First- and Second-Order Necessary Conditions via Lower-order Exact Penalty Functions

Xiaoqi Yang

Department of Applied Mathematics
The Hong Kong Polytechnic University

Joint with Zhangyou Chen (SWJTU), Kaiwen Meng (SWJTU), Zhiqing Meng (ZJUTCEM).

April 1, 2019

Outline

Consider the nonlinear programming problem (NLP):

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

where $f, g_i, h_j : R^n \to R$ are assumed to be smooth functions.

KKT conditions, originated with [**?**] and [**?**], are the well-known first-order necessary conditions for local minima of (NLP).

KKT conditions are useful in the design of optimal algorithms as one can compute a KKT point at most.

We denote by C the feasible set and by $\bar{S}$ the set of optimal solutions of (NLP).

$$\left.\begin{array}{c}\text{(NLP) has a local minimum at } \bar{x} \\ \text{plus a constraint qualification}\end{array}\right\} \Longrightarrow \text{the following KKT condition a}$$

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \mu_i \nabla g_i(\bar{x}) + \sum_{j \in J} \eta_j \nabla h_j(\bar{x}) = 0 \quad (\mu_i \geq 0).$$

Constraint qualifications include:

- LICQ [**?**]
- MFCQ [**?**]
- ACQ [**?**]
- GCQ (weakest CQ) [**?**]

Literature review

Another approach to study optimality conditions is by virtue of exact penalty functions. [**?**] and [**?**] used l_1 exact penalty functions to derive KKT necessary optimality conditions.

On the other hand, [**?**] and [**?**] used $l_p(p \in [0, 1])$ exact penalty functions (see [**?**] to derive KKT necessary optimality conditions together with some nonpositivity condition on the second-order directional derivative of the constraints.

Let $0 \leq p \leq 1$, $0^0 := 0$ and $g_{i+}(x) = \max\{g_i(x), 0\}$. A particular penalty term associated with (NLP) is of the form

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

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

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

- $p = 1$, the classical l_1 penalty function, see [**?**] and [**?**].
- $p < 1$, referred to as the lower order l_p penalty function, first introduced in [**?**] for the study of MPEC and was rediscovered from a unified augmented Lagrangian scheme by [**?**] and [**?**].

A penalty function is said to be exact if any optimal solution of (NLP) is also one for the penalty problem.

By definition, $\mathcal{F}_0$ is exact at any local minimum of (NLP). It was shown in [**?**] that $\mathcal{F}_p$ with $0 < p \leq 1$ is exact if and only if the following generalized calmness-type condition holds:

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

where $\beta(u)$ is the optimal value of the optimization problem

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

When $p = 1$, this result was established in [**?**] and [**?**].

Let f be locally Lipschitz. If the following error bound condition holds

$$\tau d(x, \bar{S}) \leq S^p(x), \ x \in X$$

then $\mathcal{F}_p(x)$ is an exact penalty function.

The exact penalty function plays a key role in deriving KKT conditions.

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

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

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

The Dini upper directional derivative of a function $\phi : R^n \to R$ at $x \in R^n$ in the direction $u \in R^n$ is defined by

$$D_+\phi(x; u) = \limsup_{t \to 0+} \frac{\phi(x + tu) - \phi(x)}{t}.$$

The generalized Clarke second-order directional derivative of a $C^{1,1}$ function is

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

If $F_p(x) = f(x) + \mu \sum_{i \in I} g_{i+}^p(x)$ is exact at $\bar{x}$, then

$$D_+ F_p(\bar{x}; u) \geq 0, \quad \forall u \in R^n.$$

Thus

$$\nabla f(\bar{x})^\top u + \mu \sum_{i \in I} D_+ g_{i+}^p(\bar{x}; u) \geq 0, \quad \forall u \in R^n.$$

Then,

$$\sum_{i \in I} D_+ g_{i+}^p(\bar{x}; u) \leq 0 \Longrightarrow \nabla f(\bar{x})^\top u \geq 0, \quad \forall u \in R^n.$$

By Farkas lemma, which says that exactly one of the following two systems has a solution:

System 1 $\quad Au \leq 0,\ c^\top u > 0,\ $ for some u,

System 2 $\quad A^\top \mu = c,\ \mu \geq 0,\ $ for some μ,

we establish that the following KKT condition holds:

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \mu_i \nabla g_i(\bar{x}) = 0 \quad (\mu_i \geq 0).$$

The case $p \in (0,1)$.

