \mid \langle v, w\rangle \leq d\psi(\bar{x})(w) \quad \forall w \in \Re^n\}. \tag{13}$$

Observing that the subderivative $d\psi(\bar{x})$ is a lower semicontinuous and positively homogeneous function, Lemma 2.1 is thus applicable in the circumstance that $0 \in \widehat{\partial}\psi(\bar{x})$ or equivalently $d\psi(\bar{x}) \geq 0$ as is true in particular when ψ has a local minimum at $\bar{x}$; see the generalized Fermat's rule [22, Theorem 10.1]. The following theorem, which has been partially shown by the authors in [18, 16], is more or less a restatement of Lemma 2.1, but fully characterizes the connections between $\ker d\psi(\bar{x})$ and $\widehat{\partial}\psi(\bar{x})$.

Theorem 2.1 *Let $\psi : \Re^n \to \overline{\Re}$ be a function with $\psi(\bar{x})$ finite. Assume that $0 \in \widehat{\partial}\psi(\bar{x})$ or equivalently $d\psi(\bar{x}) \geq 0$. The following statements are true:*

(a) $\widehat{\partial}\psi(\bar{x})^\infty = [\mathrm{dom}\, d\psi(\bar{x})]^* \subset \widehat{\partial}\psi(\bar{x}) \subset [\ker d\psi(\bar{x})]^*$, *for which the first inclusion is an equality if and only if $\widehat{\partial}\psi(\bar{x})$ is a cone, while the second inclusion is an equality if and only if $[\mathrm{dom}\, d\psi(\bar{x})]^* = [\ker d\psi(\bar{x})]^*$. If furthermore $d\psi(\bar{x})$ is convex as is true when ψ is regular at $\bar{x}$, then $\mathrm{cl}\,(\mathrm{pos}\,\widehat{\partial}\psi(\bar{x})) = [\ker d\psi(\bar{x})]^*$.*

(b) *$\widehat{\partial}\psi(\bar{x})$ has nonempty interior if and only if $\ker d\psi(\bar{x})$ is pointed. In that case, $\mathrm{cl}\,(\mathrm{pos}\,\widehat{\partial}\psi(\bar{x})) = [\ker d\psi(\bar{x})]^*$.*

(c) $\mathrm{int}\,(\mathrm{pos}\,\widehat{\partial}\psi(\bar{x})) = \{\lambda v \mid v \in \mathrm{int}\,\widehat{\partial}\psi(\bar{x}), \lambda > 0\} = \{v \in \Re^n \mid \langle v, w\rangle < 0\ \forall w \in \ker d\psi(\bar{x})\backslash\{0\}\}$.

(d) *If there exists some $\tau > 0$ such that $d(w, \ker d\psi(\bar{x})) \leq \tau d\psi(\bar{x})(w)$ for all w, then*

$$\mathrm{pos}\,\widehat{\partial}\psi(\bar{x}) = [\ker d\psi(\bar{x})]^*.$$

For a function $\psi : \Re^n \to \overline{\Re}$ and a point $\bar{x}$ with $\psi(\bar{x})$ finite, it follows from [28, Proposition 2.1] or the definition of subderivative that, ψ has a local sharp minimum at $\bar{x}$ if and only if $d\psi(\bar{x})(w) > 0$ for all $w \in \Re^n$ with $w \neq 0$. The latter condition has the inclusion $0 \in$

int $(\widehat{\partial}\psi(\bar{x}))$ as its dual counterpart as can be easily seen from Theorem 2.1 (c) or the equality (13). These primal and dual characterizations of local sharp minima are summarized in the following corollary 2.1 for further reference.

Corollary 2.1 *The following properties are equivalent:*

(i) *ψ has a local sharp minimum at $\bar{x}$.*

(ii) *$d\psi(\bar{x})(w) > 0$ for all $w \in \Re^n$ with $w \neq 0$.*

(iii) $0 \in \operatorname{int}(\widehat{\partial}\psi(\bar{x}))$.

For a function $\psi : \Re^n \to \overline{\Re}$ and a point $\bar{x}$ with $\psi(\bar{x})$ finite, the variational description [22, Proposition 8.5] of a regular subgradient is as follows: A vector v belongs to $\widehat{\partial}\psi(\bar{x})$, if and only if, there exists some $\delta > 0$ along with a function $h : \Re^n \to \overline{\Re}$ smooth on $B_\delta(\bar{x})$ such that $h(\bar{x}) = \psi(\bar{x})$, $\nabla h(\bar{x}) = v$ and

$$\psi(x) > h(x) \quad \forall x \in B_\delta(\bar{x})\backslash\{\bar{x}\}. \tag{14}$$

However, if the regular subgradient v belongs to the interior of $\widehat{\partial}\psi(\bar{x})$, the inequality (14) can be strengthened as can be seen from the following corollary.

Corollary 2.2 *(variational description of interior of regular subdifferentials). A vector v belongs to the interior of $\widehat{\partial}\psi(\bar{x})$, if and only if, there exist some $\varepsilon > 0$ and $\delta > 0$ along with a function $h : \Re^n \to \overline{\Re}$ smooth on $B_\delta(\bar{x})$ such that $h(\bar{x}) = \psi(\bar{x})$, $\nabla h(\bar{x}) = v$ and*

$$\psi(x) > h(x) + \varepsilon\|x - \bar{x}\| \quad \forall x \in B_\delta(\bar{x})\backslash\{\bar{x}\}. \tag{15}$$

Proof. If h is smooth on $B_\delta(\bar{x})$ with $\nabla h(\bar{x}) = v$ for some $\delta > 0$, we have by [22, Exercise 8.8],

$$\widehat{\partial}(\psi - h)(\bar{x}) = \widehat{\partial}\psi(\bar{x}) - v.$$

Observing that $h(\bar{x}) = \psi(\bar{x})$, we can reformulate (15) as

$$(\psi - h)(x) > (\psi - h)(\bar{x}) + \varepsilon\|x - \bar{x}\| \quad \forall x \in B_\delta(\bar{x})\backslash\{\bar{x}\},$$

which by Corollary 2.1, amounts to $0 \in \operatorname{int}(\widehat{\partial}(\psi - h)(\bar{x}))$, or in other words, $v \in \operatorname{int}(\widehat{\partial}\psi(\bar{x}))$. This completes the proof. $\square$

We now present the primal and dual conditions for exact sharp local minima of the penalty function (1).

