Mathematical Programming 17 (1979) 251–269.
North-Holland Publishing Company

EXACT PENALTY FUNCTIONS IN NONLINEAR PROGRAMMING*

S.-P. HAN

University of Illinois, Urbana, Illinois, U.S.A.

O.L. MANGASARIAN

University of Wisconsin, Madison, Wisconsin, U.S.A.

Received 10 November 1978
Revised manuscript received 13 March 1979

It is shown that the existence of a strict local minimum satisfying the constraint qualification of [16] or McCormick's [12] second order sufficient optimality condition implies the existence of a class of exact local penalty functions (that is ones with a finite value of the penalty parameter) for a nonlinear programming problem. A lower bound to the penalty parameter is given by a norm of the optimal Lagrange multipliers which is dual to the norm used in the penalty function.

Key words: Nonlinear Programming, Penalty Functions, Exact Penalty Functions, Constraint Qualification, Second Order Optimality Conditions.

1. Introduction

We shall be concerned here with the nonlinear programming problem

$$\begin{aligned} &\text{minimize} \quad f(x), \\ &\text{subject to} \quad g(x) \le 0, \quad h(x) = 0, \end{aligned} \tag{1.1}$$

where f, g and h are functions from $\mathbf{R}^n$ into $\mathbf{R}$, $\mathbf{R}^m$ and $\mathbf{R}^k$ respectively. A point x in $\mathbf{R}^n$ satisfying the constraints $g(x) \le 0$, $h(x) = 0$ is called *feasible.* A feasible point $\bar{x}$ such that $f(\bar{x}) \le f(x)$ for all feasible $x \ne \bar{x}$ in some neighborhood $N(\bar{x})$ of $\bar{x}$ is called a *local solution* of (1.1). If $f(\bar{x}) < f(x)$ then $\bar{x}$ is called a *strict local solution* of (1.1). We shall associate with this nonlinear programming problem the following class of penalty functions.

$$P(x, \alpha) := f(x) + \alpha Q(\|g(x)_+, h(x)\|), \tag{1.2}$$

where α is a nonnegative real number, $(g(x)_+)_j = \max\{0, g_j(x)\}$, $j = 1, \dots, m$, $\|\cdot\|$ is any fixed vector norm in $\mathbf{R}^{m+k}$, and Q is some function from the nonnegative real line $\mathbf{R}_+$ into itself with the following properties

$$Q(0) = 0, \tag{1.3a}$$

* Sponsored by the United States Army under Contract No. DAAG29-75-C-0024 and by the National Science Foundation under Grant No. MCS74-20584 A02.

$$Q(\zeta) > 0 \quad \text{for} \quad \zeta > 0, \tag{1.3b}$$

$$\infty > Q'(0+) := \lim_{\zeta \to 0+} \frac{Q(\zeta) - Q(0)}{\zeta} > 0. \tag{1.3c}$$

Obviously the third condition of (1.3) is equivalent to $Q'(0)$ being positive and finite when Q is differentiable at 0. Included in this class of penalty functions is the classical exact penalty function

$$P_1(x, \alpha) := f(x) + \alpha\left(\sum_{j=1}^{m} g_j(x)_+ + \sum_{j=1}^{k} |h_j(x)|\right) \tag{1.4}$$

which is obtained from (1.2) by setting $Q(\zeta) = \zeta$ and using the one norm. With some exceptions [1, 2, 4] most of the literature on exact penalty functions is generally devoted to this particular penalty function [3, 10, 14, 17, 22, 23] and is mainly concerned with conditions that ensure that $P_1(x, \alpha)$ has a local (global) minimum at a local (global) minimum of (1.1) for all sufficiently large but *finite* values of α. The best known among these conditions is probably the one due to Pietrzykowski [17] which requires the linear independence of the gradients of all the equality constraints and of the active inequality constraints, that is those inequalities satisfied as equalities at the point being considered. One of our principal results, Theorem 4.4, is more natural than Pietrzykowski's result which it subsumes. It is more natural because it merely requires the constraint qualification of [16]. This constraint qualification besides ensuring the satisfaction of the Karush–Kuhn–Tucker conditions at local minima of (1.1) has been shown to be a necessary and sufficient condition for the constraints of (1.1) to be stable under small perturbations [20]. In this sense this constraint qualification may be viewed as the minimum requirement for a problem to be numerically well-posed. In [3] by using the completely different approach of multifunction theory a local minimum for $P_1(x, \alpha)$ is also established under a "controllability condition" which turns out to be equivalent to the constraint qualification of [16]. Our generalization of the penalty function $P_1(x, \alpha)$ to the class $P(x, \alpha)$, which subsumes that of [2], is not merely generalization for its own sake but in order to allow us to handle special functions Q of other norms in (1.2) and in particular the infinity and two norms which can be used to obtain improved quasi-Newton computational algorithms [6, 7, 8, 19]. We also note that the classical exterior penalty function [5], which can also be obtained from (1.2) by using the two norm and letting $Q(\zeta) = \zeta^2$, violates however the requirement (1.3) because $Q'(0) = 0$. This is as expected because it is well known that for the classical exterior penalty function the penalty parameter α is not finite. (See however, an interesting exception to this for linear programs in [1] and references therein.) Using instead $Q(\zeta) = \zeta$ or $Q(\zeta) = \zeta + \zeta^2$ with the two norm would however result with an *exact* penalty function which would again be nondifferentiable.

Because of the significant role played in this paper by the constraint qualification of [16], Section 2 of this paper will be devoted to the derivation of an equivalent statement of this constraint qualification which will be used in deriving one of our principal results, Theorem 4.4. Section 3 is devoted to second order sufficient optimality conditions which also play an important role in establishing the existence of exact penalty minimum points. In particular we derive a second order sufficient optimality condition of the Fritz John type (Theorem 3.1) which subsumes McCormick's well known second order sufficient optimality condition [5, 12]. We also give an equivalent formulation (3.6) of McCormick's second order condition (3.9) which may be used to derive second order optimality conditions for quadratic programming without any knowledge of the optimal Lagrange multipliers (Corollary 3.6). Section 4 contains our principal results pertaining to the class of exact penalty function $P(x, \alpha)$ defined by (1.2). Theorem 4.1 shows that the existence of an exact penalty function minimum point implies the existence of a minimum point to the nonlinear programming problem (1.1). Theorem 4.2 shows that local solutions of exact penalty functions within the class given by (1.2) are the same. Theorem 4.4 shows that for sufficiently large but finite α, $P(x, \alpha)$ has a local minimum point at any strict local minimum point $\bar{x}$ of (1.1) which satisfies the constraint qualification of [16]. In Theorem 4.6 we show that McCormick's second order sufficiency conditions imply that $P(x, \alpha)$ has a strict local minimum for all values of the penalty parameter α that are larger than 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 (1.2). In Theorem 4.8 we show that the existence of a local minimum of $P(x, \alpha)$ for all sufficiently large α implies, under suitable assumptions, the satisfaction of the Karush–Kuhn–Tucker conditions [11] for problem (1.1). In our final theorem, Theorem 4.9, we treat the convex case and again establish the fact that the generalized Slater constraint qualification [13] implies that $P(x, \alpha)$ has a global minimum for all values of the penalty parameter larger or equal to the lower bound established in Theorem 4.6. We note that Theorem 4.6 and its corollary, Corollary 4.7, subsume and sharpen Theorem 2 of [2], while Theorems 4.9 and 4.1 subsume and sharpen Theorem 1 of [2].

To simplify notation a vector is either a row or a column vector depending on the context. For example, the inner product of two vectors x and y is written simply as xy rather than $x^{\mathrm{T}}y$.

