Non-Lipschitzian Reformulation Method for Constrained Optimization Problems

XIAOQI YANG

Department of Applied Mathematics,
The Hong Kong Polytechnic University
Email: mayangxq@polyu.edu.hk

The (Fourth) South Pacific Continuous Optimization Meeting
University of South Australia, 8-13 February 2015
Outline

1. Non-Lipschitzian Reformulation
2. First-order Necessary Conditions and KKT-type Penalty Terms
3. Second-order Necessary Conditions
4. Interior-Point $\ell_p$-Penalty Method
5. Conclusions
1. Non-Lipschitzian Reformulation

Consider the nonlinear programming problem (NLP): 

\[ \begin{align*} 
& \min \quad f(x) \\
& \text{s.t.} \quad g_i(x) \leq 0, \quad i \in I := \{1, \ldots, m\}, \\
& \quad h_j(x) = 0, \quad j \in J := \{m + 1, \ldots, m + q\}, 
\end{align*} \]

where \( f, g_i, h_j : \mathbb{R}^n \to \mathbb{R} \) are assumed to be smooth functions.

We denote by \( C \) the feasible set of (NLP).
Karush-Kuhn-Tucker (KKT) condition for a local minimum $\bar{x}$ of (NLP), originated with Kuhn and Tucker (1951) and Karush (1939), holds if there exists a vector $\lambda \in \mathbb{R}^{m+q}$, called a KKT multiplier, such that

$$\nabla f(\bar{x}) + \sum_{i \in I} \lambda_i \nabla g_i(\bar{x}) + \sum_{j \in J} \lambda_j \nabla h_j(\bar{x}) = 0, \quad \lambda_i \geq 0, \quad \lambda_i g_i(\bar{x}) = 0 \quad \forall i \in I.$$
Second-Order Condition (SON), originated with Ioffe (1979), holds at a local minimum $\bar{x}$ of (NLP) if

$$
\sup_{\lambda \in \text{KKT}(\bar{x})} \langle w, \nabla^2_{xx} L(\bar{x}, \lambda) w \rangle \geq 0 \quad \forall w \in \mathcal{V}(\bar{x}),
$$

where $\text{KKT}(\bar{x})$ is the set of all KKT multipliers at $\bar{x}$ and the critical cone $\mathcal{V}(\bar{x})$ at $\bar{x}$ is defined by

$$
\mathcal{V}(\bar{x}) := \left\{ w \in \mathbb{R}^n \left| \begin{array}{c}
\langle \nabla f(\bar{x}), w \rangle \leq 0 \\
\langle \nabla g_i(\bar{x}), w \rangle \leq 0 \quad \forall i \in I(\bar{x}) \\
\langle \nabla h_j(\bar{x}), w \rangle = 0 \quad \forall j \in J
\end{array} \right. \right\}.
$$
Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ be a lower semicontinuous function such that

$$C = \{ x \in \mathbb{R}^n \mid \phi(x) = 0 \}.$$ 

The general penalty problem is

$$\min_{x \in \mathbb{R}^n} f(x) + \mu \phi(x).$$

**Definition 1.1** We say that the penalty function $f + \mu \phi$ is exact at $\bar{x}$ if, $f + \mu \phi$ admits a local minimum at $\bar{x}$ with some finite penalty parameter $\mu > 0$.

**Definition 1.2** We say that the penalty term $\phi$ is of KKT-type at $\bar{x}$ if the KKT condition holds at $\bar{x}$ whenever the penalty function $f + \mu \phi$ is exact at $\bar{x}$. 
Lower order penalty: Let $0 \leq p \leq 1$ and $0^0 := 0$. Let

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

while the $l_p$ penalty function associated with (NLP) is of the form

$$\mathcal{F}_p(x) := f(x) + \mu S^p(x).$$

- $p = 1$, $\mathcal{F}_1(x)$ is the classical $l_1$ penalty function, see Eremin (1967) and Zangwill (1967).