Theorem 2.2 *Let $\bar{x} \in C$. The penalty function (1) has an exact local sharp minimum at $\bar{x}$ if and only if $-\nabla f(\bar{x}) \in \operatorname{int}(\operatorname{pos}(\widehat{\partial}\phi(\bar{x}))$ or equivalently (8) holds. In that case,*

$$\mathrm{SPP}(f, \phi, \bar{x}) = \gamma_{\widehat{\partial}\phi(\bar{x})}(-\nabla f(\bar{x})). \tag{16}$$

Proof. Let $\bar{x} \in C$. Observing that ϕ has a global minimum at $\bar{x}$, we have $0 \in \widehat{\partial}\phi(\bar{x})$ and hence Theorem 2.1 is applicable. By [22, Exercise 8.8], we have for each $\mu > 0$,

$$\widehat{\partial}(f + \mu\phi)(\bar{x}) = \nabla f(\bar{x}) + \mu\widehat{\partial}\phi(\bar{x}). \tag{17}$$

By definition, the penalty function (1) has an exact local sharp minimum at $\bar{x}$ if and only if there exists some $\mu > 0$ such that $f + \mu\phi$ has a local sharp minimum at $\bar{x}$, or equivalently by (17) and Corollary 2.1, there exists some $\mu > 0$ such that $-\nabla f(\bar{x}) \in \mu \operatorname{int}(\widehat{\partial}\phi(\bar{x}))$. The latter condition amounts to $-\nabla f(\bar{x}) \in \operatorname{int}(\operatorname{pos}(\widehat{\partial}\phi(\bar{x}))$, or in other words, (8) holds as can be seen from Theorem 2.1. Moreover, we have by definition

$$\mathrm{SPP}(f, \phi, \bar{x}) = \inf\{\mu > 0 \mid -\nabla f(\bar{x}) \in \mu \operatorname{int}(\widehat{\partial}\phi(\bar{x}))\}.$$

Since $\widehat{\partial}\phi(\bar{x})$ is a closed and convex set containing the origin, we have the following equality in the case of $\operatorname{int}(\widehat{\partial}\phi(\bar{x})) \neq \emptyset$:

$$\inf\{\mu > 0 \mid -\nabla f(\bar{x}) \in \mu \operatorname{int}(\widehat{\partial}\phi(\bar{x}))\} = \inf\{\mu > 0 \mid -\nabla f(\bar{x}) \in \mu\widehat{\partial}\phi(\bar{x})\},$$

which is equal to $\gamma_{\widehat{\partial}\phi(\bar{x})}(-\nabla f(\bar{x}))$. This completes the proof. □

3 Applications to the ℓ_p Penalty Functions

Throughout this section, let $\bar{x} \in C$ be a fixed feasible point of (NLP). By applying the results presented in the last section, we shall answer the question as to when the ℓ_p $(0 \leq p \leq 1)$ penalty function (2) has an exact local sharp minimum at $\bar{x}$. In what follows, we intend to answer the question in two ways. In Section 3.1, we shall calculate the closed cone $\ker dS^p(\bar{x})$ and/or its counterpart $\widehat{\partial}S^p(\bar{x})$, and characterize exact local sharp minima of the ℓ_p penalty function by virtue of the original data of (NLP) and without the help of any constraint qualification. In Section 3.2, by assuming some constraint qualifications, we shall show that $\bar{x}$ is a local sharp minimum of (NLP) if and only if the ℓ_p penalty function has an exact local sharp minimum at $\bar{x}$ with p ranging in some interval depending on the constraint qualification assumed.

3.1 Exact Local Sharp Minima without Constraint Qualifications

As the calculations for $\ker dS^p(\bar{x})$ and $\widehat{\partial} S^p(\bar{x})$ are different for various p's, we shall consider in what follows two cases: (i) $\frac{1}{2} < p \leq 1$; and (ii) $p = \frac{1}{2}$.

3.1.1 The case of $\frac{1}{2} < p \leq 1$

Denote the active inequality index set of (NLP) at $\bar{x}$ by

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

and the first-order linearized tangent cone to C at $\bar{x}$ by

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

Moreover, let

$$\mathrm{KKT}(\bar{x}) := \left\{ \lambda \in \Re^{m+q} \;\middle|\; \begin{array}{c} \nabla f(\bar{x}) + \sum_{i\in I} \lambda_i \nabla g_i(\bar{x}) + \sum_{j\in J} \lambda_j \nabla h_j(\bar{x}) = 0 \\ \lambda_i = 0 \; \forall i \in I\backslash I(\bar{x}), \quad \lambda_i \geq 0 \; \forall i \in I(\bar{x}) \end{array} \right\}.$$

We recall two formulas from [18] as follows:

$$dS(\bar{x})(w) = \sum_{i\in I(\bar{x})} \max\{\langle \nabla g_i(\bar{x}), w\rangle, 0\} + \sum_{j\in J} |\langle \nabla h_j(\bar{x}), w\rangle|, \quad \forall w \in \Re^n, \tag{18}$$

and

$$\widehat{\partial} S(\bar{x}) = \left\{ \sum_{i\in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) + \sum_{j\in J} \lambda_j \nabla h_j(\bar{x}) \mid 0 \leq \lambda_i \leq 1 \; \forall i \in I(\bar{x}), \; -1 \leq \lambda_j \leq 1 \; \forall j \in J \right\}. \tag{19}$$