2. Equivalent forms of the constraint qualification

We begin by recalling the following definition of the constraint qualification.

Definition 2.1. [13, 16]. Let $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $I = \{i \mid g_i(\bar{x}) = 0, i = 1, \ldots, m\}$. The constraints $g(x) \leq 0$, $h(x) = 0$ are said to satisfy the constraint qualification of [16] at $\bar{x}$ if g is differentiable at $\bar{x}$, h is continuously differentiable at $\bar{x}$, and

$$\begin{aligned}&\nabla h_i(\bar{x}),\quad i=1,\dots,k \text{ are linearly independent}\\ &\qquad\text{and, there exists a } z\in\mathbf{R}^n \text{ such that}\\ &\nabla g_i(\bar{x})z<0,\quad i\in I,\\ &\nabla h_i(\bar{x})z=0,\quad i=1,\dots,k.\end{aligned}\tag{2.1}$$

It can be shown by using theorems of the alternative [12] that (2.1) is equivalent to the following condition.

There exist *no* u_i, $i\in I$ and v_i, $i=1,\dots,k$ such that

$$\begin{aligned}&\sum_{i\in I} u_i\nabla g_i(\bar{x})+\sum_{i=1}^{k} v_i\nabla h_i(\bar{x})=0,\\ &u_i\geq 0,\quad i\in I,\\ &(u_i,\ i\in I,\ v_i,\ i=1,\dots,k)\neq 0.\end{aligned}\tag{2.2}$$

We state and prove now an alternate formulation of (2.1) that will be needed in deriving our exact penalty results.

Theorem 2.2 (Constraint qualification equivalence). *Let* $g(\bar{x})\leq 0$, $h(\bar{x})=0$, $I=\{i \mid g_i(\bar{x})=0,\ i=1,\dots,m\}$ *and let* g *and* h *be continuously differentiable at* $\bar{x}$. *The constraint qualification* (2.1) *is satisfied at* $\bar{x}$ *if and only if there exists an open neighborhood* $N(\bar{x};\epsilon)$ *of* $\bar{x}$ *such that*

For each bounded function $b(x)$: $N(\bar{x};\epsilon)\to\mathbf{R}^k$ *there exists a bounded function* $d(x)$: $N(\bar{x};\epsilon)\to\mathbf{R}^n$ *such that for all* x *in* $N(\bar{x};\epsilon)$

$$\begin{aligned}&\nabla g_i(x)\,\mathrm{d}(x)\leq -1,\quad i\in I,\\ &\nabla h_i(x)\,\mathrm{d}(x)=b_i(x),\quad i=1,\dots,k.\end{aligned}\tag{2.3}$$

Proof. (2.3)$\Rightarrow$(2.1): Just set $b(x)=0$ and $x=\bar{x}$ in (2.3) and note that for each b in $\mathbf{R}^k$, $\nabla h_i(\bar{x})z=b_i$, $i=1,\dots,k$, has a solution z in $\mathbf{R}^n$. (2.1)$\Rightarrow$(2.3): Because $\nabla h_i(\bar{x})$, $i=1,\dots,k$ are linearly independent it follows that $k\leq n$. Choose $n-k$ vectors in $\mathbf{R}^n$, $w^1, w^2,\dots,w^{n-k}$ such that $\{\nabla h_1(\bar{x}),\dots,\nabla h_k(\bar{x}), w^1,\dots,w^{n-k}\}$ are linearly independent. Define the $n\times n$ matrix function $A(x)$ as follows:

$$A(x)=\begin{bmatrix}\nabla h_1(x)\\ \vdots\\ \nabla h_k(x)\\ w^1\\ \vdots\\ w^{n-k}\end{bmatrix}$$

Since $A(\bar{x})$ is nonsingular there exists an $\epsilon>0$ such $A^{-1}(x)$ exists and is bounded in $N(\bar{x};\epsilon)$. By (2.1) there exists a vector $\bar{z}$ in $\mathbf{R}^n$ such that

$$\nabla g_i(\bar{x})\bar{z} < 0, \quad i \in I,$$
$$\nabla h_i(\bar{x})\bar{z} = 0, \quad i = 1, \dots, k.$$

Define $z(x) = A^{-1}(x)c$ where

$$c = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ w^1\bar{z} \\ \vdots \\ w^{n-k}\bar{z} \end{bmatrix} \in \mathbf{R}^n$$

Clearly $z(\bar{x}) = \bar{z}$ and $z(x)$ is continuous in $N(\bar{x}; \epsilon)$. Thus we can shrink ϵ, if necessary, so that

$$\nabla g_i(x)z(x) \leq -\frac{\gamma}{2} \quad \text{for} \quad x \in N(\bar{x}; \epsilon) \quad \text{and} \quad i \in I,$$

where

$$-\gamma = \max_{i \in I} \{\nabla g_i(\bar{x})\bar{z}\} < 0.$$

We also have that

$$\nabla h_i(x)z(x) = 0 \quad \text{for} \quad x \in N(\bar{x}; \epsilon) \quad \text{and} \quad i = 1, \dots, k.$$

Let $b(x)$ be any given bounded function from $N(\bar{x}; \epsilon)$ into $\mathbf{R}^k$, let

$$\bar{b}(x) = \begin{bmatrix} b(x) \\ 0 \end{bmatrix} \in \mathbf{R}^n,$$

and let $y(x) = A(x)^{-1}\bar{b}(x)$. The function $y(x)$ is bounded in $N(\bar{x}; \epsilon)$ and furthermore

$$\nabla h_i(x)y(x) = b_i(x), \quad i = 1, \dots, k.$$

Let

$$d(x) = \beta z(x) + y(x)$$

where

$$\beta = \frac{2(1+\lambda)}{\gamma}$$

and

$$\lambda = \max_{i \in I} \sup_{x \in N(\bar{x};\epsilon)} \{\nabla g_i(x)y(x)\}.$$

Hende $d(x)$ is bounded and satisfies (2.3).

We note that the more stringent constraint qualification used by Pietrzykowski [17], namely that the gradients $\nabla g_i(\bar{x})$, $i \in I$, $\nabla h_1(\bar{x}), \dots, \nabla h_k(\bar{x})$, are linearly independent, implies the constraint qualification (2.2) and hence its equivalents (2.1) and (2.3).

3. Second order sufficient optimality conditions

We first derive in this section a second order sufficient optimality condition of the Fritz John type for problem (1.1) which subsumes the standard second order sufficiency condition of McCormick [12].

Theorem 3.1 (Generalized second order sufficiency). *Let $\bar{x}$ be a local solution of* (1.1) *or let $(\bar{x}, \bar{u}_0, \bar{u}, \bar{v}) \in \mathbf{R}^{n+1+m+k}$ satisfy the Fritz John necessary optimality conditions for problem* (1.1):

$$\begin{aligned} &\bar{u}_0\nabla f(\bar{x}) + \sum_{i=1}^{m} \bar{u}_i\nabla g_i(\bar{x}) + \sum_{j=1}^{k} \bar{v}_j\nabla h_j(\bar{x}) = 0, \\ &(\bar{u}_0, \bar{u}) \geq 0, \quad (\bar{u}_0, \bar{u}, \bar{v}) \neq 0, \\ &\bar{u}g(\bar{x}) = 0, \quad g(\bar{x}) \leq 0, \quad h(\bar{x}) = 0. \end{aligned} \tag{3.1}$$

Let f, g and h be twice differentiable at $\bar{x}$, let $I = \{i \mid g_i(\bar{x}) = 0, i = 1, \ldots, m\}$ and let

$$\left.\begin{aligned} &\nabla f(\bar{x})x \leq 0 \\ &\nabla g_i(\bar{x})x \leq 0, \quad i \in I \\ &\nabla h_i(\bar{x})x = 0, \quad i = 1, \ldots, k \\ &x \neq 0 \end{aligned}\right\} \Rightarrow x\nabla_{11}L^0(\bar{x}, \bar{u}_0, \bar{u}, \bar{v})x > 0 \tag{3.2}$$

where

$$L^0(x, u_0, u, v) = u_0f(x) + ug(x) + vh(x) \tag{3.3}$$

and $\nabla_{11}L^0(x, u_0, u, v)$ denotes the $n \times n$ Hessian of $L(x, u_0, u, v)$ with respect to its first argument x. Then $\bar{x}$ is a strict local minimum of (1.1).