- $p < 1$, $\mathcal{F}_p(x)$ is referred to as the lower order $l_p$ penalty function, first introduced in Luo et al. (1996) for the study of MPEC and was rediscovered from a unified augmented Lagrangian scheme by Huang and Yang (2003) and Rubinov and Yang (2003).

The $l_p$ penalty problem (NLP($p$)) is

$$\min_{x \in R^n} f(x) + \mu S^p(x).$$
$l_1$ exact penalty function implies the KKT condition and the SON condition, see Clarke (1983) and Rockafellar (1989), respectively.

Can $l_p(p < 1)$ exact penalty function be used for the KKT condition and the SON condition?
2. **First-order Necessary Conditions and KKT-type Penalty Terms**

If $q = 0$ and all $g_i$’s are concave, then the KKT condition of (NLP) holds. See Hestenes (1970).
The generalized Clarke second-order directional derivative of a $C^{1,1}$ function is defined by

$$g^{oo}(x; w) = \limsup_{y \to x, t \to 0^+} \frac{\nabla g(y + tu)^T w - \nabla g(y)^T w}{t},$$


A $C^{1,1}$ function $g$ is concave if and only if

$$g^{oo}(x; w) \leq 0, \forall x, w \in \mathbb{R}^n.$$

Let

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

\[ I(\bar{x}, w) : = \{ i \in I \mid g_i(\bar{x}) = 0, \langle \nabla g_i(\bar{x}), w \rangle = 0 \}. \]

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

\[
L_C(\bar{x}) := \left\{ w \in \mathbb{R}^n \left| \begin{array}{c}
\langle \nabla g_i(\bar{x}), w \rangle \leq 0 \ \forall i \in I(\bar{x}) \\
\langle \nabla h_j(\bar{x}), w \rangle = 0 \ \forall j \in J
\end{array} \right. \right\}.
\]
By calculating the upper Dini-derivative of $\mathcal{F}_{1/2}(x)$ and using a second-order Taylor expansion, Yang and Meng (2007) showed that $S^\frac{1}{2}(x)$ is of KKT-type at $\bar{x}$, that is the KKT condition holds, if, for every $w \in L_C(\bar{x})$,

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

These results have been extended to semi-infinite program and generalized semi-infinite program and the paper is submitted to a JOTA special issue dedicated to Elijah (Lucien) Polak’s 85th birthday. See Yang, Chen and Zhou (2015).
By calculating the contingent directional derivative of $F_{1/2}(x)$ at $\bar{x}$, Meng and Yang (2010) showed that if, for every $w \in L_C(\bar{x})$, there exists some $z \in R^n$ such that

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

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

then the KKT condition holds.
Meng and Yang (2015) employ the following tools from Variational Analysis, see Rockafellar and Wets (1998). For any $f : \mathbb{R}^n \to \mathbb{R}$ and a point $\bar{x}$ with $f(\bar{x})$ finite,

- For any $w \in \mathbb{R}^n$, the subderivative of $f$ at $\bar{x}$ for $w$ is defined by
  \[
  df(\bar{x})(w) := \liminf_{\tau \to 0^+, w' \to w} \frac{f(\bar{x} + \tau w') - f(\bar{x})}{\tau}.
  \]

- A vector $v \in \mathbb{R}^n$ is a regular subgradient of $f$ at $\bar{x}$, written $v \in \hat{\partial} f(\bar{x})$, if
  \[
  f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(\|x - \bar{x}\|).
  \]

- The relation between subderivative and regular subdifferential is
  \[
  \hat{\partial} f(\bar{x}) = \{ v \in \mathbb{R}^n \mid \langle v, w \rangle \leq df(\bar{x})(w) \forall w \in \text{dom} df(\bar{x}) \}.\]
Lemma 2.1 Suppose that the function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ has a local minimum at $\bar{x}$ with $\psi(\bar{x})$ finite. Then we have\(^a\)

$$[\text{dom} \, \text{d} \psi(\bar{x})]^* \subset \hat{\partial} \psi(\bar{x}) \subset [\text{ker} \, \text{d} \psi(\bar{x})]^*.$$  