See [22, Example 7.28, Exercise 8.31 and Corollary 10.9] for more details. In view of (18) and the definition of $L_C(\bar{x})$, we have

$$\ker dS(\bar{x}) = L_C(\bar{x}). \tag{20}$$

Let

$$K := \left\{ \sum_{i\in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) + \sum_{j\in J} \lambda_j \nabla h_j(\bar{x}) \mid \lambda_i \geq 0 \quad \forall i \in I(\bar{x}), \quad \lambda_j \in \Re \quad \forall j \in J \right\}.$$

In view of (19) and the definition of positive hull, we have $\mathrm{pos}(\widehat{\partial} S(\bar{x})) = K$. By the definition of $L_C(\bar{x})$, we get from the Farkas lemma (cf. [22, Lemma 6.45]) that $L_C(\bar{x})^* = K$. That is, we have

$$\mathrm{pos}(\widehat{\partial} S(\bar{x})) = L_C(\bar{x})^* = K. \tag{21}$$

Theorem 3.1 *Let $\frac{1}{2} < p \le 1$. The following properties are equivalent:*

(a) *The ℓ_p penalty function has an exact local sharp minimum at $\bar{x}$.*

(b) *The linear system*

$$\langle \nabla f(\bar{x}), w\rangle \le 0, \quad \langle \nabla g_i(\bar{x}), w\rangle \le 0 \;\; \forall i \in I(\bar{x}), \quad \langle \nabla h_j(\bar{x}), w\rangle = 0 \;\; \forall j \in J$$

has a unique solution $w = 0$.

(c) $\mathrm{span}(\{\nabla g_i(\bar{x}) \mid i \in I(\bar{x})\} \cup \{\nabla h_j(\bar{x}) \mid j \in J\}) = \Re^n$, *and the KKT condition holds at $\bar{x}$ with some multiplier satisfying the strict complementarity condition, i.e., there exists some $\lambda \in \mathrm{KKT}(\bar{x})$ such that $\lambda_i > 0$ for all $i \in I(\bar{x})$.*

These properties entail that $\mathrm{SPP}(f, S^p, \bar{x}) = 0$ *if $\frac{1}{2} < p < 1$, and*

$$\mathrm{SPP}(f, S, \bar{x}) = \min\{\|\lambda\|_\infty \mid \lambda \in \mathrm{KKT}(\bar{x})\}, \tag{22}$$

where $\|\lambda\|_\infty$ denotes the infinity norm of λ.

Proof. First we consider the case of $p = 1$. The equivalence of (a) and (b) follows immediately from Theorem 2.2 and (20). By [7, Lemma 2.3] and (21), $-\nabla f(\bar{x})$ belongs to the relative interior of $\mathrm{pos}(\widehat{\partial} S(\bar{x}))$ if and only if

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

for some $\lambda_i > 0$ for all $i \in I(\bar{x})$ and $\lambda_j \in \Re$ for all $j \in J$, or in other words, there exists some $\lambda \in \mathrm{KKT}(\bar{x})$ such that $\lambda_i > 0$ for all $i \in I(\bar{x})$. By Theorem 2.1 and (20), $\mathrm{pos}(\widehat{\partial} S(\bar{x}))$ has nonempty interior if and only if the cone $L_C(\bar{x})$ is pointed, i.e.,

$$\{w \in \Re^n \mid \langle \nabla g_i(\bar{x}), w\rangle = 0 \; \forall i \in I(\bar{x}), \quad \langle \nabla h_j(\bar{x}), w\rangle = 0 \; \forall j \in J\} = \{0\},$$

or in other words, $\mathrm{span}(\{\nabla g_i(\bar{x}) \mid i \in I(\bar{x})\} \cup \{\nabla h_j(\bar{x}) \mid j \in J\}) = \Re^n$. Thus, we have shown $-\nabla f(\bar{x}) \in \mathrm{int}\,(\mathrm{pos}(\widehat{\partial} S(\bar{x})))$ if and only if (c) holds. Then by Theorem 2.2, the equivalence of (a) and (c) follows. By Theorem 2.2 and (19), we have

$$\begin{aligned}
\mathrm{SPP}(f, S, \bar{x}) &= \gamma_{\widehat{\partial} S(\bar{x})}(-\nabla f(\bar{x})) \\
&= \inf\{\tau \ge 0 \mid \lambda \in \mathrm{KKT}(\bar{x}), 0 \le \lambda_i \le \tau \; \forall i \in I(\bar{x}), \; -\tau \le \lambda_j \le \tau \; \forall j \in J\} \\
&= \inf\{\|\lambda\|_\infty \mid \lambda \in \mathrm{KKT}(\bar{x})\}.
\end{aligned}$$

According to [22, Theorem 1.9], the problem of minimizing $\|\lambda\|_\infty$ over $\mathrm{KKT}(\bar{x})$ attains its minimum. Thus, (22) follows.