Lemma

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

(i) *If* $h(x) < 0$, *then* $D_+\bar{h}(x; d) = 0$;

(ii) *If* $h(x) = 0$ *and* $\langle \nabla h(x), d \rangle < 0$, *then* $D_+\bar{h}(x; d) = 0$;

(iii) *If* $p \in (0.5, 1)$, $h(x) = 0$, $\langle \nabla h(x), d \rangle = 0$ *and* $h^{\circ\circ}(x; d)$ *is finite, then* $D_+\bar{h}(x; d) = 0$;

(iv) *If* $p = 0.5$, $h(x) = 0$ *and* $\langle \nabla h(x), d \rangle = 0$, *then* $D_+\bar{h}(x; d) \leq \sqrt{\max\{\frac{1}{2} h^{\circ\circ}(x; d), 0\}}$;

(v) *If* $p \in (0, 0.5)$, $h(x) = 0$, $\langle \nabla h(x), d \rangle = 0$ *and* $h^{\circ\circ}(x; d) < 0$, *then* $D_+\bar{h}(x; d) = 0$.

By estimating the upper Dini-directional derivative of $\mathcal{F}_p(x)$, we have

Theorem

[?] If $F_p(x)$ is exact at $\bar{x}$ and in addition, one of the following conditions is satisfied:

(i) $p \in (\frac{1}{2}, 1]$, *$g_i (i \in I)$ and $h_j (j \in I)$ are $C^{1,1}$,*

(ii) $p = \frac{1}{2}$ *and, for every $w \in L_C(\bar{x})$, it follows that*

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

(iii) $p \in [0, 1/2)$, $q = 0$ *(i.e., there is no equality constraint) and, for every $w \in L_C(\bar{x})$ with $w \neq 0$, it follows that*

$$g_i^{\circ\circ}(\bar{x}; w) < 0, \qquad \forall i \in I(\bar{x}, w),$$

then $\mathrm{KKT}(\bar{x}) \neq \emptyset$.

Let $M : R^n \rightrightarrows R^s$ be a set-valued map and $(x, y) \in gphM$. The contingent derivative of M at (x, y) is defined by the set-valued map $DM(x, y) : R^n \rightrightarrows R^s$ such that

$$gph(DM(x, y)) = T_{gphM}(x, y).$$

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

$$KerDM(x) = \{u \in R^n \mid 0 \in DM(x)(u)\}.$$

Now, define an optimality indication set of (NLP) with respect to C and $\bar{x}$ as follows:

$$\Pi(C,\bar{x}) := \{p \in [0,1] \mid KerDS^p(\bar{x})^* \subset KerDS(\bar{x})^*\}.$$

By estimating the contingent derivative of $\mathcal{F}_p(x)$, we have

Theorem

[?] If there exists $p \in \Pi(C,\bar{x})$ such that the l_p penalty function $\mathcal{F}_p$ is exact at $\bar{x}$, then $\mathrm{KKT}(\bar{x}) \neq \emptyset$.

In what follows,

- we distinguish a point $\bar{x} \in C$ for consideration;
- let $\phi : R^n \to R_+ \cup \{+\infty\}$ be a lower semicontinuous function such that

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

- ϕ is called a **penalty term** associated with (NLP)
- The function of the form

$$f + \mu\phi$$

is called a **penalty function** associated with (NLP), where μ, a positive number, is often referred to as the **penalty parameter**.

Definition

We say that the penalty term ϕ 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 $\bar{x}$.

Theorem

Consider the following conditions:

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

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

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

Then (i) $\Longrightarrow$ (ii) $\Longleftrightarrow$ (iii).

Theorem

Let $0 \leq p < 1$. *Consider the following conditions:*

(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* $\bar{x}$.

Then (i) $\Longrightarrow$ (ii) $\Longleftrightarrow$ (iii).

In what follows,

- we distinguish a point $\bar{x} \in C$ for consideration;
- let $\phi : R^n \to R_+ \cup \{+\infty\}$ be a lower semicontinuous function such that

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

- ϕ is called a **penalty term** associated with (NLP)
- The function of the form

$$f + \mu\phi$$

is called a **penalty function** associated with (NLP), where μ, a positive number, is often referred to as the **penalty parameter**.

Definition

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**.

[exactness of penalty function at $\bar{x} \Longrightarrow \bar{x}$ being a local minimum of (NLP)]

It is well-known that [1]

$$\mathcal{F}_p \text{ with } p = 1 \text{ is exact at } \bar{x} \Longrightarrow \text{KKT condition at } \bar{x}.$$

But in general,

$$\mathcal{F}_p \text{ with } 0 < p < 1 \text{ is exact at } \bar{x} \not\Longrightarrow \text{KKT condition at } \bar{x}.$$

[1] See Theorem 4.8 of [**?**].

