Exact Penalty Functions
a Lower Bound to the Penalty Parameter

S. M. H. Janesch

Department of Mathematics-CFM-UFSC
Caixa Postal 476
88040-900, Florianópolis-SC-Brazil
sjanesch@mtm.ufsc.br

Abstract

We consider a class of nondifferentiable penalty functions associated
to nonlinear programming problem with equality and inequality con-
straints. We analyze the relationships between the solutions of the con-
strained problem and the solutions of the penalty function. Assuming
first-order sufficient optimality condition of the constrained problem,
we show the exactness of the class of penalty functions and give an
estimative of the size of the exact penalty parameter. Moreover, under
assumption, exactness of the penalty function, we provide a lower bound
to the penalty parameter in terms of the Lagrange multipliers of the
nonlinear programming problem.

Mathematics Subject Classification: 49M30, 49M37, 90C30

Keywords: Exact penalty function, nonlinear programming, constrained
optimization, optimality conditions.

1 Introduction

Considerable attention has been given to the penalty methods that use exact
penalty functions, see [4]. These methods transform the constrained nonlinear
programming problem,

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h(x) = 0, \\
& \quad g(x) \leq 0,
\end{align*}
\]  

(1)
where \( f, h \) and \( g \) are functions from \( \mathbb{R}^n \) into \( \mathbb{R}, \mathbb{R}^{m_1} \) and \( \mathbb{R}^{m_2} \) respectively, with \( f, h, \) and \( g \in C^1 \), into a single unconstrained problem or into a finite sequence of unconstrained problems, avoiding thus the infinite sequential process of the classical penalty methods.

Let \( x^* \) be a local minimum point of (1) and let \( \rho > 0 \) be a penalty parameter. We remember that a penalty function \( Q(x, \rho) \) is exact at \( x^* \), if there exists a finite penalty parameter \( \bar{\rho} > 0 \), such that \( x^* \) is a local minimum point of \( Q(x, \bar{\rho}) \), and hence of \( Q(x, \rho) \) for any \( \rho \geq \bar{\rho} \).

We associate with problem (1) the class of penalty functions, introduced by Han and Mangasarian [6] in 1979,

\[
P(x, \rho) = f(x) + \rho \| h(x), g^+(x) \|,
\]

where \( \rho > 0 \) is the penalty parameter and \( \| \cdot \| \) is any fixed vector norm in \( \mathbb{R}^{m_1+m_2} \). Throughout the paper, the notation \( \| u, v \| \) means the norm of \( (u, v) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \) with \( u = (u_1, \ldots, u_{m_1})^T \), and the notation \( u^+ \) indicates the vector with components \( u_i^+ = \max[0, u_i] \).

Included in the class of penalty functions (2) is the classical exact penalty function analyzed by Pietrzykowski [9] in 1969,

\[
P_1(x, \rho) = f(x) + \rho \left( \sum_{i=1}^{m_1} |h_i(x)| + \sum_{j=1}^{m_2} g_j^+(x) \right).
\]

The main motivation for the use of penalty methods is that of solving the constrained optimization problem by employing some unconstrained minimization algorithm. Usually, unconstrained minimization algorithms provide local minimizers or just stationary points. Thus, for nonconvex problems, it is of greater interest to investigate properties that ensure that local solution or stationary point to the penalty function is local solution or stationary point to the constrained problem.

In the literature on exact penalty functions, in general, the two important issues are: the conditions in order to ensure that the penalty function has a local minimum at a local minimum of (1) for all sufficiently large but finite \( \rho \), and to provide a lower bound on the critical penalty parameter. For example, in [9] Pietrzykowski assumed linear independence of the active constraints gradients and showed that there exists a finite penalty parameter, \( \bar{\rho} > 0 \), such that for \( \rho \geq \bar{\rho} \) any local minimizer of problem (1) is also a local minimizer of the function \( P_1(x, \rho) \). Howe [7] assumed first-order sufficient optimality conditions for problem (1), and established the existence of a finite penalty parameter associated to function \( P_1(x, \rho) \). Under second-order sufficient optimality conditions for problem (1), Han and Mangasarian [6] and Charalambous [2] provided a lower bound to the penalty parameter in terms of the optimal
Lagrange multipliers. Assuming the Mangasarian-Fromovitz constraint qualification, Han and Mangasarian [6] showed the existence of an exact penalty parameter associated to function $P(x, \rho)$.