- The first inclusion is an equality if and only if $\hat{\partial} \psi(\bar{x})$ is a cone;
- The second inclusion is an equality if and only if
  $$[\text{dom} \, \text{d} \psi(\bar{x})]^* = [\text{ker} \, \text{d} \psi(\bar{x})]^*.$$  

- If the subderivative $\text{d} \psi(\bar{x})$ is a sublinear function as is true when $\psi$ is regular at $\bar{x}$ (see Definition 7.25 of Rockafellar and Wets (1998)), then
  $$\text{clpos}(\hat{\partial} \psi(\bar{x})) = [\text{ker} \, \text{d} \psi(\bar{x})]^*. \tag{1}$$

\(^a\)The polar cone of $A$ is defined by

$$A^* = \{ z \in \mathbb{R}^n | \langle z, x \rangle \leq 0 \, \forall \, x \in A \}.$$
The closure operation cannot be removed from the left-hand side of (1) even if $\psi$ is convex.

**Example 2.1** Consider at $\bar{x} = (0, 0)$ the function

$$\psi(x) = \max_{0 \leq t \leq 1} g(x, t),$$

where $g(x, t) = tx_1 + t^2x_2$ for all $x \in \mathbb{R}^2$ and $t \in \mathbb{R}$. The equality

$$\text{clpos}(\hat{\partial}\psi(\bar{x})) = [\ker \partial\psi(\bar{x})]^* (= \{ x \mid 0 \leq x_2 \leq x_1 \})$$

holds, but when the closure operation is removed from the left-hand side, we merely have

$$\text{pos}(\hat{\partial}\psi(\bar{x})) \subset [\ker \partial\psi(\bar{x})]^*,$$

because

$$\text{pos}(\hat{\partial}\psi(\bar{x})) = \{ x \mid 0 \leq x_2 \leq x_1 \}\{ x \mid x_1 > 0, x_2 = 0 \}.$$
It is well-known, see Rockafellar and Wets (1998) that, for a polyhedral set $P$ at one of its points $\bar{x}$, there exists a neighborhood $V$ of $\bar{x}$ such that

$$P \cap V = [\bar{x} + T_P(\bar{x})] \cap V.$$ 

We now introduce such a property for a convex set in an analogous way.

**Definition 2.1** We say that a convex set $C \subset \mathbb{R}^n$ admits exactness of tangent approximation (ETA, for short) at one of its points $\bar{x}$, if $\exists$ a neighborhood $V$ of $\bar{x}$ such that

$$(clC) \cap V = [\bar{x} + T_C(\bar{x})] \cap V.$$ 

**Proposition 2.1** (Meng, Roshchina and Yang (2015)) Let $C \subset \mathbb{R}^n$ be closed and convex. The following properties are equivalent:

(i) $C$ is locally polyhedral at every $x \in C$, i.e., $(C - \{x\}) \cap V$ is a polyhedron for some polyhedral neighbourhood $V$ of $x$.

(ii) $C$ admits ETA at every $x \in C$.

(iii) $pos(C - x)$ is closed for all $x \in C$. 

We recall the **variational description of regular subgradients**:

**Lemma 2.2** *(Rockafellar and Wets (1998), Proposition 8.5).* A vector $v$ belongs to $\hat{\partial} f(\bar{x})$ 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 $\bar{x}$.

**Remark 2.1** This variational description is a contribution to the basics of variational analysis, as pointed out on p.347 of *Rockafellar and Wets (1998).*
We can obtain from Lemmas 2.1 and 2.2 the following.

**Theorem 2.1 (Meng and Yang (2015))** Consider the following conditions:

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

(ii) \(\hat{\partial}\phi(\bar{x}) \subset L_C(\bar{x})^*\).

(iii) The penalty term \(\phi\) is of KKT-type at \(\bar{x}\).