Proof. We shall assume that $\bar{x}$ is not a strict local minimum of (1.1) and exhibit a contradiction. Since $\bar{x}$ is assumed not to be a strict local minimum of (1.1), there exists a sequence of feasible points $\{x^j\}$, that is $g(x^j) \leq 0$ and $h(x^j) = 0$, converging to $\bar{x}$, such that $f(x^j) \leq f(\bar{x})$ and $x^j \neq \bar{x}$. Hence

$$0 \geq \frac{f(x^j) - f(\bar{x})}{\|x^j - \bar{x}\|} = \nabla f(\bar{x})\frac{(x^j - \bar{x})}{\|x^j - \bar{x}\|} + \frac{o(\|x^j - \bar{x}\|)}{\|x^j - \bar{x}\|},$$

$$0 \geq \frac{g_i(x^j) - g_i(\bar{x})}{\|x^j - \bar{x}\|} = \nabla g_i(\bar{x})\frac{(x^j - \bar{x})}{\|x^j - \bar{x}\|} + \frac{o(\|x^j - \bar{x}\|)}{\|x^j - \bar{x}\|}, \quad i \in I,$$

$$0 = \frac{h_i(x^j) - h_i(\bar{x})}{\|x^j - \bar{x}\|} = \nabla h_i(\bar{x})\frac{(x^j - \bar{x})}{\|x^j - \bar{x}\|} + \frac{o(\|x^j - \bar{x}\|)}{\|x^j - \bar{x}\|}, \quad i = 1, \ldots, k.$$

Hence there exists an accumulation point $\bar{s}$ of the sequence $\{s^j\} := \{x^j - \bar{x}/\|x^j - \bar{x}\|\}$ such that

$$\begin{aligned} &\|\bar{s}\| = 1, \quad \nabla f(\bar{x})\bar{s} \leq 0, \quad \nabla g_i(\bar{x})\bar{s} \leq 0, \quad i \in I, \\ &\nabla h_i(\bar{x})\bar{s} = 0, \quad i = 1, \ldots, k. \end{aligned} \tag{3.4}$$

Making use of the twice differentiability property now gives

$$0 \geq \frac{f(x^j)-f(\bar{x})}{\|x^j-\bar{x}\|^2} = \frac{\nabla f(\bar{x})s^j}{\|x^j-\bar{x}\|} + \tfrac{1}{2}s^j\nabla^2 f(\bar{x})s^j + \frac{o(\|x^j-\bar{x}\|)}{\|x^j-\bar{x}\|},$$

$$0 \geq \frac{g_i(x^j)-g_i(\bar{x})}{\|x^j-\bar{x}\|^2} = \frac{\nabla g_i(\bar{x})s^j}{\|x^j-\bar{x}\|} + \tfrac{1}{2}s^j\nabla^2 g_i(\bar{x})s^j + \frac{o(\|x^j-\bar{x}\|)}{\|x^j-\bar{x}\|}, \quad i \in I,$$

$$0 = \frac{h_i(x^j)-h_i(\bar{x})}{\|x^j-\bar{x}\|^2} = \frac{\nabla h_i(\bar{x})s^j}{\|x^j-\bar{x}\|} + \tfrac{1}{2}s^j\nabla^2 h_i(\bar{x})s^j + \frac{o(\|x^j-\bar{x}\|)}{\|x^j-\bar{x}\|}, \quad i = 1, \ldots, k.$$

Multiplication of the above relations respectively by $\bar{u}_0$, $\bar{u}_i$, $i \in I$, $\bar{v}_i$, $i = 1, \ldots, k$, summing and making use of the first equality of the Fritz John conditions (3.1) which must hold when $\bar{x}$ is a local solution of (1.1) [13, 16] gives

$$0 \geq \tfrac{1}{2}s^j\nabla_{11}L(\bar{x}, \bar{u}_0, \bar{u}, \bar{v})s^j + \frac{o(\|x^j-\bar{x}\|)}{\|x^j-\bar{x}\|}.$$

Hence the accumulation point $\bar{s}$ of $\{s^j\}$ satisfies

$$\bar{s}\nabla_{11}L(\bar{x}, \bar{u}_0, \bar{u}, \bar{v})\bar{s} \leq 0.$$

This inequality together with (3.4) contradict (3.2).

We state now a paraphrase of McCormick's second order sufficient optimality conditions which may have certain advantages over the standard way [5, 12] these conditions are stated. We will show that the paraphrase and standard statements are equivalent, and we discuss below some of the advantages of the paraphrase.

Theorem 3.2 (Paraphrase of McCormick's second order sufficiency). *Let $(\bar{x}, \bar{u}, \bar{v}) \in \mathbf{R}^{n+m+k}$ satisfy the Karush–Kuhn–Tucker necessary optimality conditions for problem* (1.1)

$$\nabla f(\bar{x}) + \sum_{j=1}^{m} \bar{u}_j\nabla g_j(\bar{x}) + \sum_{j=1}^{k} \bar{v}_j\nabla h_j(\bar{x}) = 0, \qquad \bar{u} \geq 0, \quad \bar{u}g(\bar{x}) = 0, \quad g(\bar{x}) \leq 0, \quad h(\bar{x}) = 0. \tag{3.5}$$

Let f, g and h be twice differentiable at $\bar{x}$, let $I = \{i \mid g_i(\bar{x}) = 0, i = 1, \ldots, m\}$ and let

$$\left.\begin{array}{ll} \nabla f(\bar{x})x \leq 0, & \\ \nabla g_i(\bar{x})x \leq 0, & i \in I \\ \nabla h_i(\bar{x})x = 0, & i = 1, \ldots, k \\ x \neq 0 & \end{array}\right\} \Rightarrow x\nabla_{11}L(\bar{x}, \bar{u}, \bar{v})x > 0, \tag{3.6}$$

where

$$L(x, u, v) = f(x) + ug(x) + vh(x). \tag{3.7}$$

Then $\bar{x}$ is a strict local minimum of (1.1).

Remark 3.3. Theorem 3.1 subsumes Theorem 3.2 because whenever the Karush–Kuhn–Tucker conditions (3.5) are satisfied, so are the Fritz John conditions (3.1) with $\bar{u}_0 = 1$. The following simple example shows that there are indeed cases which are covered by Theorem 3.1 and not by Theorem 3.2:

$$\begin{aligned} &\text{minimize} && x_1, \\ &\text{subject to} && x_1^2 - x_2 \le 0, \\ &&& x_1^2 + x_2 \le 0. \end{aligned} \tag{3.8}$$

The origin in $\mathbf{R}^2$ is the only feasible point and hence is a strict local solution. Theorem 3.1 can be used to verify the uniqueness of the solution because the Fritz John conditions are satisfied, whereas because the Karush–Kuhn–Tucker conditions are not satisfied, Theorem 3.2 cannot be employed. The same example (3.8) can be used to show that the origin is not a local minimum of $P_1(x, \alpha)$ as defined in (1.4) for this problem. Hence the second order Fritz John conditions cannot guarantee the existence of a local minimum for $P_1(x, \alpha)$. We will show however in Theorem 4.6 that McCormick's second order sufficient optimality conditions are sufficient to ensure that all exact penalty functions as defined by (1.2) have a strict local minimum.