The purposes of this paper are to relate solutions of the constrained problem with solutions of the penalty function associated, and give an estimate of the size of the exact penalty parameter. More specifically, in Section 2, the Theorem 2.2 shows that point stationary of Karush-Kuhn-Tucker (KKT) for problem (1) is also stationary of $P(x, \rho)$, for sufficiently large $\rho$. In the Theorem 2.4, we show that stationary point of $P(x, \rho)$, that is feasible for problem (1), is a stationary point of KKT for constrained problem. The Theorem 2.8, provides conditions for that a stationary point of the penalty function be a feasible point for problem (1). In Theorem 2.9 we show, under suitable assumptions, that local minimizer of $P(x, \rho)$ is also local minimizer of the constrained problem. The Theorem 2.10, shows the exactness of $P(x, \rho)$ and provides a lower bound to the penalty parameter. The Section 3 contains our principal result, the Theorem 3.3. It shows that if $P(x, \rho)$ is exact then a lower bound for penalty parameter is given by a constant times a norm of the optimal Lagrange multipliers. This norm is dual to the norm used in the definition of the exact penalty function $P(x, \rho)$.

2 Local solutions and stationary points

In this section, we relate stationary point and local minimizer of problem (1) with a stationary point and a local minimizer of the penalty function (2), respectively. Initially, we review some concepts and notations that will be useful.

We denote by $DP(x^*, \rho; d)$ the directional derivative of $P(x, \rho)$ at $x^*$ along the direction $d$. We say that $x^* \in \mathbb{R}^n$ is a stationary point of $P(x, \rho)$, if $DP(x^*, \rho; d) \geq 0$ for all $d \in \mathbb{R}^n$.

A point $x^* \in \mathbb{R}^n$ is called a stationary point of KKT for the problem (1), if there exists $(\lambda^*, \mu^*) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ such that

$$\nabla f(x^*) + J_h(x^*)^T \lambda^* + J_g(x^*)^T \mu^* = 0,$$

$$h(x^*) = 0, \quad g(x^*) \leq 0,$$

$$\mu^* \geq 0 \text{ and } \mu^*_j g_j(x^*) = 0 \text{ for } j = 1, \ldots, m_2,$$

where

$$J_h(x^*) = [\nabla h_1(x^*) \ldots \nabla h_{m_1}(x^*)]^T \quad \text{and} \quad J_g(x^*) = [\nabla g_1(x^*) \ldots \nabla g_{m_2}(x^*)]^T.$$

For problem (1) we define $J(x)$ to be index set of active inequality constraints at $x$, this is

$$J(x) = \{ j \mid g_j(x) = 0 \}.$$
We denote by $\mathcal{F}$ the feasible points set for problem (1):

$$\mathcal{F} = \{ x \in \mathbb{R}^n \mid h(x) = 0, \ g(x) \leq 0 \}.$$ 

If $\| \cdot \|$ is a norm in $\mathbb{R}^n$ then there is a correspondent norm $\| \cdot \|'$, called the dual norm, which is defined by

$$\| x \|' = \max_{y \neq 0} \frac{y^T x}{\| y \|}, \text{ for all } x \in \mathbb{R}^n.$$ 

For example, the dual of $\| \cdot \|_1$ is $\| \cdot \|_\infty$. From the definition of dual norm, it follows the so-called generalized Cauchy inequality,

$$|x^T y| \leq \| x \|' \| y \|, \text{ for all } x, y \in \mathbb{R}^n.$$ 

To relate a stationary point of problem (1) with a stationary point of the penalty function, we need the expression of the directional derivative of the penalty function in a feasible point for problem (1).

**Proposition 2.1** ([8], p.16). Let $P(x, \rho)$ be the function defined by (2), and let $x^* \in \mathcal{F}$. The directional derivative of $P(x, \rho)$ at $x^*$ along the direction $d \in \mathbb{R}^n$ is given by

$$DP(x^*, \rho; d) = \nabla f(x^*)^T d + \rho \| J_h(x^*) d, J_{g^+}(x^*) d \|,$$

where

$$J_h(x^*) d = [\nabla h_1(x^*)^T d, \ldots, \nabla h_{m_1}(x^*)^T d]^T,$$

$$J_{g^+}(x^*) d = [\nabla g_{1^+}(x^*)^T d, \ldots, \nabla g_{m_2^+}(x^*)^T d]^T \text{ and}$$

$$\nabla g_{j^+}(x^*)^T d = \begin{cases} 
0, & \text{if } g_j(x^*) < 0 \\
\max \{ \nabla g_j(x^*)^T d, 0 \}, & \text{if } g_j(x^*) = 0.
\end{cases}$$

**Theorem 2.2.** Let $\| \cdot \|$ be any given norm in $\mathbb{R}^{m_1+m_2}$ and $P(x, \rho)$ be its corresponding penalty function defined as in (2), and let $\| \cdot \|'$ be its dual norm. Let $x^*$ be a stationary point of KKT for problem (1) with $(\lambda^*, \mu^*) \in \mathbb{R}^{m_1} \times \mathbb{R}_{+}^{m_2}$ the Lagrange multipliers correspondent. If $\rho \geq \| \lambda^*, \mu^* \|'$ then $x^*$ is a stationary point of $P(x, \rho)$.