Definition

We say that the penalty term ϕ 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 $\bar{x}$.

We will employ the tools from Variational Analysis, see [**?**].
For any $f : R^n \to \overline{R}$ and a point $\bar{x}$ with $f(\bar{x})$ finite,

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

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

- For any $w \in R^n$, the subderivative (or Hadamard directional derivative) of f at $\bar{x}$ for w is defined by

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

-

$$\widehat{\partial} f(\bar{x}) = \{v \in R^n \mid \langle v, w \rangle \leq df(\bar{x})(w)\ \forall w \in \mathrm{dom} df(\bar{x})\}.$$

Lemma

Suppose that the function $\psi : R^n \to \overline{R}$ has a local minimum at $\bar{x}$ with $\psi(\bar{x})$ finite. Then we have

$$[\mathrm{dom} d\psi(\bar{x})]^* \subset \widehat{\partial}\psi(\bar{x}) \subset [\ker d\psi(\bar{x})]^*. \tag{1}$$

Moreover,

- *The first inclusion in (1) is an equality if and only if the regular subdifferential $\widehat{\partial}\psi(\bar{x})$ is a cone;*
- *The second inclusion in (1) is an equality if and only if $[\mathrm{dom} d\psi(\bar{x})]^* = [\ker d\psi(\bar{x})]^*$;*
- *If the subderivative $d\psi(\bar{x})$ is a sublinear function as is true when ψ is regular at $\bar{x}$ (see Definition 7.25 of [?]), then*

$$\mathrm{clpos}(\widehat{\partial}\psi(\bar{x})) = [\ker d\psi(\bar{x})]^*. \tag{2}$$

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

Lemma

*([**?**], Proposition 8.5). 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 $\bar{x}$.*

Remark

*This variational description is a contribution to the basics of variational analysis, as pointed out on p.347 of [**?**].*

We can obtain from Lemmas 9 and 10 the following.

Theorem

Consider the following conditions:

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

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

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

Then (i) $\Longrightarrow$ (ii) $\Longleftrightarrow$ (iii).

Theorem

Let $0 \le p < 1$. Consider the following conditions:

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

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

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

Then (i) $\Longrightarrow$ (ii) $\Longleftrightarrow$ (iii).

Remark

In the case of $p = 0$, (i) and (ii) are equivalent, and moreover Theorem 12 recovers a well-known result that the GCQ $[T_C(\bar{x})^* = L_C(\bar{x})^*]$ is the weakest one ensuring KKT conditions.

Remark

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.

By a direct calculation using the chain rule for second subderivatives of piecewise linear-quadratic functions [2], 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 $dS^{\frac{1}{2}}(\bar{x})(w)$

$$= \frac{\sqrt{2}}{2}\sqrt{\max_{\rho \in \mathrm{KKT}_0(\bar{x}),\, \|\rho\|_\infty = 1} \left\langle \left[\sum_{i\in I} \rho_i \nabla^2 g_i(\bar{x}) + \sum_{j \in J} \rho_j \nabla^2 h_j(\bar{x})\right] w, w \right\rangle},$$

where

$$\mathrm{KKT}_0(\bar{x}) := \left\{ \rho \;\middle|\; \begin{array}{l} \sum_{i\in I} \rho_i \nabla g_i(\bar{x}) + \sum_{j\in J} \rho_j \nabla h_j(\bar{x}) = 0 \\ \rho_i \geq 0 \quad \forall i \in I(\bar{x}),\ \rho_i = 0\ \forall i \in I \backslash I(\bar{x}) \end{array} \right\}$$

denotes the degenerate KKT multiplier set at $\bar{x}$.
But we have no idea the explicit formula of $\widehat{\partial} S^p(\bar{x})$, though we are sure that

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

[2] See Chapter 13 of [?]