Next we consider the case of $\frac{1}{2} < p < 1$. Since $0 \le S(x) \le S^p(x)$ for all x near $\bar{x}$, we have $\ker dS^p(\bar{x}) \subset \ker dS(\bar{x})$. By (21), we thus have $\ker dS^p(\bar{x}) \subset L_C(\bar{x})$. Let $t_k \to 0+$ and let $w \in L_C(\bar{x})$ be given. In the case of $i \notin I(\bar{x})$ or $i \in I(\bar{x})$ with $\langle \nabla g_i(\bar{x}), w\rangle < 0$, we have

$$\frac{\max\{g_i(\bar{x}+t_k w), 0\}}{t_k^{1/p}} \to 0.$$

By the second-order Taylor expansion, we have

$$\frac{h_j(\bar{x}+t_k w)}{t_k^{1/p}} = \frac{1}{2} t_k^{2-1/p} \langle w, \nabla^2 h_j(\bar{x}) w\rangle + \frac{o(t_k^2)}{t_k^2} t_k^{2-1/p} \to 0 \quad \forall j \in J,$$

and in the case of $i \in I(\bar{x})$ with $\langle \nabla g_i(\bar{x}), w\rangle = 0$,

$$\frac{\max\{g_i(\bar{x}+t_k w), 0\}}{t_k^{1/p}} = \max\{\frac{1}{2} t_k^{2-1/p} \langle w, \nabla^2 h_j(\bar{x}) w\rangle + \frac{o(t_k^2)}{t_k^2} t_k^{2-1/p}, 0\} \to 0.$$

Therefore, we have

$$\frac{S(\bar{x}+t_k w)}{t_k^{1/p}} \to 0 \quad \forall w \in L_C(\bar{x}),$$

or equivalently,

$$\frac{S^p(\bar{x}+t_k w)}{t_k} \to 0 \quad \forall w \in L_C(\bar{x}).$$

This entails that $L_C(\bar{x}) \subset \ker dS^p(\bar{x})$. Thus, we have $\ker dS^p(\bar{x}) = L_C(\bar{x})$ and more generally

$$\ker dS^{\tilde{p}}(\bar{x}) = L_C(\bar{x}) \quad \forall \tilde{p} \in (\frac{1}{2}, 1].$$

Let $w \in \operatorname{dom} dS^p(\bar{x})$. By the definition of subderivative, we can find sequences $t_k \to 0+$ and $w_k \to w$ such that

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

implying that $S(\bar{x}+t_k w_k) \to 0$. Then for any p' with $p < p' \le 1$, we have

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

which implies that $w \in \ker dS^{p'}(\bar{x})$ and hence $\operatorname{dom} dS^p(\bar{x}) \subset \ker dS^{p'}(\bar{x}) = L_C(\bar{x})$. Therefore, we have

$$\ker dS^p(\bar{x}) = \operatorname{dom} dS^p(\bar{x}) = L_C(\bar{x}).$$

By Theorem 2.1 (a), we have $\widehat{\partial} S^p(\bar{x}) = L_C(\bar{x})^*$, implying that $\widehat{\partial} S^p(\bar{x})$ is a cone. Applying Theorem 2.2, statement (a) holds if and only if $\langle \nabla f(\bar{x}), w\rangle > 0$ for all $w \in \ker dS^p(\bar{x}) = L_C(\bar{x})$ with $w \neq 0$, or if and only if $-\nabla f(\bar{x}) \in \mathrm{int}\,(\mathrm{pos}(\widehat{\partial} S^p(\bar{x}))) = \mathrm{int}\,(\widehat{\partial} S^p(\bar{x})) = \mathrm{int}\,(L_C(\bar{x})^*)$. Now following the same arguments as for the proof of the case $p = 1$, statements (a), (b) and (c) are equivalent with each other. Moreover, in view of the facts that $\widehat{\partial} S^p(\bar{x})$ is a cone and that $-\nabla f(\bar{x}) \in \mathrm{int}\,(\widehat{\partial} S^p(\bar{x}))$, we obtain from Theorem 2.2 that $\mathrm{SPP}(f, S^p, \bar{x}) = \gamma_{\widehat{\partial} S^p(\bar{x})}(-\nabla f(\bar{x})) = 0$. This completes the proof. $\square$

Remark 3.1 *Theorem 3.1 (c) in its current form is new, though an equivalent condition in an abstract form was presented in [24, Theorem 3.4] to show that it is sufficient for $\bar{x}$ to be a local sharp minimum of (NLP). In [11], it was shown that Theorem 3.1 (b) is sufficient for the ℓ_1 penalty function to have a strict local minimum at $\bar{x}$, i.e., there exist some $\mu > 0$ and $\delta > 0$ such that*

$$f(x) + \mu S(x) > f(\bar{x}) + \mu S(\bar{x}) \quad \forall x \in B_\delta(\bar{x})\backslash\{\bar{x}\}.$$

While in [4, Theorem 4.8], it was shown that the ℓ_1 penalty function has an exact local sharp minimum at $\bar{x}$ if Theorem 3.1 (b) holds and $\mathrm{KKT}(\bar{x}) \neq \emptyset$. The latter condition turns out to be unnecessary in our result.

3.1.2 The case of $p = \frac{1}{2}$

The calculus rules [22, Example 13.16 and Proposition 13.19] yield that

$$d^2 S(\bar{x} \mid 0)(w) = \begin{cases} \max\limits_{\rho \in \mathrm{KKT}_0(\bar{x}) \cap \mathbb{B}} \left\langle \left[\sum\limits_{i \in I} \rho_i \nabla^2 g_i(\bar{x}) + \sum\limits_{j \in J} \rho_j \nabla^2 h_j(\bar{x}) \right] w, w \right\rangle & \text{if } w \in L_C(\bar{x}), \\ +\infty & \text{otherwise,} \end{cases}$$

where $\mathbb{B} := \{x \in \Re^{m+q} \mid |x_i| \leq 1 \ \forall i\}$ denotes the closed unit ball induced by the infinity norm, and

$$\mathrm{KKT}_0(\bar{x}) := \left\{ \rho \in \Re^{m+q} \;\middle|\; \begin{array}{c} \sum\limits_{i \in I} \rho_i \nabla g_i(\bar{x}) + \sum\limits_{j \in J} \rho_j \nabla h_j(\bar{x}) = 0 \\ \rho_i = 0 \ \forall i \in I \backslash I(\bar{x}), \quad \rho_i \geq 0 \ \forall i \in I(\bar{x}) \end{array} \right\}.$$