**Proof.** Using the expression for the directional derivative of $P(x, \rho)$, the fact of $x^*$ to be stationary point for problem (1), and generalized Cauchy inequality, we obtain for all $d \in \mathbb{R}^n$,

$$DP(x^*, \rho; d) = \nabla f(x^*)^T d + \rho \| J_h(x^*) d, J_{g^+}(x^*) d \|$$

$$= -(\lambda^*)^T J_h(x^*) d - (\mu^*)^T J_{g^+}(x^*) d + \rho \| J_h(x^*) d, J_{g^+}(x^*) d \|$$

$$\geq -(\lambda^*)^T J_h(x^*) d - (\mu^*)^T J_{g^+}(x^*) d + \rho \| J_h(x^*) d, J_{g^+}(x^*) d \|$$

$$\geq - \| \lambda^*, \mu^* \|' \| J_h(x^*) d, J_{g^+}(x^*) d \| + \rho \| J_h(x^*) d, J_{g^+}(x^*) d \|$$

$$= (\rho - \| \lambda^*, \mu^* \|') \| J_h(x^*) d, J_{g^+}(x^*) d \|.$$
Thus, for \( \rho \geq \|\lambda^*, \mu^*\| \) we have
\[
DP(x^*, \rho; d) \geq 0, \quad \text{for all } d \in \mathbb{R}^n.
\]
Therefore, \( x^* \) is a stationary point of \( P(x, \rho) \).

While a point stationary of KKT gives rise to a stationary point of \( P(x, \rho) \), for sufficiently large \( \rho \), the reverse is not true. The next example illustrates this situation.

**Example 2.3.** Consider the problem of minimize \( f(x) = 0 \) subject to \( x^3 + 3x^2 + 3 = 0 \). Associate to this problem we have the penalty function using the one norm. The point \( x^* = 0 \) is a local minimizer of the \( P_1(x, \rho) = \rho|x^3 + 3x^2 + 3| \), for all \( \rho > 0 \). But \( x^* \) is an infeasible point for the constrained problem, so it can’t be a stationary point.

To relate a stationary point of the penalty function with a stationary point for problem (1), we will assume feasibility.

**Theorem 2.4.** Let \( P(x, \rho) \) be the penalty function defined by (2). If \( x^* \in \mathcal{F} \) is a stationary point of \( P(x, \rho) \) then \( x^* \) is a stationary point of KKT for problem (1).

**Proof.** Consider the set
\[
A = \{d \in \mathbb{R}^n \mid \nabla h_i(x^*)^T d = 0, \quad i = 1, \ldots, m_1, \quad \nabla g_j(x^*)^T d \leq 0, \quad j \in J(x^*),
\]
\[
\nabla f(x^*)^T d < 0\}.
\]
We have, by Farkas’ lemma, that if \( A = \emptyset \) then there exists \( (\lambda^*, \mu^*) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \) such that \( x^* \) is a stationary point of KKT for problem (1). We going to show that \( A = \emptyset \). As \( x^* \in \mathcal{F} \), then by Proposition 2.1 we can write
\[
DP(x^*, \rho; d) = \nabla f(x^*)^T d + \rho\|J_h(x^*)d, J_g(x^*)d\|.
\]
Thus for \( d \) satisfying \( \nabla h_i(x^*)^T d = 0, \quad i = 1, \ldots, m_1 \) and \( \nabla g_j(x^*)^T d \leq 0, \quad j \in J(x^*) \), we get \( DP(x^*, \rho; d) = \nabla f(x^*)^T d \). Since \( x^* \) is stationary point of \( P(x, \rho) \), it follows that \( \nabla f(x^*)^T d \geq 0 \) and consequently \( A = \emptyset \). Therefore the conditions of KKT are satisfied at \( x^* \). \hfill \Box

Others proofs of the Theorems 2.2 and 2.4 can be found depending on the chosen norm used in the penalty function. For \( \| \cdot \|_\infty \), see [1], [3] and [4], and for \( \| \cdot \|_p \) \( (p \geq 1) \), see [3] and [4]. Our proofs do not depend on the chosen norm.

The Theorem 2.4 ensures that stationary point of \( P(x, \rho) \), that is feasible for problem (1), is a stationary point KKT for (1). The Theorem 2.8 will
provide conditions for that a stationary point of the penalty function be a feasible point for problem (1). In order to show its, we need the propositions below. Before, we recall the definitions of subdifferential and monotonic norm.