Proposition

$S^{\frac{1}{2}}$ is of KKT-type at $\bar{x}$ if one of the following conditions is satisfied:

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

$$\langle w, \nabla^2 g_i(\bar{x}) w \rangle \leq 0 \ \forall i \in I(\bar{x}, w), \quad \langle w, \nabla^2 h_j(\bar{x}) w \rangle = 0 \ \forall j \in J. \tag{3}$$

(ii) *For every $w \in L_C(\bar{x})$, there exists some $z \in R^n$ such that*

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

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

$$\max_{\lambda \in \mathrm{KKT}_0(\bar{x})} \left\{ \sum_{i \in I} \lambda_i \langle w, \nabla^2 g_i(\bar{x}) w \rangle + \sum_{j \in J} \lambda_j \langle w, \nabla^2 h_j(\bar{x}) w \rangle \right\} = 0. \tag{4}$$

- Condition (3) was originally given in [?]. In general, we have
  - LICQ $\not\Longrightarrow$ (3). Consider $x_2^2 - x_1 \leq 0$ and $\bar{x} = (0,0)$.
  - (3) $\not\Longrightarrow$ LICQ. Consider $x^3 \leq 0$ and $\bar{x} = 0$.
- Condition (4) is newly obtained, and we have

$$\text{MFCQ} \Longrightarrow (4),$$

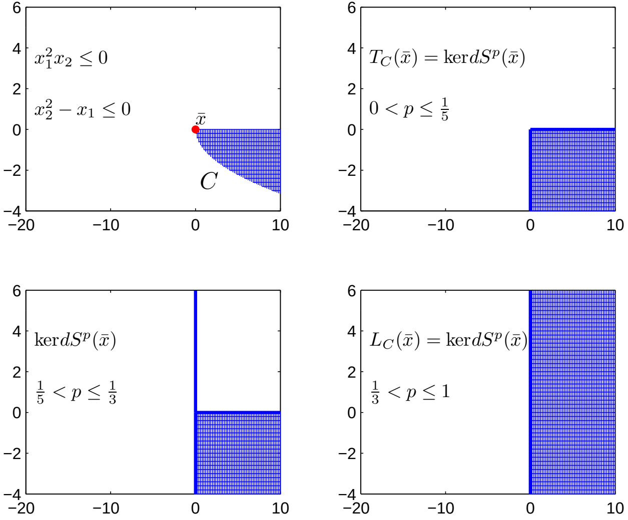
because the MFCQ at $\bar{x} \Longleftrightarrow \mathrm{KKT}_0(\bar{x}) = \{0\}$.

Example

Let $\bar{x} = (0,0)$ and let

$$C = \left\{ x \in R^n \;\middle|\; \begin{array}{c} x_1^2 x_2 \leq 0 \\ x_2^2 - x_1 \leq 0 \end{array} \right\}.$$

- Neither the GCQ nor (3) is satisfied at $\bar{x}$.
- (4) holds and $\mathrm{KKT}_0(\bar{x}) = R_+ \times \{0\}$.


$x_1^2 x_2 \leq 0$
$x_2^2 - x_1 \leq 0$
$\bar{x}$
C
$T_C(\bar{x}) = \ker dS^p(\bar{x})$
$0 < p \leq \frac{1}{5}$
$\ker dS^p(\bar{x})$
$\frac{1}{5} < p \leq \frac{1}{3}$
$L_C(\bar{x}) = \ker dS^p(\bar{x})$
$\frac{1}{3} < p \leq 1$

$T_C(\bar{x}) = R_+ \times (-R_+)$, $L_C(\bar{x}) = R_+ \times R$, and

$$\ker dS^p(\bar{x}) = \begin{cases} R_+ \times (-R_+) & \text{if } 0 < p \leq \frac{1}{5}, \\ R_+ \times (-R_+) \cup \{0\} \times R_+ & \text{if } \frac{1}{5} < p \leq \frac{1}{3}, \\ R_+ \times R & \text{if } \frac{1}{3} < p \leq 1. \end{cases}$$

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

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

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

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

where the convention $\sup \emptyset := -\infty$ is used.

- l_1 exactness $\Longrightarrow$ (SON). See Corollary 4.5 of [**?**].

For any w and z, let

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

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

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

The parabolic subderivative of f at $\bar{x}$ for w with respect to z is defined by, see [?]

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

Theorem

Let $\bar{x}$ be a local minimum of (NLP). Suppose that the penalty function $f + \mu\phi$ is exact at $\bar{x}$. If

$$L_C^2(\bar{x} \mid w) \subset \mathrm{clconv}[\mathrm{ker} d^2\phi(\bar{x})(w \mid \cdot)] \qquad \forall w \in \mathcal{V}(\bar{x}), \tag{5}$$

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

$$L_C^2(\bar{x} \mid w) \subset \mathrm{ker} d^2 S^p(\bar{x})(w \mid \cdot) \qquad \forall w \in L_C(\bar{x}) \tag{6}$$

to hold, which is slightly stronger than (5) since in general $\mathrm{ker} d^2 S^p(\bar{x})(w \mid \cdot)$ is not a closed and convex set and $\mathcal{V}(\bar{x}) \subsetneq L_C(\bar{x})$.