By the definition of the second-subderivative, we have

$$d^2 S(\bar{x} \mid 0)(w) = 2[dS^{\frac{1}{2}}(\bar{x})(w)]^2 \quad \forall w \in \Re^n,$$

which gives

$$dS^{\frac{1}{2}}(\bar{x})(w) = \begin{cases} \dfrac{\sqrt{2}}{2}\sqrt{\displaystyle\max_{\rho\in \mathrm{KKT}_0(\bar{x})\cap\mathbb{B}} \left\langle \left[\sum_{i\in I}\rho_i\nabla^2 g_i(\bar{x}) + \sum_{j\in J}\rho_j\nabla^2 h_j(\bar{x})\right] w, w\right\rangle} & \text{if } w\in L_C(\bar{x}),\\ +\infty & \text{otherwise.}\end{cases} \tag{23}$$

With the formula (23) at hand, we can describe the kernel of $dS^{\frac{1}{2}}(\bar{x})$ by virtue of the original data of (NLP) in several ways as follows.

Proposition 3.1 *$w \in \ker dS^{\frac{1}{2}}(\bar{x})$ if and only if one of the following conditions holds:*

(a) *$w \in L_C(\bar{x})$ and*

$$\left\langle \left[\sum_{i\in I}\rho_i\nabla^2 g_i(\bar{x}) + \sum_{j\in J}\rho_j\nabla^2 h_j(\bar{x})\right] w, w\right\rangle \le 0 \quad \forall \rho\in \mathrm{KKT}_0(\bar{x}).$$

(b) *$w \in L_C(\bar{x})$ and there exists some $z\in\Re^n$ such that*

$$\langle \nabla g_i(\bar{x}), z\rangle + \langle w, \nabla^2 g_i(\bar{x})w\rangle \le 0, \quad \forall i\in I(\bar{x}) \text{ with } \langle\nabla g_i(\bar{x}), w\rangle = 0,$$

$$\langle \nabla h_j(\bar{x}), z\rangle + \langle w, \nabla^2 h_j(\bar{x})w\rangle = 0, \quad \forall j\in J.$$

(c) *$w \in L_C(\bar{x})$ and*

$$\left\langle \left[\sum_{i\in I}\rho_i\nabla^2 g_i(\bar{x}) + \sum_{j\in J}\rho_j\nabla^2 h_j(\bar{x})\right] w, w\right\rangle \le 0 \quad \forall \rho\in \mathcal{A},$$

$$\left\langle \left[\sum_{j\in J}\rho_j\nabla^2 h_j(\bar{x})\right] w, w\right\rangle = 0 \quad \forall \rho\in \mathcal{B},$$

where $\mathcal{B}$ is a basis of the linear subspace $\{\rho\in\Re^{m+q} \mid \rho_i = 0\ \forall i\in I,\ \sum_{j\in J}\rho_j\nabla h_j(\bar{x}) = 0\}$ and $\mathcal{A}$ is finite subset of $\mathrm{KKT}_0(\bar{x})$ such that

$$\mathrm{KKT}_0(\bar{x}) = \mathrm{con}\,(\mathrm{pos}\,\mathcal{A}) + \mathrm{span}\,\mathcal{B}. \tag{24}$$

Proof. Let

$$F(w,\rho) := \left\langle \left[\sum_{i\in I}\rho_i\nabla^2 g_i(\bar{x}) + \sum_{j\in J}\rho_j\nabla^2 h_j(\bar{x})\right] w, w\right\rangle.$$

By (23), $w\in\ker dS^{\frac{1}{2}}(\bar{x})$ if and only if $w\in L_C(\bar{x})$ and

$$\max_{\rho\in \mathrm{KKT}_0(\bar{x})\cap\mathbb{B}} F(w,\rho) = 0. \tag{25}$$

As the function $F(w, \rho)$ is linear in ρ, (25) holds if and only if

$$\max_{\rho \in \mathrm{KKT}_0(\bar{x})} F(w, \rho) = 0. \tag{26}$$

That is, $w \in \ker dS^{\frac{1}{2}}(\bar{x})$ if and only if (a) holds. From the duality theorem of linear programming (see [15]), (26) holds if and only if there exists some $z \in \Re^n$ such that

$$\langle \nabla g_i(\bar{x}), z \rangle + \langle w, \nabla^2 g_i(\bar{x}) w \rangle \leq 0, \quad \forall i \in I(\bar{x}) \text{ with } \langle \nabla g_i(\bar{x}), w \rangle = 0,$$
$$\langle \nabla h_j(\bar{x}), z \rangle + \langle w, \nabla^2 h_j(\bar{x}) w \rangle = 0, \quad \forall j \in J.$$

This suggests that (a) and (b) are equivalent. By the linearity of $F(w, \rho)$ in ρ again, and the construction of $\mathcal{A}$ and $\mathcal{B}$, we obtain the equivalence of (a) and (c). This completes the proof. $\square$

Applying Theorem 2.2 and Proposition 3.1, we can characterize exact local sharp minima of the $\ell_{\frac{1}{2}}$ penalty function as follows.