Let \( \phi: \mathbb{R}^n \to \mathbb{R} \) be a convex function. The subdifferential \( \partial \phi(x) \) is defined for each \( x \in \mathbb{R}^n \) by

\[
\partial \phi(x) = \{ z \in \mathbb{R}^n \mid \phi(x + y) \geq \phi(x) + z^T y \text{ for all } x + y \in \mathbb{R}^n \}.
\]

Let \( x \in \mathbb{R}^n \). We define \( |x| = (|x_1|, \ldots, |x_n|)^T \) and if \( |x_i| \leq |y_i| \), for all \( i = 1, \ldots, n \), we call that \( |x| \leq |y| \). A norm \( \| \cdot \| \) in \( \mathbb{R}^n \) is monotonic norm if \( |x| \leq |y| \) implies that \( \|x\| \leq \|y\| \), for all \( x, y \in \mathbb{R}^n \).

**Proposition 2.5** ([5], p.369). Let \( x^* \in \mathbb{R}^n \), and \( P(x, \rho) = f(x) + \rho \| h(x), g^+(x) \| \) be the penalty function, where \( \| \cdot \| \) is a monotonic norm in \( \mathbb{R}^{m_1 + m_2} \). The directional derivative of \( P(x, \rho) \) at \( x^* \) along the direction \( d \in \mathbb{R}^n \) is given by

\[
DP(x^*, \rho; d) = \nabla f(x^*)^T d + \rho \max_{(\alpha, \beta) \in \partial \| h(x^*), g^+(x^*) \|} (J_h(x^*)d, J_g(x^*)d),
\]

where

\[
J_h(x^*)d = [\nabla h_1(x^*)^T d, \ldots, \nabla h_{m_1}(x^*)^T d]^T \quad \text{and} \quad J_g(x^*)d = [\nabla g_1(x^*)^T d, \ldots, \nabla g_{m_2}(x^*)d]^T.
\]

**Proposition 2.6.** Let \( \| \cdot \| \) be any monotonic norm in \( \mathbb{R}^{m_1 + m_2} \) with \( \| \cdot \|' \) its dual norm, and let \( \phi: \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \to \mathbb{R}^+ \) be the convex function defined by \( \phi(x, y) = \|x, y^+\| \). If \( (\alpha, \beta) \in \partial \phi(x, y) \) then the vector \( (\alpha, \beta) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \) satisfies:

(i) \( \beta_j \geq 0 \) for all \( j = 1, \ldots, m_2 \),

(ii) If \( y_j < 0 \) then \( \beta_j = 0 \),

(iii) \( \alpha_i x_i \geq 0 \) for all \( i = 1, \ldots, m_1 \),

(iv) \( (\alpha, \beta)^T (x, y) = \|x, y^+\| \), and

(v) \( \|\alpha, \beta\|' \leq 1 \).

**Proof.** From \( (\alpha, \beta) \in \partial \phi(x, y) \), we have

\[
(\alpha, \beta)^T (u, v) + \|x, y^+\| \leq \|x + u, (y + v)^+\| \text{ for all } (u, v) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}.
\]

Since the norm is monotonic and applying the triangle inequality, we can write

\[
(\alpha, \beta)^T (u, v) \leq \|u, v^+\| \text{ for all } (u, v) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}.
\]

Taking the vector \((u, v) = (0, -e_j)\) in (5), where \( e_j \) is a vector of zeros except for 1 at the \( j \)th position, and using the fact that \( \|0, (-e_j)^+\| = 0 \), we have

\[
\beta_j \geq 0 \text{ for all } j = 1, \ldots, m_2.
\]
To show (ii), suppose that $\beta_j \neq 0$. It follows from the part (i) that $\beta_j > 0$. For $(u, v) = (0, -y_j e_1)$ in (4) and using the fact that the norm is monotonic, we obtain $\beta_j y_j \geq \|x, y^+\| - \|x, (y_j y_j e_j)^+\| \geq 0$. As $\beta_j > 0$, it follows that $y_j \geq 0$.

To prove (iii), let us take $(u, v) = (-x, e_1, 0)$ in (4). Thus

$$\|x - x, e_1, y^+\| \geq -\alpha_i x_i + \|x, y^+\|.$$ 

Since the norm is monotonic, we have $\alpha_i x_i \geq 0$ for all $i = 1, \ldots, m_1$.