Remark 3.4. The standard way of stating the second order sufficiency condition is to replace the implication (3.6) by the following equivalent one

$$\left.\begin{aligned} \nabla g_i(\bar{x})x &= 0, \quad i \in J \\ \nabla g_i(\bar{x})x &\le 0, \quad i \in K \\ \nabla h_i(\bar{x})x &= 0, \quad i = 1, \dots, k \\ x &\ne 0. \end{aligned}\right\rangle \Rightarrow x\nabla_{11}L(\bar{x}, \bar{u}, \bar{v})x > 0, \tag{3.9}$$

where J and K are the following subsets of I:

$$\begin{aligned} J &= \{i \mid g_i(\bar{x}) = 0,\ \bar{u}_i > 0,\ i = 1, \dots, m\}, \\ K &= \{i \mid g_i(\bar{x}) = 0,\ \bar{u}_i = 0,\ i = 1, \dots, m\}. \end{aligned} \tag{3.10}$$

That implication (3.9) is equivalent to implication (3.6) can be easily established as shown by the following theorem.

Theorem 3.5 (Equivalence of (3.6) and (3.9)). *Under the assumptions of Theorem* 3.2 *implications* (3.6) *and* (3.9) *are equivalent.*

Proof. We will show that, under the assumptions of Theorem 3.2, the sets S and T in $\mathbf{R}^n$ satisfying the conditions on the left-hand side of implications (3.6) and (3.9) respectively are equal.

We first show that $S \subset T$. We assume that S is nonempty, otherwise the implication is trivially true. Let x be in S. Clearly, we only need to show that for $j \in J$, $\nabla g_j(\bar{x})x = 0$. By (3.5) we have that

$$\nabla f(\bar{x})x + \sum_{j\in I} \bar{u}_j \nabla g_j(\bar{x})x + \sum_{j=1}^{k} \bar{v}_j \nabla h_j(\bar{x})x = 0.$$

Because $\nabla h_j(\bar{x})x = 0$ for $j = 1, \ldots, k$ and $\bar{u}_j = 0$ for $j \in K$, we have

$$\nabla f(\bar{x})x + \sum_{j\in J} \bar{u}_j \nabla g_j(\bar{x})x = 0.$$

Because each term in the above equation is nonpositive and $\bar{u}_j > 0$ for $j \in J$, we then have

$$\nabla g_j(\bar{x})x = 0 \quad \text{for} \quad j \in J.$$

We now prove that $T \subset S$. Again we assume that T is nonempty and let x be any point in T. It suffices to show that $\nabla f(\bar{x})x \le 0$. As before, we have

$$\nabla f(\bar{x})x + \sum_{j\in I} \bar{u}_j \nabla g_j(\bar{x})x + \sum_{j=1}^{k} \bar{v}_j \nabla h_j(\bar{x}) = 0.$$

Clearly $\nabla f(\bar{x})x = 0$ because all the other terms are zeros. The proof is then complete.

We give now an interpretation of the implication (3.6). The set of x in $\mathbf{R}^n$ satisfying the left-hand side conditions of (3.6) can be seen [15] to be the set of directions along which the linearized problem, obtained by linearizing (1.1) around $\bar{x}$, has nonunique solutions. In order to have uniqueness for the nonlinear problem, implication (3.6) requires that the Hessian of the Lagrangian be positive definite along these directions. Besides having this simple interpretation, implication (3.6) is also simpler than (3.9) because the left-hand side conditions of (3.6) do not require any information on the multiplier vector $\bar{u}$ whereas the corresponding conditions of (3.9) do. As an example of the usefulness of this fact we give below a sufficient condition for the existence of a strict local minimum point for a quadratic programming problem which does not require the knowledge of any of the multipliers.

Corollary 3.6 (Sufficient conditions for a strict local minimum in quadratic programming). *Let $\bar{x}$ be a local solution of the quadratic program*

$$\begin{aligned} &\text{minimize} && \tfrac{1}{2}xQx + px,\\ &\text{subject to} && Ax \le b,\\ & && Cx = d, \end{aligned} \tag{3.11}$$

where Q, A and C are $n \times n$, $m \times n$ and $k \times n$ matrices respectively with Q symmetric, and p, b and d are vectors in $\mathbf{R}^n$, $\mathbf{R}^m$ and $\mathbf{R}^k$ respectively. Let $I = \{i \mid A_i\bar{x} = b_i,\ i = 1, \ldots, m\}$. If

$$\left.\begin{array}{l}(\bar{x}Q+p)x \leq 0 \\ \quad A_i x \leq 0, \quad i \in I \\ \quad Cx = 0\end{array}\right\rangle \Rightarrow xQx > 0, \tag{3.12}$$

then $\bar{x}$ *is a strict local minimum of* (3.11).

4. Exact penalty functions

We derive in this section our principal results which relate local (global) solutions of the penalty function (1.2) to local (global) solutions of the nonlinear programming problem (1.1). Our vehicle for deriving many of the results of this section will be the classical exact penalty function $P_1(x, \alpha)$ defined in (1.4). Because we wish to establish these results for the more general penalty function of (1.2) we establish an important equivalence between members of the class of penalty functions given by (1.2) in Theorem 4.2 below. Before doing this we establish the sufficiency of the existence of an exact penalty minimum point for the existence of a minimum point to the nonlinear programming problem. This theorem was given in [14] without proof.

Theorem 4.1 (Sufficiency of exact penalty minimum). *If there exists an* $\bar{\alpha} \geq 0$ *such that for all* $\alpha \geq \bar{\alpha}$, $P(\bar{x}, \alpha) \leq P(x, \alpha)$ *for all* x *in some set* Y *containing* $\bar{x}$ *and some feasible point of* (1.1), *then* $\bar{x}$ *solves* (1.1) *subject to the extra condition that* $x \in Y$.

Proof. We first show by contradiction that $\bar{x}$ must be feasible for problem (1.1). If $\bar{x}$ is infeasible then $Q(\|g(\bar{x})_+, h(\bar{x})\|) > 0$. Choose any feasible point $\hat{x}$ which is also in Y and let

$$\alpha > \max\left\{\frac{f(\hat{x}) - f(\bar{x})}{Q(\|g(\bar{x})_+, h(\bar{x})\|)}, \bar{\alpha}\right\}.$$

We then have

$$f(\hat{x}) = P(\hat{x}, \alpha) \geq P(\bar{x}, \alpha) = f(\bar{x}) + \alpha Q(\|g(\bar{x})_+, h(\bar{x})\|) > f(\hat{x}),$$

where the last inequality follows from the choice of α. This gives a contradiction and hence $\bar{x}$ is feasible for (1.1). To show that $\bar{x}$ is optimal for (1.1) let x be any other feasible point for (1.1) which is also in Y and let $\alpha \geq \bar{\alpha}$. Then

$$f(\bar{x}) = P(\bar{x}, \alpha) \leq P(x, \alpha) = f(x),$$

and hence $\bar{x}$ solves (1.1) with the added restriction that $x \in Y$.

We show now that local solutions of exact penalty functions of the class given by (1.2) are the same.