Theorem 3.2 *The following properties are equivalent:*

(a) *The $\ell_{\frac{1}{2}}$ penalty function has an exact local sharp minimum at $\bar{x}$.*

(b) *The system of finitely many quadratic forms mixed with linear equations and inequalities*

$$\begin{aligned}
&\langle \nabla f(\bar{x}), w \rangle \leq 0,\\
&\langle \nabla g_i(\bar{x}), w \rangle \leq 0 \quad \forall i \in I(\bar{x}),\\
&\langle \nabla h_j(\bar{x}), w \rangle = 0 \quad \forall j \in J,\\
&\left\langle \left[\sum_{i \in I} \rho_i \nabla^2 g_i(\bar{x}) + \sum_{j \in J} \rho_j \nabla^2 h_j(\bar{x}) \right] w, w \right\rangle \leq 0 \quad \forall \rho \in \mathcal{A},\\
&\left\langle \left[\sum_{j \in J} \rho_j \nabla^2 h_j(\bar{x}) \right] w, w \right\rangle = 0 \quad \forall \rho \in \mathcal{B},
\end{aligned} \tag{27}$$

has a unique solution $w = 0$, where $\mathcal{A}$ and $\mathcal{B}$ are given as in Proposition 3.1 (c).

(c) *The following second-order sufficient condition holds:*

$$\sup_{\rho \in \mathrm{KKT}_0(\bar{x})} \left\langle \left[\sum_{i \in I} \rho_i \nabla^2 g_i(\bar{x}) + \sum_{j \in J} \rho_j \nabla^2 h_j(\bar{x}) \right] w, w \right\rangle \equiv +\infty > 0 \quad \forall w \in \mathcal{V}(\bar{x}) \backslash \{0\}, \tag{28}$$

where

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

denotes the critical cone of (NLP) at $\bar{x}$.

Remark 3.2 *Let the set of Fritz John multipliers of (NLP) at $\bar{x}$ be denoted by*

$$\mathrm{FJ}(\bar{x}) := \left\{ (\lambda_0, \rho) \in \Re \times \Re^{m+q} \;\middle|\; \begin{array}{c} \lambda_0 \nabla f(\bar{x}) + \sum_{i \in I} \rho_i \nabla g_i(\bar{x}) + \sum_{j \in J} \rho_j \nabla h_j(\bar{x}) = 0 \\ \rho_i = 0 \ \forall i \in I \backslash I(\bar{x}), \quad \rho_i \geq 0 \ \forall i \in I(\bar{x}) \\ \lambda_0 \geq 0, \quad (\lambda_0, \rho) \neq (0,0) \end{array} \right\}.$$

It has been shown in [3, Proposition 5.48] that, if

$$\sup_{(\lambda_0, \rho) \in \mathrm{FJ}(\bar{x})} \left\langle \left[\lambda_0 \nabla^2 f(\bar{x}) + \sum_{i \in I} \rho_i \nabla^2 g_i(\bar{x}) + \sum_{j \in J} \rho_j \nabla^2 h_j(\bar{x}) \right] w, w \right\rangle > 0 \quad \forall w \in \mathcal{V}(\bar{x}) \backslash \{0\}, \tag{29}$$

then $\bar{x}$ is a local minimum of (NLP) satisfying the second-order growth condition, i.e., there exist some $\alpha > 0$ and $\delta > 0$ such that

$$f(x) \geq f(\bar{x}) + \alpha \|x - \bar{x}\|^2 \quad \forall x \in C \cap B_\delta(\bar{x}). \tag{30}$$

In what follows, we consider three circumstances under which (29) may hold: (i) $\mathcal{V}(\bar{x}) = \{0\}$; (ii) $\mathcal{V}(\bar{x}) \neq \{0\}$ and $\mathrm{KKT}(\bar{x}) = \emptyset$; (iii) $\mathcal{V}(\bar{x}) \neq \{0\}$ and $\mathrm{KKT}(\bar{x}) \neq \emptyset$. In case (i), (29) holds automatically, and according to Theorem 3.1, it says nothing but that the ℓ_1 penalty function has an exact local sharp minimum at $\bar{x}$; In case (ii), it is easy to verify that (29) holds if and only if Theorem 3.2 (c) holds. From Theorems 3.1 and 3.2, it follows that (29) holds in case (ii) if and only if the ℓ_p penalty function has an exact local sharp minimum at $\bar{x}$ with $p = \frac{1}{2}$ but not with $p = 1$. And in case (iii), (29) reduces to the well-known second-order sufficient condition (see [13, 21] and also [3, Remark 5.49]):

$$\sup_{\rho \in \mathrm{KKT}(\bar{x})} \left\langle \left[\nabla^2 f(\bar{x}) + \sum_{i \in I} \rho_i \nabla^2 g_i(\bar{x}) + \sum_{j \in J} \rho_j \nabla^2 h_j(\bar{x}) \right] w, w \right\rangle > 0 \quad \forall w \in \mathcal{V}(\bar{x}) \backslash \{0\}.$$

According to [21, Corollary 4.5], the latter condition implies that the ℓ_1 penalty function with a finite penalty parameter has a local minimum at $\bar{x}$ satisfying the second-order growth condition, i.e., there exist some $\mu > 0$, $\alpha > 0$ and $\delta > 0$ such that

$$f(x) + \mu S(x) \geq f(\bar{x}) + \mu S(\bar{x}) + \alpha \|x - \bar{x}\|^2 \quad \forall x \in B_\delta(\bar{x}). \tag{31}$$

To sum up, corresponding to cases (i)-(iii), one of the following happens when (29) is fulfilled: (I) The ℓ_1 penalty function has an exact local sharp minimum at $\bar{x}$; (II) The ℓ_1 penalty function does not have an exact local sharp minimum at $\bar{x}$, but the $\ell_{\frac{1}{2}}$ penalty function does; (III) The ℓ_1 penalty function does not have an exact local sharp minimum at $\bar{x}$, but it has an exact local minimum satisfying the second-order growth condition (31).