Then \((i) \implies (ii) \iff (iii)\).
[(iii) \implies (ii)]: Let \( v \in \hat{\partial} \phi(\bar{x}) \). According to the variational description of regular subgradients in (Rockafellar and Wets, 1998, Proposition 8.5), there exist a neighborhood \( V \) of \( \bar{x} \) and a continuously differentiable function \( \psi : \mathbb{R}^n \to \mathbb{R} \) with \( \psi(\bar{x}) = \phi(\bar{x}) = 0 \) and \( \nabla \psi(\bar{x}) = v \) such that

\[
\psi(x) \leq \phi(x), \quad \forall x \in V.
\]

Set \( f = -\psi \). Clearly,

\[
f(x) + \phi(x) = -\psi(x) + \phi(x) \geq 0 = f(\bar{x}) + \phi(\bar{x}), \quad \forall x \in V.
\]

That is, the penalty function \( f + \phi \) admits a local minimum at \( \bar{x} \). Since \( \phi \) is a KKT-type penalty term at \( \bar{x} \), the KKT condition holds at \( \bar{x} \). Thus \( -\nabla f(\bar{x}) \in L_C(\bar{x})^* \). Thus \( v \in L_C(\bar{x})^* \), and

\[
\hat{\partial} \phi(\bar{x}) \subseteq L_C(\bar{x})^*.
\]
Theorem 2.2 (Meng and Yang (2015)) Let $0 \leq p < 1$. Consider the following conditions:

(i) $\ker dS^p(\bar{x})^* = L_C(\bar{x})^*$.

(ii) $\hat{\partial}S^p(\bar{x}) = L_C(\bar{x})^*$.

(iii) $S^p$ is a KKT-type penalty term at $\bar{x}$.

Then (i) $\implies$ (ii) $\iff$ (iii).

• In the case of $p = 0$, (i) and (ii) are equivalent, and moreover Theorem 2.2 recovers a well-known result that the GCQ, i.e.

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

is the weakest one ensuring KKT conditions, as $\ker dS^0(\bar{x}) = T_C(\bar{x})$. 
• In the case of $0 < p < 1$, we are not aware of the equivalence of (i) and (ii), although they are the same in many situations.
The degenerate KKT multiplier set at \( \bar{x} \) is defined by

\[
\text{KKT}_0(\bar{x}) := \left\{ \rho \left| \begin{array}{c}
\sum_{i \in I} \rho_i \nabla g_i(\bar{x}) + \sum_{j \in J} \rho_j \nabla h_j(\bar{x}) = 0 \\
\rho_i \geq 0 \quad \forall i \in I, \\
\rho_i = 0 \quad \forall i \in I \setminus I(\bar{x})
\end{array} \right. \right\}
\]

The second subderivative of \( f \) at \( \bar{x} \) for \( v \) and \( w \) is defined by, see Rockafellar and Wets (1998)

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

By definition, it is straightforward to verify that

\[
d^2 S(\bar{x}|0)(w) = 2[dS^{\frac{1}{2}}(\bar{x})(w)]^2.
\]
By a direct calculation using the chain rule for second subderivatives of piecewise linear-quadratic functions \(^a\), we have

\[ dS^{\frac{1}{2}}(\bar{x})(w) = +\infty \quad \forall w \notin L_C(\bar{x}), \]

and if \( w \in L_C(\bar{x}) \), we have

\[
\begin{align*}
&dS^{\frac{1}{2}}(\bar{x})(w) \\
&= \sqrt{\frac{1}{2} d^2 S(\bar{x}|0)(w)} \\
&= \sqrt{\frac{2}{2}} \sqrt{\max_{\rho \in \text{KKT}_0(\bar{x}), \|\rho\|_{\infty} = 1} \left\langle \left\{ \sum_{i \in I} \rho_i \nabla^2 g_i(\bar{x}) + \sum_{j \in J} \rho_j \nabla^2 h_j(\bar{x}) \right\} w, w \right\rangle},
\end{align*}
\]

\[ \hat{\partial} S^{\frac{1}{2}}(\bar{x}) = \{ v \mid \langle v, w \rangle \leq dS^{\frac{1}{2}}(\bar{x})(w) \quad \forall w \}. \]

But we cannot give an explicit formula for \( \hat{\partial} S^{\frac{1}{2}}(\bar{x}) \).