Theorem 4.2 (Equality of local solutions of exact penalty functions). *Let* $\|\cdot\|_\mu$ *and* $\|\cdot\|_\nu$ *denote two vector norms in* $\mathbf{R}^{m+k}$ *and let the corresponding exact penalty functions defined by* (1.2) *be denoted by* $P_\mu(x, \alpha)$ *and* $P_\nu(x, \alpha)$ *with corresponding* Q_μ *and* Q_ν *satisfying* (1.3). *If there exists an* $\bar{x}$ *in* $\mathbf{R}^n$, *an* $\bar{\alpha}_\mu \geq 0$, *and a neighborhood* $N_\mu(\bar{x})$ *of* $\bar{x}$ *containing some feasible point of* (1.1) *such that g and h are continuous on* $N_\mu(\bar{x})$ *and*

$$P_\mu(\bar{x}, \alpha) \leq P_\mu(x, \alpha) \quad \text{for all } x \in N_\mu(\bar{x}) \quad \text{and} \quad \alpha \geq \bar{\alpha}_\mu,$$

then there exists an $\bar{\alpha}_\nu \geq 0$ *and a neighborhood* $N_\nu(\bar{x})$ *containing some feasible point of* (1.1) *such that*

$$P_\nu(\bar{x}, \alpha) \leq P_\nu(x, \alpha) \quad \text{for all } x \in N_\nu(\bar{x}) \quad \text{and} \quad \alpha \geq \bar{\alpha}_\nu$$

where

$$\bar{\alpha}_\nu = \left(\frac{1+\epsilon}{1-\epsilon}\right) \frac{Q'_\mu(0+)\bar{\alpha}_\mu}{Q'_\nu(0+)\gamma_{\mu\nu}} \quad \text{for any } \epsilon \in (0, 1),$$

and $\gamma_{\mu\nu}$ *is the positive number relating the* μ-norm and the ν-norm in $\mathbf{R}^{m+k}$ *by*

$$\gamma_{\mu\nu} \leq \frac{\|y\|_\nu}{\|y\|_\mu} \quad \text{for all nonzero } y \in \mathbf{R}^{m+k}.$$

Proof. By Theorem 4.1, $\bar{x}$ is feasible for problem (1.1). Choose $\epsilon \in (0, 1)$ and $\bar{t} > 0$ such that $(1+\epsilon)Q'_\mu(0+)t \geq Q_\mu(t)$ and $Q_\nu(t) \geq (1-\epsilon)Q'_\nu(0+)t$ for $t \in [0, \bar{t}]$ and choose $N_\nu(\bar{x}) \subset N_\mu(\bar{x})$ sufficiently small such that $\|g(x)_+, h(x)\|_\nu \leq \bar{t}$ and $\|g(x)_+, h(x)\|_\mu \leq \bar{t}$ for $x \in N_\nu(\bar{x})$. This is possible because $g(\bar{x})_+ = 0$, $h(\bar{x}) = 0$ and g and h are continuous on $N_\mu(\bar{x})$. Note that $N_\nu(\bar{x})$ contains a feasible point to problem (1.1), namely the point $\bar{x}$ itself. For any $\alpha \geq \bar{\alpha}_\nu$ and $x \in N_\nu(\bar{x})$ we have

$$\begin{aligned}
P_\nu(x, \alpha) &= f(x) + \alpha Q_\nu(\|g(x)_+, h(x)\|_\nu) \\
&\geq f(x) + \alpha(1-\epsilon)Q'_\nu(0+)\|g(x)_+, h(x)\|_\nu \\
&\geq f(x) + \alpha\gamma_{\mu\nu}(1-\epsilon)Q'_\nu(0+)\|g(x)_+, h(x)\|_\mu \\
&\geq f(x) + \alpha\gamma_{\mu\nu}\left(\frac{1-\epsilon}{1+\epsilon}\right)\frac{Q'_\nu(0+)}{Q'_\mu(0+)} Q_\mu(\|g(x)_+, h(x)\|_\mu) \\
&\geq f(\bar{x}) + \alpha\gamma_{\mu\nu}\left(\frac{1-\epsilon}{1+\epsilon}\right)\frac{Q'_\nu(0+)}{Q'_\mu(0+)} Q_\mu(\|g(\bar{x})_+, h(\bar{x})\|_\mu) \\
&\qquad \text{(By definition of } \bar{\alpha}_\nu\text{)} \\
&= f(\bar{x}) \\
&= f(\bar{x}) + \alpha Q_\nu(\|g(\bar{x})_+, h(\bar{x})\|_\nu) \\
&= P_\nu(\bar{x}, \alpha).
\end{aligned}$$

To prove our next principal result, namely that when a strict local minimum of (1.1) exists satisfying (2.1), the constraint qualification of [16], a local minimum to the exact penalty function $P(x, \alpha)$ exists, we make use of the following result due to Pietrzykowski.

Lemma 4.3 (Pietrzykowski [18]). *Let f, g and h be continuous on a neighborhood of $\bar{x}$ and let $\bar{x}$ be a strict local minimum point of problem* (1.1). *There exists a number $\bar{\alpha} \geq 0$ such that for any $\alpha \geq \bar{\alpha}$ there exists a positive number $\epsilon(\alpha)$ and a vector $x(\alpha)$ in $\mathbf{R}^n$ such that*

(i) $x(\alpha) \in N(\bar{x}; \epsilon(\alpha))$
(ii) $\lim_{\alpha\to\infty} \epsilon(\alpha) = 0$
(iii) $P(x(\alpha), \alpha) \leq P(x, \alpha)$ *for all* $x \in N(\bar{x}; \epsilon(\alpha))$.

Theorem 4.4 (Strict local minimum and constraint qualification imply local minimum of exact penalty). *Let f, g and h be continuously differentiable on a neighborhood of a strict local minimum point of $\bar{x}$ of* (1.1) *and let the constraint qualification* (2.1) *hold at $\bar{x}$. Then for each norm $\|\cdot\|$ in $\mathbf{R}^{m+k}$ there exists an $\bar{\alpha} \geq 0$, such that for all $\alpha \geq \bar{\alpha}$, $\bar{x}$ is a local minimum of $P(x, \alpha)$, where $P(x, \alpha)$ is defined in* (1.2) *with Q satisfying* (1.3).

Proof. We will establish the result for $P_1(x, \alpha)$ of (1.4) and the theorem will follow, by virtue of Theorem 4.2, for all other $P(x, \alpha)$ defined by (1.2) with Q satisfying (1.3).