3.2 Exact Local Sharp Minima with Constraint Qualifications

In this subsection, we will show that under a constraint qualification, a local sharp minimum of (NLP) can be an exact local sharp minimum of the ℓ_p penalty function with p ranging in an interval depending the constraint qualification assumed. To simplify our presentations, we will not utilize all constraint qualifications available, but only take advantage of some weakest possible constraint qualifications, such as the Guignard constraint qualification (for short, GCQ), and its generalized version, called the generalized Guignard constraint qualification (for short, g-GCQ).

The GCQ originating with [9] holds at $\bar{x}$ if by definition, $T_C(\bar{x})^* = L_C(\bar{x})^*$. Among various constraint qualifications ensuring KKT conditions, the GCQ is the weakest one in the sense that the GCQ holds at $\bar{x}$ if and only if the KKT condition holds at $\bar{x}$ for any objective function having a local minimum at $\bar{x}$ subject to the same constraints; see [8] for the original version of this result, and [22, Theorem 6.11] for some new features of this result. In the following, we introduce a generalized version of the GCQ.

Definition 3.1 *We say that the g-GCQ holds at $\bar{x}$ if $T_C(\bar{x})^* = [\ker dS^{\frac{1}{2}}(\bar{x})]^*$.*

Clearly, the g-GCQ is a consequence of the GCQ, but not vice versa. Taking the single inequality constraint $x^2 \leq 0$ for instance, the g-GCQ holds at $\bar{x} = 0$ as $T_C(\bar{x}) = \ker dS^{\frac{1}{2}}(\bar{x}) = \{0\}$, but the GCQ does not hold at $\bar{x} = 0$ as $L_C(\bar{x})^* = \{0\} \neq \Re = T_C(\bar{x})^*$.

Theorem 3.3 *Assume that the GCQ holds at $\bar{x}$. The following properties are equivalent:*

(a) *$\bar{x}$ is a local sharp minimum of (NLP);*

(b) *The ℓ_p penalty function with $0 \leq p \leq 1$ has an exact local sharp minimum at $\bar{x}$.*

These properties entail that $\mathrm{SPP}(f, S^p, \bar{x}) = 0$ *if* $p < 1$, *and*

$$\mathrm{SPP}(f, S, \bar{x}) = \min\{\|\lambda\|_\infty \mid \lambda \in \mathrm{KKT}(\bar{x})\}.$$

Proof. To begin with, we recall that $\bar{x}$ is a local sharp minimum of (NLP) if and only if (7) holds or equivalently by [22, Excercise 6.22], $-\nabla f(\bar{x}) \in \operatorname{int}(T_C(\bar{x})^*)$. From the GCQ at $\bar{x}$, it then follows that (a) holds if and only if $-\nabla f(\bar{x}) \in \operatorname{int}(L_C(\bar{x})^*)$. Thus, in the case of $p = 1$, (a) and (b) are equivalent, and hence the formula for $\mathrm{SPP}(f, S, \bar{x})$ follows from Theorem 3.1.

Now let $0 \le p < 1$. Observing that $0 \le S(x) \le S^p(x) \le S^0(x)$ for all x near $\bar{x}$, we have by the definition of subderivative,

$$T_C(\bar{x}) = \ker dS^0(\bar{x}) \subset \ker dS^p(\bar{x}) \subset \operatorname{dom} dS^p(\bar{x}) \subset \ker dS(\bar{x}) = L_C(\bar{x}).$$

By the GCQ at $\bar{x}$, we thus have $[\ker dS^p(\bar{x})]^* = [\operatorname{dom} dS^p(\bar{x})]^* = L_C(\bar{x})^*$. According to Theorem 2.1, we have $\widehat{\partial} S^p(\bar{x}) = L_C(\bar{x})^*$. In view of Theorem 2.2, all the results follow readily. $\square$

Theorem 3.4 *Assume that the g-GCQ holds at $\bar{x}$. The following properties are equivalent:*

(a) *$\bar{x}$ is a local sharp minimum of (NLP);*

(b) *The ℓ_p penalty function with $0 \le p \le \frac{1}{2}$ has an exact local sharp minimum at $\bar{x}$.*

These properties entail that $\mathrm{SPP}(f, S^p, \bar{x}) = 0$ *if* $p < \frac{1}{2}$.

Proof. As in the proof of Theorem 3.3, (a) holds if and only if $-\nabla f(\bar{x}) \in \operatorname{int}(T_C(\bar{x})^*)$. By the g-GCQ at $\bar{x}$, (a) holds if and only if $-\nabla f(\bar{x}) \in \operatorname{int}[\ker dS^{\frac{1}{2}}(\bar{x})]^*$ or equivalently by [22, Excercise 6.22],

$$\langle \nabla f(\bar{x}), w \rangle > 0 \quad \forall w \in \ker dS^{\frac{1}{2}}(\bar{x}) \backslash \{0\}.$$

In view of Theorem 2.2, (a) and (b) are equivalent in the case of $p = \frac{1}{2}$.

Now let $0 \le p < \frac{1}{2}$. Observing that $0 \le S^{\frac{1}{2}}(x) \le S^p(x) \le S^0(x)$ for all x near $\bar{x}$, we have by the definition of subderivative,

$$T_C(\bar{x}) = \ker dS^0(\bar{x}) \subset \ker dS^p(\bar{x}) \subset \operatorname{dom} dS^p(\bar{x}) \subset \ker dS^{\frac{1}{2}}(\bar{x}).$$

Thus, by the assumption that $T_C(\bar{x})^* = [\ker dS^{\frac{1}{2}}(\bar{x})]^*$, we have

$$[\ker dS^p(\bar{x})]^* = [\operatorname{dom} dS^p(\bar{x})]^* = [\ker dS^{\frac{1}{2}}(\bar{x})]^*.$$

According to Theorem 2.1, we thus have $\widehat{\partial} S^p(\bar{x}) = [\ker dS^{\frac{1}{2}}(\bar{x})]^*$. In view of Theorem 2.2, all the results follow readily. $\square$