\(^a\)See Chapter 13 of Rockafellar and Wets (1998).
Proposition 2.2 \( S^{1/2} \) is of KKT-type at \( \bar{x} \) if one of the two following conditions is satisfied:

(i) For every \( w \in L_C(\bar{x}) \), it follows that

\[
\max_{\lambda \in \text{KKT}_0(\bar{x})} \left\{ \sum_{i \in I} \lambda_i \langle w, \nabla^2 g_i(\bar{x}) w \rangle + \sum_{j \in J} \lambda_j \langle w, \nabla^2 h_j(\bar{x}) w \rangle \right\} = 0. \tag{2}
\]

(ii) \( \ker dS^{1/2}(\bar{x}) = L_C(\bar{x}) \).

Moreover, we have (i) \( \iff \) (ii).
Condition (2) is newly obtained, and we have

\[ \text{MFCQ} \Rightarrow (2), \]

because the MFCQ at \( \bar{x} \iff \text{KKT}_0(\bar{x}) = \{0\} \).

**Example 2.2** Let \( \bar{x} = (0, 0) \) and let

\[ C = \left\{ x \in \mathbb{R}^n \left| \begin{array}{l} x_1^2 x_2 \leq 0 \\ x_2^2 - x_1 \leq 0 \end{array} \right. \right\}. \]

(2) holds and \( \text{KKT}_0(\bar{x}) = \mathbb{R}_+ \times \{0\} \).
\[ T_C(\bar{x}) = R_+ \times (-R_+), \quad L_C(\bar{x}) = R_+ \times R, \text{ and} \]

\[
\text{ker}dS^p(\bar{x}) = \begin{cases} 
R_+ \times (-R_+) & \text{if } 0 < p \leq \frac{1}{5}, \\
R_+ \times (-R_+) \cup \{0\} \times R_+ & \text{if } \frac{1}{5} < p \leq \frac{1}{3}, \\
R_+ \times R & \text{if } \frac{1}{3} < p \leq 1.
\end{cases}
\]
3. Second-order Necessary Conditions via Exact Penalty Functions

Denote the set of all KKT multipliers at $\bar{x}$ by $\text{KKT}(\bar{x})$ and the critical cone at $\bar{x}$ by

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

The second-order necessary condition (for short, SON), originated with Ioffe (1979), holds at a local minimum $\bar{x}$ of (NLP) if

$$\sup_{\lambda \in \text{KKT}(\bar{x})} \langle w, \nabla^2_{xx} L(\bar{x}, \lambda) w \rangle \geq 0 \quad \forall w \in \mathcal{V}(\bar{x}),$$

where the convention $\sup \emptyset := -\infty$ is used.
The parabolic subderivative of $f$ at $\bar{x}$ for $w$ with respect to $z$ is defined by, see Rockafellar and Wets (1998)

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

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

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

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

$$L^2_C(\bar{x} | w) := \left\{ z \left| \begin{array}{c} \langle \nabla g_i(\bar{x}), z \rangle + \langle w, \nabla^2 g_i(\bar{x}) w \rangle \leq 0 \quad \forall i \in I(\bar{x}, w) \\ \langle \nabla h_j(\bar{x}), z \rangle + \langle w, \nabla^2 h_j(\bar{x}) w \rangle = 0 \quad \forall j \in J \end{array} \right. \right\}.$$
Let \( w \in T_A(\bar{x}) \). The second-order tangent set to \( A \) at \( \bar{x} \) is

\[
T_A^2(\bar{x} \mid w) := \{ z \mid \exists t_k \downarrow 0 \text{ and } z_k \rightarrow z \text{ such that } \bar{x} + t_k w + \frac{1}{2} t_k^2 z_k \in A \text{ for all } k \}.
\]
\( l_1 \text{ exactness } \implies \text{(SON)}. \) See Corollary 4.5 of Rockafellar (1989), where more general results on second-order necessary conditions were obtained for a convex composite optimisation problem by virtue of twice epi-derivative and a basic constraint qualification.
For the $l_1$ exact penalty function, we can show

$$L_C^2(\bar{x} \mid w) = \ker d^2 S(\bar{x})(w \mid \cdot) \quad \forall w \in L_C(\bar{x}),$$

by applying a second-order Taylor expansion.
On the other hand, Kawasaki (1988) introduced the following second-order Guinard constraint qualification (SGCQ)