To show (iv), it suffices to do $(u, v) = (-x, -y)$ in (4) and $(u, v) = (x, y)$ in (5), thus

$$\|x, y^+\| = (\alpha, \beta)^T (x, y).$$

Now, combining the inequalities (5) and $\|u, v^+\| \leq \|u, v\|$, we find

$$(\alpha, \beta)^T (u, v) \leq \|u, v\|,$$ for all $(u, v) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$,

therefore $\|\alpha, \beta\|' \leq 1$, and proof is complete. \hfill \Box

**Proposition 2.7 ([6], p.254).** Let $x^* \in \mathcal{F}$ be satisfying the Mangasarian-Fromovitz constraint qualification, that is,

(CQ) \quad \begin{align*}
(i) & \quad \text{There exists a } d \in \mathbb{R}^n \text{ such that } \nabla h_i(x^*)^T d = 0, \; i = 1, \ldots, m_1, \\
(ii) & \quad \text{The gradients } \nabla h_i(x^*), \; i = 1, \ldots, m_1 \text{ are linearly independent.}
\end{align*}

The (CQ) is satisfied at $x^*$ if and only if there exists an open neighborhood $N(x^*; \delta)$ of $x^*$ such that for each bounded function $b(x) : N(x^*; \delta) \to \mathbb{R}^{m_1}$ there exists a bounded function $d(x) : N(x^*; \delta) \to \mathbb{R}^n$ such that for all $x \in N(x^*; \delta)$

$$\nabla h_i(x)^T d(x) = b_i(x) \text{ for } i = 1, \ldots, m_1, \text{ and } \nabla g_j(x)^T d(x) \leq -1 \text{ for } j \in J(x^*).$$

**Theorem 2.8.** Let $\| \cdot \|$ be any given monotonic norm in $\mathbb{R}^{m_1+m_2}$, and $P(x, \rho)$ be its corresponding penalty function defined as in (2). Let $x^* \in \mathcal{F}$ and suppose that (CQ) is satisfied at $x^*$. Then there exists an open neighborhood $N(x^*; \delta)$ of $x^*$ and a number $\bar{\rho} > 0$ so that the followings holds: if $x_\rho$ is a stationary point of $P(x, \rho)$, if $x_\rho \in N(x^*; \delta)$, and if $\rho \geq \bar{\rho}$ then $x_\rho \in \mathcal{F}$.

**Proof.** Suppose by contradiction that there exist a sequence of positive numbers $\{\rho_k\}$ and a sequence of points $\{x_k\}$ so that $\rho_k \to \infty$ and $x_k \to x^*$, where $x^*$ is a stationary point of $P(x, \rho)$ and $x_k \notin \mathcal{F}$. Let an open neighborhood $N(x^*; \delta)$ be defined as in Proposition 2.7 and consider the bounded function $b(x) : N(x^*; \delta) \to \mathbb{R}^{m_1}$ defined by

$$b_i(x) = \begin{cases} 
\frac{-h_i(x)}{|h_i(x)|}, & \text{if } h_i(x) \neq 0 \\
0, & \text{if } h_i(x) = 0.
\end{cases}$$
By Proposition 2.7 there exists a bounded function $d(x) : N(x^*; \delta) \rightarrow \mathbb{R}^n$ such that for all $x \in N(x^*; \delta)$,
\[
\nabla h_i(x)^T d(x) = \begin{cases} 
-1, & \text{if } h_i(x) > 0 \\
0, & \text{if } h_i(x) = 0, \text{ and } \\
1, & \text{if } h_i(x) < 0
\end{cases} \tag{6}
\]
\[
\nabla g_j(x)^T d(x) \leq -1, \ j \in J(x^*). \tag{7}
\]