Let $\bar{x}$ be a strict local minimum of (1.1) in the neighborhood $N(\bar{x}; \bar{\epsilon})$. If $I = \{i \mid g_i(\bar{x}) = 0,\ i = 1, \dots, m\}$ is empty and there are no equality constraints $h(x) = 0$, the theorem is trivially true. So assume that I is nonempty or there exists at least one equality constraint. By Lemma 4.3, for all sufficiently large α, there exist $\epsilon(\alpha) > 0$ and $x(\alpha)$ such that $x(\alpha)$ is a local minimum point of $P_1(x, \alpha)$ in $N(\bar{x}; \epsilon(\alpha))$ and $\lim_{\alpha\to\infty} \epsilon(\alpha) = 0$. Let α be sufficiently large such that $\epsilon(\alpha) \leq \bar{\epsilon}$. If for such an α the point $x(\alpha)$ is feasible for problem (1.1), then by Lemma 4.3

$$f(\bar{x}) = P_1(\bar{x}, \alpha) \geq P_1(x(\alpha), \alpha) = f(x(\alpha)).$$

Because $\bar{x}$ is a strict local minimum of (1.1), we then have that $\bar{x} = x(\alpha)$ and hence $\bar{x}$ is a local minimum of $P(x, \alpha)$. Therefore to complete the proof we only need to show that $x(\alpha)$ is feasible for all sufficiently large α. We shall assume the contrary, that is there exists a sequence of positive numbers $\{\alpha_i\} \to \infty$ such that $x(\alpha_i)$ is infeasible for problem (1.1), and exhibit a contradiction. Let a neighborhood $N(\bar{x}; \epsilon)$ be defined as in Theorem 2.2 and consider the bounded function $b(x)$: $N(\bar{x}; \epsilon) \to \mathbf{R}^k$ defined by

$$b_i(x) = \begin{cases} -h_i(x)/|h_i(x)| & \text{if } h_i(x) \neq 0, \\ 0 & \text{if } h_i(x) = 0. \end{cases}$$

By Theorem 2.2 there exists a bounded function $d(x)$: $N(\bar{x}; \epsilon) \to \mathbf{R}^n$ such that for all $x \in N(\bar{x}; \epsilon)$

$$\nabla g_i(x) d(x) \leq -1, \quad i \in I,$$

$$\nabla h_i(x)d(x) = \begin{cases} -1 & \text{if } h_i(x) > 0, \\ 0 & \text{if } h_i(x) = 0, \\ 1 & \text{if } h_i(x) < 0. \end{cases}$$

Now choose $\epsilon_1 \in (0, \epsilon]$ such that $g_i(x) < 0$ for $x \in N(\bar{x}; \epsilon_1)$ and $i \notin I$. We then have for $x \in N(\bar{x}; \epsilon_1)$ and x infeasible for (1.1) the following directional derivative for $P_1(x, \alpha)$ of (1.4) in the direction $d(x)$

$$\begin{aligned} P_1'(x, \alpha; d(x)) &= \nabla f(x)d(x) + \alpha \sum_{g_i(x)>0} \nabla g_i(x)d(x) + \alpha \sum_{g_i(x)=0} (\nabla g_i(x)d(x))_+ \\ &\quad + \alpha \sum_{h_i(x)>0} \nabla h_i(x)d(x) + \alpha \sum_{h_i(x)<0} -\nabla h_i(x)d(x) + \alpha \sum_{h_i(x)=0} |\nabla h_i(x)d(x)| \\ &\le \|\nabla f(x)\|_2 \|d(x)\|_2 - \alpha. \end{aligned}$$

Hence $P_1'(x(\alpha_i), \alpha_i; d(x(\alpha_i))) < 0$ for α_i sufficiently large. This contradicts the fact that $x(\alpha_i)$ is a local minimum of $P_1(x, \alpha_i)$. This contradiction establishes the theorem for $P_1(x, \alpha)$ and consequently for all $P(x, \alpha)$ of (1.2) with Q satisfying (1.3).

We establish next the existence of a strict local minimum of the exact penalty function at each strict local minimum of problem (1.1) which satisfies the second order sufficient optimality conditions of Theorem 3.2. In addition we are able under these assumptions to give a lower bound to the penalty parameter α. We begin by establishing a lemma.

Lemma 4.5. *Let the assumptions of Theorem* 3.2 *hold. Then for any fixed* $(u, v) \in \mathbf{R}^{m+k}$ *such that* $u > \bar{u}$ *and* $v > |\bar{v}|$, $\bar{x}$ *is a strict local minimum of the following function*

$$\varphi(x, u, v) := f(x) + \sum_{i=1}^{m} u_i g_i(x)_+ + \sum_{j=1}^{k} v_j |h_j(x)|.$$

Proof. If the lemma were false, then there exists a sequence $\{x^j\}$ converging to $\bar{x}$ such that $x^j \neq \bar{x}$ and

$$\varphi(x^j, u, v) \le \varphi(\bar{x}, u, v).$$

Hence

$$f(x^j) - f(\bar{x}) + \sum_{i=1}^{m} u_i g_i(x^j)_+ + \sum_{i=1}^{k} v_i |h_i(x^j)| \le 0.$$

By passing to a subsequence, if necessary, we have a vector s with $\|s\| = 1$ such that

$$s = \lim_{j\to\infty} \frac{x^j - \bar{x}}{\|x^j - \bar{x}\|}.$$

Therefore

$$\nabla f(\bar{x})s + \sum_{i\in I} u_i(\nabla g_i(\bar{x})s)_+ + \sum_{i=1}^{k} v_i |\nabla h_i(\bar{x})s| \le 0,$$

where $I = \{i \mid g_i(\bar{x}) = 0,\ i = 1, \dots, m\}$. Since $(\bar{x}, \bar{u}, \bar{v})$ satisfy the Karush–Kuhn–Tucker conditions (3.5) we also have that

$$\sum_{i\in I} [u_i(\nabla g_i(\bar{x})s)_+ - \bar{u}_i\nabla g_i(\bar{x})s] + \sum_{i=1}^{k} [v_i|\nabla h_i(\bar{x})s| - \bar{v}_i\nabla h_i(\bar{x})s] \le 0.$$

Because $u > \bar{u}$ and $v > |\bar{v}|$, each term in the above summation is nonnegative and hence zero. Thus it follows that

$$\begin{aligned} \nabla g_i(\bar{x})s &= 0 \quad \text{for } i \in I \text{ and } \bar{u}_i > 0,\\ \nabla g_i(\bar{x})s &\le 0 \quad \text{for } i \in I \text{ and } \bar{u}_i = 0,\\ \nabla h_i(\bar{x})s &= 0 \quad \text{for } i = 1, \dots, k. \end{aligned}$$

By the second order sufficiency condition (3.6) or equivalently (3.9) it follows that $s\nabla_{11}L(\bar{x}, \bar{u}, \bar{v})s > 0$. This implies that for sufficiently large j that

$$L(x^j, \bar{u}, \bar{v}) > L(\bar{x}, \bar{u}, \bar{v}),$$

and consequently

$$\begin{aligned} \varphi(x^j, u, v) &\ge \varphi(x^j, \bar{u}, \bar{v})\\ &\ge L(x^j, \bar{u}, \bar{v})\\ &> L(\bar{x}, \bar{u}, \bar{v})\\ &= f(\bar{x})\\ &= \varphi(\bar{x}, u, v)\\ &\ge \varphi(x^j, u, v) \end{aligned}$$

which is a contradiction. Hence the lemma is true.

To establish a lower bound for the penalty parameter we need the concept of dual norms. Recall that for any given vector norm $\|\cdot\|$ in $\mathbf{R}^l$ there is a corresponding vector norm $\|\cdot\|'$, called the dual norm of $\|\cdot\|$, which is defined by

$$\|x\|' = \sup_{\|y\|=1} yx.$$

Recall also that if $\infty \ge p, q \ge 1$ and $(1/p) + (1/q) = 1$ then for any z in $\mathbf{R}^l$ the p-norm $\|z\|_p := (\sum_{i=1}^{l} |z_i|^p)^{1/p}$ and the q-norm $\|z\|_q$ are dual to each other. For a positive definite and symmetric $l \times l$ matrix A we may define a vector norm $\|z\|_A$ by $\|z\|_A = (zAz)^{1/2}$. The dual norm of $\|\cdot\|_A$ is $\|\cdot\|_{A^{-1}}$. For a detailed discussion on the duality of norms readers are referred to Rockafellar [21, Chapter 15] or Householder [9, pp. 39–45]. We note here that it follows from the definition of dual norms that if two norms $\|\cdot\|$ and $\|\cdot\|'$ are dual to each other then for any x and y we have

$$|xy| \le \|x\|\,\|y\|'.$$

This is known as the generalized Cauchy inequality and will be needed in the proof of the following theorem.