Theorem

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

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

$$\begin{cases} \langle w, \nabla^2 g_i(\bar{x})z \rangle + \dfrac{1}{3} g_i^{(3)}(\bar{x})(w, w, w) \leq 0 & \forall\, i \in I(\bar{x}, w, z), \\ \langle w, \nabla^2 h_j(\bar{x})z \rangle + \dfrac{1}{3} h_j^{(3)}(\bar{x})(w, w, w) = 0 & \forall\, j \in J, \end{cases} \tag{7}$$

(iii) $p \in [0, \frac{2}{3})$, $q = 0$ *(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*

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

Remark

(a) *Let $p = 1$. By applying the second-order Taylor expansion we have*

$$L_C^2(\bar{x} \mid w) = \ker d^2 S(\bar{x})(w \mid \cdot) \quad \forall w \in L_C(\bar{x}), \tag{8}$$

*which implies that condition (6) holds. This recovers a well-known result that the (SON) condition holds at $\bar{x}$ when the l_1 penalty function is exact at $\bar{x}$, see [**?**].*

(b) *Let $0 \leq p < 1$. It can be shown that*

$$\mathrm{ker} d^2 S^p(\bar{x})(w \mid \cdot) \subset L_C^2(\bar{x} \mid w) \quad \forall w \in \mathrm{ker} dS^p(\bar{x}). \qquad (9)$$

Thus, condition (5) holds if and only if

$$L_C^2(\bar{x} \mid w) = \mathrm{clconv}[\mathrm{ker} d^2 S^p(\bar{x})(w \mid \cdot)] \quad \forall w \in \mathcal{V}(\bar{x}). \qquad (10)$$

Moreover, it is clear that

$$T_C^2(\bar{x} \mid w) = \mathrm{ker} d^2 S^0(\bar{x})(w \mid \cdot), \quad \forall w \in T_C(\bar{x}).$$

Condition (10) *with $p = 0$ reduces to the so-called* SGCQ, *originated with [?], which holds at $\bar{x}$ if by definition*

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

(c) *It was shown by [**?**] that if the linear independent constraint qualification (for short, LICQ) holds at* $\bar{x}$*, then*

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

and hence (6) holds for any $p \in [0, 1]$.
Simple example can be given to demonstrate that condition (7) may not hold even if the LICQ holds at $\bar{x}$.

Consider the following semi-infinite program, denoted as (SIP):

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

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \times \Omega \to \mathbb{R}$ are smooth functions, and Ω is a nonempty and compact set of parameters in $\mathbb{R}^m$.

Let x^* be a locally optimal solution of (SIP),

$$X := \{x \in R^n : g(x,t) \leq 0, t \in \Omega\}$$

be the feasible set and, for $x \in X$, let

$$\Omega_x := \{t \in \Omega : g(x,t) = 0\}.$$

Literature review

Three types of optimality conditions for (SIP):

$$0 \in \mathrm{conv}\{\nabla f(x^*), \nabla_x g(x^*, t)\ (t \in \Omega_{x^*})\},$$

(see Fritz John (1948), Pschenichnyi (1971), Hettich and Jongen (1978), and Borwein (1981),)

$$0 \in \nabla f(x^*) + \mathrm{cl\ cone}\{\nabla_x g(x^*, t)\ (t \in \Omega_{x^*})\},$$

(see [**?**], Li, et al (2000),)

$$0 \in \nabla f(x^*) + \mathrm{cone}\{\nabla_x g(x^*, t)\ (t \in \Omega_{x^*})\},$$

(see Pschenichnyi (1971), Lopez and Vercher (1983), Hettich and Kortanek (1993), Zheng and Yang (2007).)

Literature review

By a Farkas lemma, see [**?**],

$$0 \in \nabla f(x^*) + \mathrm{cl}\ \mathrm{cone}\{\nabla_x g(x^*, t)\ (t \in \Omega_{x^*})\}, \tag{11}$$

is equivalent to

$$\langle \nabla f(x^*), d \rangle \geq 0,\ \forall d \in D(x^*), \tag{12}$$

where $D(x) = \{0 \neq d \in R^n : \langle \nabla_x g(x, t), d \rangle \leq 0\ \forall t \in \Omega_x\}$.

In this talk, we will study (11) but using the form of (12).

Literature review

[?] introduced the following l_1 integral penalty function

$$f(x) + \rho \int_{\Omega(x)} g(x,t)\mathrm{d}\mu(t),$$

where $\Omega(x) := \{t \in \Omega : g(x,t) > 0\}$, but too weak penalty for infeasibility.