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

As we have

\[ T_C^2(\bar{x} \mid w) = \text{ker}d^2S^0(\bar{x})(w \mid \cdot), \quad \forall w \in T_C(\bar{x}), \]

the SGCQ reduces to

\[ L_C^2(\bar{x} \mid w) = \text{clconv}[\text{ker}d^2S^0(\bar{x})(w \mid \cdot)] \quad \forall w \in \mathcal{V}(\bar{x}). \]
So, we are looking at the following second-order constraint qualification:

\[ L_C^2(\bar{x} | w) \subset \text{clconv} [\ker d^2 \phi(\bar{x})(w | \cdot)] \quad \forall w \in \mathcal{V}(\bar{x}). \]
Theorem 3.1 (Meng and Yang (2015)) Let $\bar{x}$ be a local minimum of (NLP). Suppose that the penalty function $f + \mu \phi$ is exact at $\bar{x}$. If

$$L_C^2(\bar{x} \mid w) \subset \text{clconv} [\ker d^2 \phi(\bar{x})(w \mid \cdot)] \quad \forall w \in \mathcal{V}(\bar{x}),$$

then the SON condition holds, and in particular when $L_C^2(\bar{x} \mid w) = \emptyset$, the supremum in the SON condition is $+\infty$. 
Let $\bar{x} \in C$ and let $\phi = S^p$.

We shall give sufficient conditions in terms of the original data for the inclusion

$$L^2_C(\bar{x} \mid w) \subset \ker d^2 S^p(\bar{x})(w \mid \cdot) \quad \forall w \in L_C(\bar{x})$$

(4)

to hold, which is slightly stronger than (3) since in general $\ker d^2 S^p(\bar{x})(w \mid \cdot)$ is not a closed and convex set and $\mathcal{V}(\bar{x}) \subsetneq L_C(\bar{x})$.  


Theorem 3.2 (Meng and Yang (2015)) Let $\bar{x}$ be a local minimum of (NLP). Suppose that the $l_p$ penalty function is exact at $\bar{x}$. If, in addition, one of the following conditions is satisfied:

(i) $p \in \left(\frac{2}{3}, 1\right]$,

(ii) $p = \frac{2}{3}$ and, for every $z \in L^2_C(\bar{x} \mid w)$, it follows that

\[
\begin{align*}
    \langle w, \nabla^2 g_i(\bar{x}) z \rangle + \frac{1}{3} g_i^{(3)}(\bar{x})(w, w, w) &\leq 0 \quad \forall i \in I(\bar{x}, w, z), \\
    \langle w, \nabla^2 h_j(\bar{x}) z \rangle + \frac{1}{3} h_j^{(3)}(\bar{x})(w, w, w) &= 0 \quad \forall j \in J,
\end{align*}
\]

(iii) $p \in \left[0, \frac{2}{3}\right)$, $q = 0$ (i.e., there is no equality constraint) and, for every $z \in L^2_C(\bar{x} \mid w)$ with $(w, z) \neq 0$, it follows that

\[
\langle w, \nabla^2 g_i(\bar{x}) z \rangle + \frac{1}{3} g_i^{(3)}(\bar{x})(w, w, w) < 0 \quad \forall i \in I(\bar{x}, w, z),
\]

then (4) holds and so does the SON condition.
4. **Interior-Point $\ell_p$-Penalty Method**