Theorem 4.6 (Second order sufficiency implies strict local minimum of exact penalty). *Let the assumptions of Theorem* 3.2 *hold. Let Q satisfy* (1.3) *and be convex on $\mathbf{R}_+$, let $\|\cdot\|$ be any given vector norm in $\mathbf{R}^{m+k}$ and $P(x,\alpha)$ be its corresponding exact penalty function defined as in* (1.2), *and let $\|\cdot\|'$ be its dual norm. Then for any $\alpha > \bar{\alpha}$ where*

$$\bar{\alpha} = \frac{\|\bar{u}, \bar{v}\|'}{Q'(0+)}$$

the point $\bar{x}$ is a strict local minimum of $P(x, \alpha)$.

Proof. For α satisfying the above inequality we can find $(u, v) \in \mathbf{R}^{m+k}$ such that $u > \bar{u}$, $v > |\bar{v}|$ and

$$\alpha Q'(0+) \geq \|u, v\|' > \|\bar{u}, \bar{v}\|'.$$

By Lemma 4.5 there exists a neighborhood $N(\bar{x})$ of $\bar{x}$ such that for $x \neq \bar{x}$ and $x \in N(\bar{x})$

$$\varphi(\bar{x}, u, v) < \varphi(x, u, v),$$

where φ is defined in the same lemma. Hence by the convexity of Q and by the generalized Cauchy inequality we have that for any $x \in N(\bar{x})$, $\alpha > \bar{\alpha}$ and $x \neq \bar{x}$ that

$$\begin{aligned} P(x, \alpha) &\geq f(x) + \alpha Q'(0+)\|g(x)_+, h(x)\| \\ &\geq f(x) + \|u, v\|'\|g(x)_+, h(x)\| \\ &\geq f(x) + \sum_{j=1}^{m} u_j g_j(x)_+ + \sum_{j=1}^{k} v_j |h_j(x)| \\ &> \varphi(\bar{x}, u, v) \\ &= P(\bar{x}, \alpha). \end{aligned}$$

It is interesting to note that if $Q'(0+) = 1$ then

$$\bar{\alpha}_1 = \|\bar{u}, \bar{v}\|_\infty, \qquad \bar{\alpha}_2 = \|\bar{u}, \bar{v}\|_2, \qquad \bar{\alpha}_\infty = \|\bar{u}, \bar{v}\|_1.$$

We also note that for any real number $p > 1$ and any $l \times l$ positive diagonal matrix D, the quantity $\|Dz\|_p$ is a norm of the vector z in $\mathbf{R}^l$ and its dual norm is $\|D^{-1}z\|_q$, where $(1/p) + (1/q) = 1$. The following corollary then is a consequence of using the norm

$$\left\| D\begin{pmatrix} g(x)_+ \\ h(x) \end{pmatrix} \right\|_p$$

in the last theorem and setting $Q(\zeta) = \zeta$.

Corollary 4.7. *Let the assumptions of Theorem* 3.2 *hold. Let* $Q(\zeta)=\zeta$, *let* $P(x,\alpha)$ *be the exact penalty function defined by* (1.2) *with the norm* $\|Dz\|_p$, *that is*

$$P(x,\alpha)=f(x)+\alpha\left(\left(\sum_{i=1}^{m}(\beta_i g_i(x)_+)^p+\sum_{i=1}^{k}(\gamma_i|h_i(x)|)^p\right)\right)^{1/p}$$

where $(\beta_1,\dots,\beta_m,\gamma_1,\dots,\gamma_k)$ *are the positive elements of the positive diagonal matrix D. Then for any* $\alpha>\bar{\alpha}$ *where*

$$\bar{\alpha}=\left(\sum_{i=1}^{m}\left(\frac{\bar{u}_i}{\beta_i}\right)^q+\sum_{i=1}^{k}\left(\frac{|\bar{v}_i|}{\gamma_i}\right)^q\right)^{1/q},\quad q=1-\frac{1}{p}$$

the point $\bar{x}$ *is a strict local minimum of* $P(x,\alpha)$.

Note that this corollary gives a wider range for β_i and γ_i than that of [2, Theorem 2] in which

$$\alpha=\bar{\alpha}=1,\qquad \beta_i>(m^*)^{1/q}\bar{u}_i,$$

where equalities were not considered and m^* is the number of active inequality constraints at $\bar{x}$.

We establish now the fact the Karush–Kuhn–Tucker conditions (3.5) for problem (1.1) are under suitable conditions satisfied at local minima of $P(x,\alpha)$.

Theorem 4.8. *If there exists an* $\bar{\alpha}\ge 0$ *such that for all* $\alpha\ge\bar{\alpha}$, $P(\bar{x},\alpha)\le P(x,\alpha)$ *for all x in some open neighborhood* $N(\bar{x})$ *which contains some feasible point of* (1.1), *and if f, g and h are differentiable at* $\bar{x}$, *then* $\bar{x}$ *and some* $(\bar{u},\bar{v})\in\mathbf{R}^{m+k}$ *satisfy the Karush–Kuhn–Tucker conditions* (3.5) *for problem* (1.1).

Proof. By Theorem 4.1 $\bar{x}$ is feasible for problem (1.1) and hence $g(\bar{x})\le 0$ and $h(\bar{x})=0$. By Theorem 4.2 $\bar{x}$ is a local minimum point of $P_1(x,\alpha)$ for any $\alpha>\bar{\alpha}$ and consequently $(\bar{x},\bar{y}=0,\ \bar{z}=0)\in\mathbf{R}^{n+m+k}$ constitute a local solution to the following problem for any $\alpha>\bar{\alpha}$:

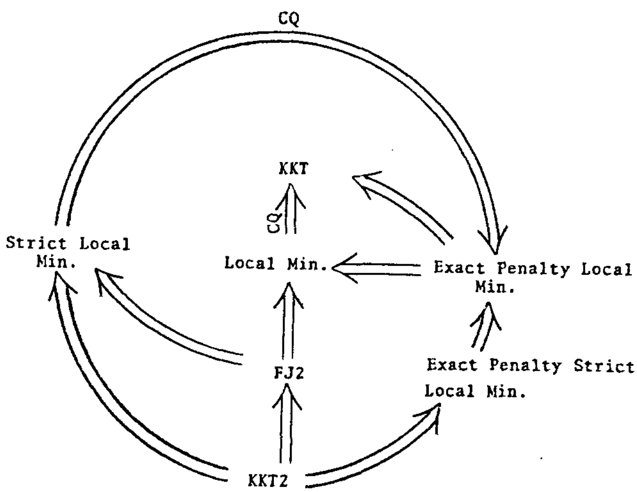
$$\begin{aligned}
\underset{(x,y,z)\mathbf{R}^{n+m+k}}{\text{minimize}}\quad & f(x)+\alpha(ey+lz),\\
\text{subject to}\quad & g(x)-y\le 0,\\
& -y\le 0,\\
& h(x)-z\le 0,\\
& -h(x)-z\le 0,\\
& -z\le 0,
\end{aligned}\tag{4.1}$$

where e and l are vectors of ones in $\mathbf{R}^m$ and $\mathbf{R}^k$ respectively. Note that the Arrow–Hurwicz–Uzawa constraint qualification [13] is satisfied at $\bar{x}$, $\bar{y}=0$, $\bar{z}=0$. (In fact it is satisfied at all feasible points of (4.1) for which g and h are differentiable.) Hence there exist $(\bar{w}, \bar{r}, \bar{s}, \bar{t}, \bar{q}) \in \mathbf{R}^{m+m+k+k+k}$ such that $(\bar{x}, \bar{y}=0, \bar{z}=0, \bar{w}, \bar{r}, \bar{s}, \bar{t}, \bar{q})$ satisfy the Karush–Kuhn–Tucker conditions for problem (4.1) which turn out to be precisely the Karush–Kuhn–Tucker conditions (3.5) for problem (1.1) upon making the identifications $\bar{u} = \bar{w}$ and $\bar{v} = \bar{s} - \bar{t}$.