Let $p > 0$. For (SIP), [?] also introduced the following l_p integral penalty function

$$f(x) + \rho \int_{\Omega} g_+^p(x,t)\mathrm{d}\mu(t)$$

and established the convergence of the solution sequence of the penalty problems to an optimal solution of (SIP).

[?] established the exact l_1 integral penalty function

$$f(x) + \rho \int_{\Omega(x)} g(x,t)\mathrm{d}\mu(t) \Bigg/ \int_{\Omega(x)} \mathrm{d}\mu(t) \,.$$

Penalty Functions

A pth-order max-type penalty function for SIP is defined as,

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

Let μ be a non-negative regular Borel measure defined on Ω with the support of μ being equal to Ω, that is $\mathrm{supp}(\mu) = \Omega$, where the support of μ is defined as the set of the points $t \in \Omega$ such that any open neighbourhood V of t has a positive measure:

$$\mathrm{supp}\,(\mu) := \{t \in \Omega : \mu(V) > 0, \text{ for any open neighbourhood } V \text{ of } t\}.$$

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

$$\begin{aligned} F^p_{int}(x) &= f(x) + \rho \int_\Omega g^p_+(x,t)\,\mathrm{d}\mu(t), \\ \bar{F}^p_{int}(x) &= f(x) + \rho \left(\int_\Omega g_+(x,t)\,\mathrm{d}\mu(t)\right)^p. \end{aligned}$$

Exactness of $F^p_{int}(x) \Longrightarrow$ that of $\bar{F}^p_{int}(x) \Longrightarrow$ that of $F^p_{max}(x)$.

The case $p = 1$.

(SIP) can be rewritten as

$$\min f(x) \text{ s.t. } \max_{t \in \Omega} g(x, t) \leq 0. \tag{13}$$

The exactness of F^1_{max} is equivalent to saying that problem (13) has an l_1 exact penalty function in the usual sense, see Clarke (1983). Thus, if F^1_{max} is exact, then the KKT-type optimality condition (12).

The case $p \in (0,1)$.

Let $h : \mathbb{R}^n \to \mathbb{R}$. The upper Dini-directional derivative of h at a point x in the direction $d \in \mathbb{R}^n$ is defined by

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

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

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

Let $D(x) = \{0 \neq d \in R^n : \langle \nabla_x g(x,t), d \rangle \leq 0 \ \forall t \in \Omega_x\}$ and let

$$\begin{aligned} \Omega_x^{=}(d) &:= \{t \in \Omega_x : \langle \nabla_x g(x,t), d \rangle = 0\}, \\ \Omega_x^{<}(d) &:= \{t \in \Omega_x : \langle \nabla_x g(x,t).d \rangle < 0\}, \end{aligned}$$

The case $p \in (0,1)$.

Lemma

[?] Let $\bar{h}(x) = (\max\{h(x), 0\})^p$ with $p \in]0,1[$ and h be continuously differentiable at x.

(i) *If $h(x) < 0$, then $D_+\bar{h}(x; d) = 0$;*

(ii) *If $h(x) = 0$ and $\langle \nabla h(x), d\rangle < 0$, then $D_+\bar{h}(x; d) = 0$;*

(iii) *If $p \in (0.5, 1)$, $h(x) = 0$, $\langle \nabla h(x), d\rangle = 0$ and $h^{\circ\circ}(x; d)$ is finite, then $D_+\bar{h}(x; d) = 0$;*

(iv) *If $p = 0.5$, $h(x) = 0$ and $\langle \nabla h(x), d\rangle = 0$, then $D_+\bar{h}(x; d) \leq \sqrt{\max\{\frac{1}{2}h^{\circ\circ}(x; d), 0\}}$;*

(v) *If $p \in (0, 0.5)$, $h(x) = 0$, $\langle \nabla h(x), d\rangle = 0$ and $h^{\circ\circ}(x; d) < 0$, then $D_+\bar{h}(x; d) = 0$.*

The case $p \in (0,1)$.

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

Theorem

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

$$\langle \nabla f(x^*), d\rangle \geq 0, \forall d \in D(x^*).$$

The case $p \in (0,1)$.

Next we employ the exactness of $\bar{F}_{int}(p \in (0,1))$ to develop the optimality condition (12) of (SIP).

Theorem

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

$$\langle \nabla f(x^*), d\rangle \geq 0, \forall d \in D(x^*).$$

The case $p \in (0,1)$.

We need the following lemma in the proof of the above theorem.