Now choose $\delta_1 \in (0, \delta]$ such that $g_j(x) < 0$ for $x \in N(x^*; \delta_1)$ and $j \not\in J(x^*)$. Hence, for $x \in N(x^*; \delta_1)$ if $g_j(x) \geq 0$ then $j \in J(x^*)$. Let $x \in N(x^*; \delta_1)$ and $x \not\in \mathcal{F}$. Then there must exist at least an index $i$ such that $|h_i(x)| > 0$ or an index $j \in J(x^*)$ such that $g_j(x) > 0$. Using Proposition 2.5, and relations (6) and (7) we can write
\[
DP(x, \rho; d(x)) = \nabla f(x)^T d(x) + \rho \max_{(\alpha, \beta) \in \partial \|[h(x), g^+(x)]\|} [\alpha^T J_h(x) d(x) + \beta J_g(x) d(x)]
\]
\[
\leq \nabla f(x)^T d(x) + \rho \max_{(\alpha, \beta) \in \partial \|[h(x), g^+(x)]\|} \left[ \sum_{i \in I_+} -\alpha_i + \sum_{i \in I_-} \alpha_i \right]
\]
\[
\sum_{j \in J_+} -\beta_j + \sum_{j \in J_-} \beta_j \nabla g_j(x)^T d(x)
\]
where $I_+ = \{i \mid h_i(x) > 0\}$, $I_- = \{i \mid h_i(x) < 0\}$, $J_+ = \{j \mid g_j(x) \geq 0\} \subseteq J(x^*)$, and $J_- = \{j \mid g_j(x) < 0\}$. By Proposition 2.6 we have $\beta_j \geq 0$ for $j = 1, \ldots, m_2$,
$\beta_j = 0$ if $g_j(x) < 0$, $\alpha_i \geq 0$ for $h_i(x) > 0$, and $\alpha_i \leq 0$ for $h_i(x) < 0$. Note that $(\alpha, \beta)$ is a nonzero vector. If $(\alpha, \beta)$ is a zero vector then we would have $\|h(x), g^+(x)\| = (\alpha, \beta)^T(h(x), g(x)) = 0$, this is, $x \in \mathcal{F}$, a contradiction. Thus, for all $x \in N(x^*; \delta_1)$ with $x \not\in \mathcal{F}$ we obtain
\[
DP(x, \rho; d(x)) \leq \nabla f(x)^T d(x) - \rho a, \text{ where } a > 0
\]
\[
\leq \|\nabla f(x)\|_2 \|d(x)\|_2 - \rho a.
\]
Hence $DP(x^k, \rho_k; d(x^k)) < 0$ for $\rho_k$ sufficiently large. This contradicts the fact that $x^k$ is a stationary point of $P(x, \rho_k)$. \hspace{1cm} \Box

**Theorem 2.9.** Let $\| \cdot \|$ be any given monotonic norm in $\mathbb{R}^{m_1 + m_2}$, and $P(x, \rho)$ be its corresponding penalty function defined as in (2). Let $x^* \in \mathcal{F}$ and suppose that (CQ) is satisfied at $x^*$. Then there exists an open neighborhood $N(x^*; \delta)$ of $x^*$ and a number $\bar{\rho} > 0$ such that for all $\rho \geq \bar{\rho}$, if $x_\rho \in N(x^*; \delta)$ is a local minimum point of $P(x, \rho)$ then $x_\rho$ is also a local minimum of problem (1).
Proof. As \( x^* \in \mathcal{F} \) and (CQ) is satisfied at \( x^* \), by Theorem 2.8, there exists an open neighborhood \( N(x^*; \delta) \) and a \( \bar{\rho} > 0 \) such that for all \( \rho \geq \bar{\rho} \), if \( x_\rho \in N(x^*; \delta) \) is a local minimizer of \( P(x, \rho) \) then \( x_\rho \in \mathcal{F} \). Hence, for \( \rho \geq \bar{\rho} \) there exists an open neighborhood \( N(x_\rho; \epsilon) \) such that

\[
 f(x_\rho) = P(x_\rho, \rho) \leq P(x, \rho) \text{ for all } x \in N(x_\rho; \epsilon).
\]

Therefore,

\[
 f(x_\rho) \leq f(x) \text{ for all } x \in N(x_\rho; \epsilon) \cap \mathcal{F},
\]

this is, \( x_\rho \) is a local minimizer of problem (1).

Under first-order sufficient optimality conditions for problem (1), the next theorem establishes the exactness of \( P(x, \rho) \). These conditions were considered by Howe in [7], where he showed that the function \( P_1(x, \rho) \) is exact, ensuring only the existence of finite penalty parameter. In our proof, we work directly with the function \( P(x, \rho) \) and give an estimative of the size of the critical penalty parameter.

**Theorem 2.10.** Let \( \| \cdot \| \) be any given norm in \( \mathbb{R}^{m_1 + m_2} \) and \( P(x, \rho) \) be its corresponding penalty function defined as in (2), and let \( \| \cdot \|' \) be its dual norm. Let \( x^* \in \mathcal{F} \), suppose that \( \nabla f(x^*) d > 0 \) for each \( d \neq 0 \) satisfying \( \nabla h_i(x^*) d = 0 \) for \( i = 1, \ldots, m_1 \) and \( \nabla g_j(x^*) d \leq 0 \) for \( j \in J(x^*) \). Then:

(i) There exists \( (\lambda^*, \mu^*) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}_+ \) such that \( x^* \) is a stationary point KKT for problem (1).

(ii) For \( \rho > \| \lambda^*, \mu^* \|' \), \( x^* \) is a strict local minimizer of \( P(x, \rho) \).