4 Conclusions

In this paper, we conducted variational analysis on local sharp minima of a nonlinear programming problem via exact penalization. Utilizing some generalized differentiation tools such as subderivatives and regular subdifferentials, we first presented some primal and dual characterizations for exact local minima of a general penalty function associated with the nonlinear programming problem. We then applied these characterizations to the ℓ_p $(0 \le p \le 1)$ penalty function. By virtue of the original data of the nonlinear programming problem, we gave primal and dual characterizations for exact local sharp minima of the ℓ_p penalty function with $p \in (\frac{1}{2}, 1]$ and $p = \frac{1}{2}$ respectively. By assuming the Guignard constraint qualification (resp. the generalized Guignard constraint qualification), we showed that a local sharp minimum of the nonlinear programming problem can be an exact local sharp minimum of the ℓ_p penalty function with $p \in [0, 1]$ (resp. $p \in [0, \frac{1}{2}]$). Moreover, we gave some formulas for calculating the smallest penalty parameter for a penalty function to have an exact local sharp minimum.

Acknowledgements The authors would like to thank the two anonymous reviewers for their valuable comments and suggestions, which have helped to improve the paper. The first author was supported by the National Natural Science Foundation of China [11201383]. The second author was supported by the Research Grants Council of Hong Kong (PolyU 5295/12E), and the National Natural Science Foundation of China [11431004].

References

[1] F. Al-Khayyal and J. Kyparisis. Finite convergence of algorithms for nonlinear programs and variational inequalities. *Journal of Optimization Theory and Applications*, 70(2):319–332, 1991.

[2] A. Auslender. Stability in mathematical programming with nondifferentiable data. *SIAM Journal on Control and Optimization*, 22(2):239–254, 1984.

[3] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York, 2000.

[4] J. V. Burke. An exact penalization viewpoint of constrained optimization. *SIAM Journal on Control and Optimization*, 29(4):968–998, 1991.

[5] L. Cromme. Strong uniqueness. a far-reaching criterion for the convergence analysis of iterative procedures. *Numerische Mathematik*, 29:179–193, 1978.

[6] I. I. Eremin. The penalty method in convex programming. *Cybernetics and Systems Analysis*, 3(4):53–56, 1967.

[7] Y. P. Fang, K. W. Meng, and X. Q. Yang. Piecewise linear multicriteria programs: The continuous case and its discontinuous generalization. *Oper. Res.*, 60(2):398–409, March 2012.

[8] 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.

[9] M. Guignard. Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. *SIAM Journal on Control*, 7(2):232–241, 1969.

[10] R. Hettich. A review of numerical methods for semi-infinite optimization. In *Semi-infinite programming and applications*, pages 158–178. Springer, 1983.

[11] S. Howe. New conditions for exactness of a simple penalty function. *SIAM Journal on Control*, 11(2):378–381, 1973.

[12] X. X. Huang and X. Q. Yang. A unified augmented Lagrangian approach to duality and exact penalization. *Mathematics of Operations Research*, 28(3):533–552, 2003.

[13] 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.

[14] Z. Q. Luo, J. S. Pang, and D. Ralph. *Mathematical Pprograms with Equilibrium Constraints*. Cambridge University Press, Cambridge, 1996.

[15] O. L. Mangasarian. *Nonlinear Programming*. Mc Graw-Hill, New York, 1969.

[16] K. W. Meng, V. Roshchina, and X. Q. Yang. On local coincidence of a convex set and its tangent cone. *Journal of Optimization Theory and Applications*, 164(1):123–137, 2015.

[17] K. W. Meng and X. Q. Yang. Optimality conditions via penalty function. *SIAM Journal on Optimization*, 20(6):3208–3231, 2010.

[18] K. W. Meng and X. Q. Yang. First- and second-order necessary conditions via exact penalty functions. *Journal of Optimization Theory and Applications*, 165(3):720–752, 2015.

[19] B. T. Polyak. Sharp minima. Institute of control sciences lecture notes, Moscow, USSR, 1979. In *IIASA workshop on generalized Lagrangians and their applications, IIASA, Laxenburg, Austria*, 1979.

[20] B. T. Polyak. *Introduction to Optimization*. Optimization Software, Publications Division (New York), 1987.

[21] R. T. Rockafellar. Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives. *Mathematics of Operations Research*, 14:462–484, 1989.

[22] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer-Verlag, Berlin, 1998.

[23] A. M. Rubinov and X. Q. Yang. *Lagrange-Type Functions in Constrained Non-Convex Optimization*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[24] G. Still and M. Streng. Optimality conditions in smooth nonlinear programming. *Journal of Optimization Theory and Applications*, 90(3):483–515, 1996.

[25] M. Studniarski. Necessary and sufficient conditions for isolated local minima of nonsmooth functions. *SIAM Journal on Control and Optimization*, 24(5):1044–1049, 1986.

[26] M. Studniarski. Characterizations of strict local minima for some nonlinear programming problems. *Nonlinear Analysis: Theory, Methods & Applications*, 30(8):5363–5367, 1997.

[27] M. Studniarski and A. W. A. Taha. A characterization of strict local minimizers of order one for nonsmooth static minmax problems. *Journal of Mathematical Analysis and Applications*, 259(2):368 – 376, 2001.

[28] D. E. Ward. Characterizations of strict local minima and necessary conditions for weak sharp minima. *Journal of Optimization Theory and Applications*, 80(3):551–571, 1994.

[29] X. Q. Yang and Z. Q. Meng. Lagrange multipliers and calmness conditions of order p. *Mathematics of Operations Research*, 32(1):95–101, 2007.

[30] W. I. Zangwill. Non-linear programming via penalty functions. *Management Science*, 13(5):344–358, 1967.