Proposition

Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ *be a non-negative function,* $f : \mathbb{R}_+ \to \mathbb{R}$ *be a continuous and strictly increasing function and* $\lambda_0 \in \mathbb{R}_+$*. Then*

$$\limsup_{\lambda \to \lambda_0} f(g(\lambda)) \leq f(\limsup_{\lambda \to \lambda_0} g(\lambda)).$$

We will also consider the following generalized semi-infinite program, denoted as (GSIP),

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

where Ω is a compact subset of $\mathbb{R}^m$,

$$\Omega(x) := \{t \in \mathbb{R}^m : v_i(x,t) \le 0, i = 1, \cdots, l\}$$

and the functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and $v_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}(i = 1, \cdots, l)$ are smooth.

[?] associated (GSIP) with an (SIP) problem via augmented Lagrangians of the lower level problem.

The lower level problem associated with (GSIP) is

$$Q(x) \qquad \max_{t\in\Omega} g(x,t) \quad \text{s.t.} \quad v_i(x,t) \leq 0, i = 1, \cdots, l.$$

Let $\mathrm{val}Q(x)$ be the optimal value of the problem $Q(x)$. It is clear that

$$x \in X_{(\mathrm{GSIP})} \quad \text{iff} \quad \mathrm{val}Q(x) \leq 0.$$

Let $\bar{f}(x,\mu,c) = f(x)$ and, for $(x,\mu,c) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}_{++}$,

$$\bar{g}(x,t,\mu,c) = g(x,t) - \frac{1}{2c}\sum_{i=1}^{l}\{([cv_i(x,t)+\mu_i]_+)^2 - \mu_i^2\}.$$

Then $\bar{g}$ is of $C^{1,1}$, see Hiriart-Urruty et al (1984).

Next we recall some concepts from [**?**].

Problem $Q(x)$ is said to satisfy the *quadratic growth condition* iff there is a $c \geq 0$ such that $\bar{g}(x, t, 0, c)$ is bounded above as a function of $t \in \Omega$.

Problem $Q(x)$ is said to be *stable of degree* 2 iff there is a neighbourhood U of the origin in $\mathbb{R}^l$ and a C^2 function $\pi_x : U \to \mathbb{R}$ such that

$$\nu(x, u) \leq \pi_x(u), \forall u \in U, \text{ and } \nu(x, 0) = \pi_x(0).$$

Let $\bar{H}(x, \mu, c) := \max_{t \in \Omega} \bar{g}(x, t, \mu, c)$.

Lemma

[?] Under the quadratic growth condition of $Q(x)$, we have

$$\mathrm{val}\, Q(x) = \min_{(\mu,c) \in \mathbb{R}^l \times \mathbb{R}_{++}} \bar{H}(x, \mu, c)$$

iff the problem $Q(x)$ is stable of degree 2.

Consider the following (SIP) problem, denoted as (SIPg),

$$\min_{(x,\mu,c)\in\mathbb{R}^n\times\mathbb{R}^l\times\mathbb{R}_{++}} \bar{f}(x,\mu,c) \quad \text{s.t.} \quad \bar{g}(x,t,\mu,c)\leq 0, t\in\Omega.$$

Therefore we have

Proposition

Assume that, for all $x\in\mathbb{R}^n$, $Q(x)$ satisfies the quadratic growth condition and is stable of degree 2. Then problems (GSIP) and (SIPg) have the same optimal value, i.e., val(GSIP) = val(SIPg), and furthermore,
(i) *if $\hat{x}$ is a locally optimal solution of (GSIP), then there exists $(\hat{\mu},\hat{c})\in\mathbb{R}^l\times\mathbb{R}_{++}$ such that $(\hat{x},\hat{\mu},\hat{c})$ is a locally optimal solution of (SIPg);*
(ii) *if $(\hat{x},\hat{\mu},\hat{c})$ is a locally optimal solution of (SIPg), then $\hat{x}$ is a locally optimal solution of (GSIP).*

For $p \in (0,1)$, let

$$G^p_{int}(x,\mu,c) := \bar{f}(x,\mu,c) + \rho \int_\Omega \bar{g}^p_+(x,t,\mu,c)\,\mathrm{d}\mu(t).$$

By applying previous Theorem for SIP, we have.