**Proof.** The part (i) follows directly from Farkas’ lemma. To show (ii), we assume that \( x^* \) is not a strict local minimizer of \( P(x, \rho) \). Then there exists a sequence \( \{x^k\} \) converging to \( x^* \) such that \( P(x^k, \rho) \leq P(x^*, \rho) \) for all \( k > 0 \). Thus,

\[
 P(x^k, \rho) - P(x^*, \rho) = f(x^k) - f(x^*) + \rho \| h(x^k), g^+(x^k) \| \leq 0. \quad (8)
\]

Define the sequence \( d^k = \frac{(x^k - x^*)}{\|x^k - x^*\|} \) and take convergent subsequence such that \( \{d^k\} \) converges to \( d \) with \( \|d\| = 1 \). Taking the limit in (8), we have

\[
 \nabla f(x^*) d + \rho \| h(x^*) d, J_g(x^*) d \| \leq 0. \quad (9)
\]

By part (i), there exists \( (\lambda^*, \mu^*) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}_+ \) such that

\[
 \nabla f(x^*) + J_h(x^*) \lambda^* + J_g(x^*) \mu^* = 0. \quad (10)
\]

Combining (9), (10) and the inequality \( (\mu^*)^T J_g(x^*) d \leq (\mu^*)^T J_{g^+}(x^*) d \), we find

\[
 \rho \| J_h(x^*) d, J_{g^+}(x^*) d \| - (\lambda^*)^T J_h(x^*) d - (\mu^*)^T J_{g^+}(x^*) d \leq 0.
\]
Now, applying the generalized Cauchy inequality, we obtain

\[(\rho - \|\lambda^*, \mu^*\|')\|J_h(x^*)d, J_g(x^*)d\| \leq 0.\]

Because of \(\rho > \|\lambda^*, \mu^*\|'\), it follows that \(J_h(x^*)d = 0\) and \(J_g(x^*)d = 0\), which imply \(\nabla h_i(x^*)^Ty = 0\) for \(i = 1, \ldots, m_1\) and \(\nabla g_j(x^*)^Ty \leq 0\) for \(j \in J(x^*)\). It follows from hypothesis that \(\nabla f(x^*)^Ty > 0\). Therefore,

\[\nabla f(x^*)^Ty + \rho\|J_h(x^*)d, J_g(x^*)d\| > 0,\]

and thus we get a contradiction with (9).

\[\square\]

3 A lower bound to the penalty parameter

In this section we show our main result. Assuming, the exactness of the function \(P(x, \rho)\), we given an estimate of the size of the critical penalty parameter.

For this, we need the below propositions.

Proposition 3.1 ([8], p.9). Let \(x^* \in \mathbb{R}^n\) and \(I(x^*) = \{i | h_i(x^*) = 0\}\). Let \(DP_1(x^*, \rho; d)\) be the directional derivative of penalty function \(P_1(x, \rho)\) defined in (3). The following two statements are equivalent:

(i) For all \(d \in \mathbb{R}^n\), \(DP_1(x^*, \rho; d) \geq 0\).

(ii) There exist scalars \(\nu_i \in [-1, 1], i \in I(x^*), \) and \(\tau_j \in [0, 1], j \in J(x^*)\) such that

\[\nabla f(x^*) + \rho \left( \sum_{i \in I(x^*)} \nu_i \nabla h_i(x^*) + \sum_{i \notin I(x^*)} \sigma_i \nabla h_i(x^*) + \sum_{j \in J(x^*)} \tau_j \nabla g_j(x^*) + \sum_{j \notin J(x^*)} \gamma_j \nabla g_j(x^*) \right) = 0,\]

where \(\sigma_i = \text{sign}(h_i(x^*))\) and \(\gamma_j = \begin{cases} 0, & \text{if } g_j(x^*) < 0 \\ 1, & \text{if } g_j(x^*) > 0. \end{cases}\)

Proposition 3.2. Let \(\| \cdot \|_a\) and \(\| \cdot \|_b\) be norms in \(\mathbb{R}^n\), with \(\| \cdot \|_a\) and \(\| \cdot \|_b\) their respective dual norms. If \(\| \cdot \|_a \leq c\| \cdot \|_b\) for some \(c \in \mathbb{R}_+\) then \(\| \cdot \|_b' \leq c\| \cdot \|_a'\).

Proof. Let \(x \in \mathbb{R}^n\). We have

\[\|x\|_b' = \max_{y \neq 0} \frac{y^Ty}{\|y\|_b} \leq \max_{y \neq 0} \frac{y^Ty}{\|y\|_a} = c \max_{y \neq 0} \frac{y^Ty}{\|y\|_a} = c\|x\|_a'.\]

\[\square\]
**Theorem 3.3.** Let \( x^* \) be a local minimizer for problem (1), and let \( \| \cdot \| \) be any given vector norm in \( \mathbb{R}^{m_1 + m_2} \) and \( P(x, \rho) \) its corresponding penalty function defined as in (2), and let \( \| \cdot \|' \) be its dual norm. If there exists a \( \bar{\rho} > 0 \) such that for all \( \rho \geq \bar{\rho} \), \( x^* \) is a local minimizer of the penalty function \( P(x, \rho) \) then:

(i) There exists \( (\lambda^*, \mu^*) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}_+ \) such that \( x^* \) is a stationary point KKT for problem (1).