Using Theorem 4.8 one may interpret the existence of a local minimum to the exact penalty function as a constraint qualification which ensures the satisfaction of the Karush–Kuhn–Tucker conditions at local minima of (1.1).

We sketch in Fig. 1 an outline of the relations obtained in this paper for convenient reference.

Our concluding result generalizes Zangwill's result [23] and is restricted to the convex case. As in Theorem 4.6 an estimate of the size of the penalty parameter α can be obtained in terms of the optimal Lagrange multipliers of the original problem (1.1).



CQ: Constraint qualification (2.1).
Strict Local Min.: Strict local minimum of problem (1.1).
Local Min.: Local minimum of problem (1.1).
Exact Penalty Local Min.: Local minimum of the exact penalty function (1.2).
Exact Penalty Strict Local Min.: Strict local minimum of (1.2).
KKT: First order Karush-Kuhn-Tucker conditions (3.5) for problem (1.1).
FJ2: Second order Fritz John conditions of Theorem 3.1 for problem (1.1).
KKT2: Second order Karush-Kuhn-Tucker conditions of Theorem 3.2 for problem (1.1)

Fig. 1. Summary of results.

Theorem 4.9. *Let $\bar{x}$ be a solution of* (1.1), *f and g be convex on $\mathbf{R}^n$ and h be linear. Let $g(x) < 0$ and $h(x) = 0$ for some x in $\mathbf{R}^n$. For any given vector norm $\|\cdot\|$ in $\mathbf{R}^{m+k}$ let $P(x, \alpha)$ be its corresponding exact penalty function defined as in* (1.2) *with Q satisfying* (1.3) *and being convex on $\mathbf{R}_+$. Then $P(\bar{x}, \alpha) \le P(x, \alpha)$ for all x in $\mathbf{R}^n$ and $\alpha \ge \bar{\alpha}$ where*

$$\bar{\alpha} = \frac{\|\bar{u}, \bar{v}\|'}{Q'(0+)}$$

and $\|\cdot\|'$ is the dual norm of $\|\cdot\|$.

Proof. Because Q is convex on $\mathbf{R}_+$ we have that $Q(t) \ge Q'(0+)t$ for $t \ge 0$. For $\alpha \ge \bar{\alpha}$ and any $x \in \mathbf{R}^n$ we have that

$$\begin{aligned}
P(\bar{x}, \alpha) &= f(\bar{x}) + \alpha Q(\|g(\bar{x})_+, h(\bar{x})\|) \\
&= f(\bar{x}) \\
&= f(\bar{x}) + \bar{u}g(\bar{x}) + \bar{v}h(\bar{x}) \\
&\le f(x) + \bar{u}g(x) + \bar{v}h(x) && \text{(By Theorem 5.4.8 [12])} \\
&\le f(x) + \bar{u}g(x)_+ + \bar{v}h(x) \\
&\le f(x) + \|\bar{u}, \bar{v}\|'\|g(x)_+, h(x)\| \\
&\le f(x) + \frac{\|\bar{u}, \bar{v}\|'}{Q'(0+)} Q(\|g(x)_+, h(x)\|) && \text{(By convexity of } Q) \\
&\le f(x) + \alpha Q(\|g(x)_+, h(x)\|) && \text{(By choice of } \alpha) \\
&= P(x, \alpha).
\end{aligned}$$

We note once more that the penalty function of Corollary 4.7 can be used in the above theorem with the same lower bound $\bar{\alpha}$ on α as that given in that corollary. This again is a sharper bound than that of [2, Theorem 1].

References

[1] D.P. Bertsekas, "Necessary and sufficient conditions for a penalty method to be exact", *Mathematical Programming* 9 (1975) 87–99.

[2] C. Charalambous, "A lower bound for the controlling parameters of the exact penalty functions", *Mathematical Programming* 15 (1978) 278–290.

[3] S. Dolecki and S. Rolewicz, "Exact penalty for local minima", *SIAM Journal on Control and Optimization* 17 (1979).

[4] J.P. Evans, F.J. Gould and J.W. Tolle, "Exact penalty functions in nonlinear programming", *Mathematical Programming* 4 (1973) 72–97.

[5] A.V. Fiacco and G.P. McCormick, *Nonlinear programming: sequential unconstrained minimization techniques* (Wiley, New York, 1968).

[6] U.M. Garcia Palomeres and O.L. Mangasarian, "Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems", *Mathematical Programming* 11 (1976) 1–13.

[7] S.-P. Han, "Superlinearly convergent variable metric algorithms for general nonlinear programming problems", *Mathematical Programming* 11 (1976) 263–282.

[8] S.-P. Han, "A global convergent method for nonlinear programming", *Journal of Optimization Theory and Applications* 22 (1977) 297–309.
[9] A.S. Householder, *The theory of matrices in numerical analyais* (Blaisdell, New York, 1964).
[10] S.P. Howe, "New conditions for exactness of a simple penalty function", *SIAM Journal on Control* 11 (1973) 378–381.
[11] W. Karush, "Minima of functions of several variables with inequalities as side conditions", Master of Science Dissertation, Department of Mathematics, University of Chicago (Chicago, December 1939).
[12] G.P. McCormick, "Second order conditions for constrained minima", *SIAM Journal on Applied Mathematics* 15 (1967) 641–652.
[13] O.L. Mangasarian, *Nonlinear programming* (McGraw-Hill, New York, 1969).
[14] O.L. Mangasarian, "Nonlinear programming theory and computation", in: J.J. Modor and S.E. Elmaghraby, eds., *Handbook of Operations Research* (Van Nostrand Reinhold, New York, 1978), pp. 245–265.
[15] O.L. Mangasarian, "Uniqueness of solution in linear programming", *Linear Algebra and its Applications* 25 (1979) 151–162.
[16] O.L. Mangasarian and S. Fromovitz, "The Fritz John necessary optimality conditions in the presence of equality constraints", *Journal of Mathematical Analysis and Applications* 17 (1967) 34–47.
[17] T. Pietrzykowski, "An exact potential method for constrained maxima", *SIAM Journal on Numerical Analysis* 6 (1969) 299–304.
[18] T. Pietrzykowski, "The potential method for conditioeenal maxima in the locally compact metric spaces", *Numerische Mathematik* 14 (1970) 325–329.
[19] M.J.D. Powell, "The convergence of variable metric methods for nonlinearly constrained optimization calculations" in: O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds., *Nonlinear programming 3* (Academic Press, New York, 1978) pp. 27–64.
[20] S.M. Robinson, "Stability theory for systems of inequalities, part II: Differentiable nonlinear systems", *SIAM Journal on Numerical Analysis* 13 (1976) 497–513.
[21] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, N.J. 1970).
[22] K. Truemper, "Note on finite convergence of exterior penalty functions", *Management Science* 21 (1975) 600–606.
[23] W.I. Zangwill, "Nonlinear programming via penalty functions", *Management Science* 13 (1967) 344–358.