Theorem

Let the assumptions of the previous Proposition hold. Let $\hat{x}$ be a locally optimal solution of (GSIP) and G^p_{int} be exact at the point $(\hat{x}, \hat{\mu}, \hat{c})$. Then, under one of the following assumptions,
(i) $p \in (0.5, 1)$,
(ii) $p = 0.5$ *and* $\bar{g}^{\circ\circ}_{(x,\mu,c)}(\hat{x}, t, \hat{\mu}, \hat{c}; d) \leq 0$ *for all* $d \in D(\hat{x}, \hat{\mu}, \hat{c})$ *and* $t \in \Omega_{(\hat{x},\hat{\mu},\hat{c})}$ *with* $\langle \nabla_{(x,\mu,c)}\bar{g}(\hat{x}, t, \hat{\mu}, \hat{c}), d\rangle = 0$, *and*
(iii) $p \in (0, 0.5)$ *and* $\bar{g}^{\circ\circ}_{(x,\mu,c)}(\hat{x}, t, \hat{\mu}, \hat{c}; d) < 0$ *for all* $d \in D(\hat{x}, \hat{\mu}, \hat{c})$ *and* $t \in \Omega_{(\hat{x},\hat{\mu},\hat{c})}$ *with* $\langle \nabla_{(x,\mu,c)}\bar{g}(\hat{x}, t, \hat{\mu}, \hat{c}), d\rangle = 0$,
we have

$$\langle \nabla f(\hat{x}), d_1\rangle \geq 0,$$

for all $d_1 \in \mathbb{R}^n$ *satisfying* $\langle \nabla_x g(\hat{x}, t) - \nabla_x^T v(\hat{x}, t)\hat{\mu}, d_1\rangle \leq 0, t \in \Omega_{(\hat{x},\hat{\mu},\hat{c})}$.

We now compute the generalized second-order directional derivative $\bar{g}^{\circ\circ}_{(x,\mu,c)}(\hat{x}, t, \hat{\mu}, \hat{c}; d)$ for $d \in D(\hat{x}, \hat{\mu}, \hat{c})$ and $t \in \Omega(\hat{x}, \hat{\mu}, \hat{c})$.

Lemma

Let $d \in D(\hat{x}, \hat{\mu}, \hat{c})$ and $t \in \Omega(\hat{x}, \hat{\mu}, \hat{c})$. Then the following formula holds:

$$\bar{g}^{\circ\circ}_{(x,\mu,c)}(\hat{x}, t, \hat{\mu}, \hat{c}; d) = d_1^T[\nabla^2_{xx} g(\hat{x}, t) - \sum_{i=1}^{l} \hat{\mu}_i \nabla^2_{xx} v_i(\hat{x}, t)] d_1$$

$$- \sum_{i \in \hat{I}^+_{(\hat{x},\hat{\mu},\hat{c})}(t)} (\sqrt{\hat{c}} d_1^T \nabla_x v_i(\hat{x}, t) + \frac{d_{2i}}{\sqrt{\hat{c}}})^2 + \sum_{i=1}^{l} \frac{d_{2i}^2}{\hat{c}},$$

where $\hat{I}^+_{(\hat{x},\hat{\mu},\hat{c})}(t) = \{i \in \{i, \cdots, l\} : \hat{c} v_i(\hat{x}, t) + \hat{\mu}_i > 0\}$.

We have the following corollary.

Corollary

Assume that the following conditions hold:
(i) $G_{int}^{\frac{1}{2}}(x,\mu,c)$ *is exact at* $(\hat{x},\hat{\mu},\hat{c})$;
(ii) $g(\cdot,t)$ *and* $-v_i(\cdot,t)$ $(i=1,\cdots,l)$ *are concave for each* $t\in\Omega$; *and*
(iii) $\hat{I}^{+}_{(\hat{x},\hat{\mu},\hat{c})}(t)=\{1,\cdots,l\}$ *and* $\langle\nabla_x v_i(\hat{x},t),d_1\rangle=0$ *for* $d\in D(\hat{x},\hat{\mu},\hat{c}), t\in\Omega(\hat{x},\hat{\mu},\hat{c})$ *and* $i\in\hat{I}^{+}(t)$. *Then we have*

$$\langle\nabla f(\hat{x}),d_1\rangle\geq 0,$$

for all $d_1\in\mathbb{R}^n$ *satisfying* $\langle\nabla_x g(\hat{x},t)-\nabla_x^T v(\hat{x},t)\hat{\mu},d_1\rangle\leq 0, t\in\Omega_{(\hat{x},\hat{\mu},\hat{c})}$.

• For (NLP), we discussed the first-order optimality conditions by Dini-directional derivative, contingent directional derivative and subderivative respectively.

• For (SIP) and (GSIP), we investigated the first-order optimality conditions by Dini-directional derivative.