Consider the $\ell_p (p \leq 1)$ penalty problem of the following form with inequality constraints only

$$\min_{x \in \mathbb{R}^n} \rho f(x) + \sum_{i \in I} (\max\{g_i(x), 0\})^p.$$

We introduce the following $p$-order relaxation constrained problem:\n
$$\text{(RCP)} \min_{x, s} \phi_p(x, s; \rho) := \rho f(x) + \sum_{i \in I} s_i$$

\[\text{s.t. } s_i \geq 0 \text{ and } s_i^{1/p} - g_i(x) \geq 0, \ i \in I.\]

- (RCP) shares the same differentiability as (NLP);
- It can be shown that the $\ell_p$ penalty problem for (RCP) is always exact. For the case $p = 1$, see Curtis (2010).

\[^a\text{See Tian, Yang and Meng (2014), Interior-point } l_{1/2} \text{ penalty function method, JIMO (to appear)}\]
The primal-dual interior-point method is used to solve the $p$-order relaxation problem (RCP), which is to solve a sequence of logarithmic barrier subproblems

\[
\text{(LBCP)} \quad \min_{x,s} \rho f(x) + \sum_{i \in I} s_i - \mu \sum_{i \in I} \log s_i - \mu^{1/p} \sum_{i \in I} \log \left( s_i^{1/p} - g_i(x) \right)
\]

subject to $s_i > 0$ and $s_i^{1/p} - g_i(x) > 0$, $i \in I$,

where $\mu > 0$ is the barrier parameter.

- Barrier parameter $\mu^{1/p}$ is set for the term $\sum_{i \in I} \log \left( s_i^{1/p} - g_i(x) \right)$, which provides better numerical results than $\mu$ and can be justified by the first-order conditions.
The first-order necessary conditions of the barrier subproblem (LBCP) are given as follows

\begin{align}
\rho \nabla f(x) + A(x)y &= 0, \quad (5a) \\
e - 1/pY s^{1/p-1} - u &= 0, \quad (5b) \\
Y(s^{1/p} - c(x)) - \mu^{1/p}e &= 0, \quad (5c) \\
Us - \mu e &= 0. \quad (5d)
\end{align}

where \(y, u \in \mathbb{R}^m\) are Lagrange multipliers, \(Y := \text{diag}(y), U := \text{diag}(u)\) and \(A(x) := [\nabla g_1(x), \cdots, \nabla g_m(x)]\).

- Modified Newton method is used for finding a search direction, see Benson, Shanno and Vanderbei (2004).
Numerical Experiments

- We refer to our algorithm as IPLOP method, which stands for the Interior-Point Lower-Order Penalty method;
- We use 266 inequality constrained problems from the CUTEr collection, COPS, MITT and GLOBAL Library test sets as our test problems;
- The existing interior-point $\ell_1$-penalty method (PIPAL-a and PIPAL-c methods in PIPAL1.0 developed by Curtis (2010)) is used to compare the performance with the proposed method;
Using the performance profiles of Dolan and Moré (2002), we plot the following figures. For example, the plots $\pi_s(\tau)$ in the left one denote the scaled performance profile

$$\pi_s(\tau) := \frac{\text{no. of problems } \hat{p} \text{ where } \log_2(r_{\hat{p},s}) \leq \tau}{\text{total no. of problems}}, \quad \tau \geq 0,$$

where $\log_2(r_{\hat{p},s})$ is the scaled performance ratio between the iteration number to solve problem $\hat{p}$ by solver $s$ over the fewest iteration number required by the solvers of the IPLOP method with different $p$. It is clear that $\pi_s(\tau)$ is the probability for solver $s$ that a scaled performance ratio $\log_2(r_{\hat{p},s})$ is within a factor $\tau \geq 0$ of the best possible ratio.
- The left one is plotted by the number of iterations;
- The right one is plotted by the values of \( \frac{1}{\rho} \).
• The left one is plotted by the number of iterations;

• The right one is plotted by the values of $\frac{1}{\rho}$. 
5. Conclusions

In this talk, we partly answer the question as whether and how optimality conditions of NLPs can be derived from exactness of penalty functions.

- We define KKT-type penalty terms, and give their characterizations and some sufficient conditions.
- We derive the SON condition from exactness of penalty functions.
- We design an interior point $l_p$ penalty method.
References