(ii) \( \bar{\rho} \geq \frac{a}{b} \| \lambda^*, \mu^* \|' \), where \( a \) and \( b \) are the positive numbers such that

\[
a \| z \|_1 \leq \| z \| \leq b \| z \|_1 \text{ for all } z \in \mathbb{R}^{m_1 + m_2}.
\]

**Proof.** Since \( x^* \) is a local minimizer \( P(x, \bar{\rho}) \), and hence \( x^* \) is a stationary point of function \( P(x, \bar{\rho}) \), we have that \( DP(x^*, \bar{\rho}; d) \geq 0 \) for all \( d \in \mathbb{R}^n \). As \( x^* \in \mathcal{F} \), by Proposition 2.1 we can write

\[
DP(x^*, \bar{\rho}; d) = \nabla f(x^*)^T d + \bar{\rho} \| J_h(x^*) d, J_g^+(x^*) d \| \geq 0 \text{ for all } d \in \mathbb{R}^n. \tag{11}
\]

From equivalence of norms in finite a dimension vector space, there are positive numbers \( a \) and \( b \) such that \( a \| \cdot \|_1 \leq \| \cdot \| \leq b \| \cdot \|_1 \). Combining this inequality with (11), we obtain

\[
\nabla f(x^*)^T d + \bar{\rho} b \| J_h(x^*) d, J_g^+(x^*) d \|_1 \geq 0 \text{ for all } d \in \mathbb{R}^n,
\]

this is, the directional derivative of \( P_1(x, \bar{\rho} b) \) at \( x^* \) along the direction \( d \). Since \( x^* \in \mathcal{F} \), by Proposition 3.1 there exists \( (\alpha, \beta) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}_+ \) with \( \alpha_i \in [-1, 1] \), \( \beta_j \in [0, 1] \) if \( g_j(x^*) = 0 \), and \( \beta_j = 0 \) if \( g_j(x^*) < 0 \) satisfying

\[
\nabla f(x^*) + \bar{\rho} \left( \sum_{i=1}^{m_1} \alpha_i \nabla h_i(x^*) + \sum_{j=1}^{m_2} \beta_i \nabla g_j(x^*) \right) = 0.
\]

Fixing \( \lambda_i^* = \bar{\rho} \alpha_i \) for \( i = 1, \ldots, m_1 \), and \( \mu_j^* = \bar{\rho} \beta_j \) for \( j = 1, \ldots, m_2 \) in the above equation, we can write

\[
\nabla f(x^*) + J_h(x^*)^T \lambda^* + J_g(x^*)^T \mu^* = 0,
\]

\( \mu^* \geq 0 \) and \( \mu_j^* g_j(x^*) = 0 \) for \( j = 1, \ldots, m_2 \).

As \( x^* \in \mathcal{F} \), we have that \( x^* \) is a point KKT for problem (1) and \( (\lambda^*, \mu^*) \) are his Lagrange multipliers associated.

Now, applying the Proposition 3.2 to vector \( (\lambda^*, \mu^*) = \bar{\rho} (\alpha, \beta) \) and using the facts \( \alpha_i \in [-1, 1] \) and \( \beta_j \in [0, 1] \), we get

\[
\frac{a}{b} \| \lambda^*, \mu^* \|' = \frac{a}{b} \| \bar{\rho} \alpha, \bar{\rho} \beta \|' = a \bar{\rho} \| \alpha, \beta \|' \leq \bar{\rho} \| \alpha, \beta \|'_1 = \bar{\rho} \| \alpha, \beta \|_{\infty} \leq \bar{\rho},
\]

and proof is complete. \( \square \)
It is interesting to observe that independently of the conditions used to guarantee the exactness $P(x, \rho)$, the Theorem 3.3 provides an estimative of the size of the critical penalty parameter. Moreover, the theorem shows that exactness of $P(x, \rho)$ can be viewed as a regularity condition which ensures the Karush-Kuhn-Tucker conditions at local minimizer of (1). This regularity condition is weaker, it is implied by most constraint qualifications, for example, linear independence and the Mangasarian-Fromovitz conditions.

ACKNOWLEDGEMENT. The author is thankful to Dr. Rosana Pérez Mera for the careful reading of the manuscript.

References


Received: December 16, 2005