Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, 27514. This work was supported in part by the Office of Naval Research under grant number N000-14-67-A-0321-0003 (NR047-096).

NEW CONDITIONS FOR EXACTNESS OF A SIMPLE PENALTY FUNCTION

by

Stephen Howe

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 778

October, 1971

New Conditions for Exactness of a Simple Penalty Function

by

Stephen Howe

Abstract

We consider penalty function methods for finding the maximum of a function f over the set

$$S_0 = \{x \in R^n: g_i(x) \leq 0 \text{ for } i=1,\ldots,m \text{ and } h_j(x)=0 \text{ for } j=1,\ldots,p\}.$$

New conditions, extending earlier work done by Pietrzykowski, are presented under which the penalty function

$$P(x,\mu) = \mu f(x) - \sum_{i=1}^{m} U(g_i(x)) - \sum_{j=1}^{p} |h_j(x)|$$

is locally exact. The relationships among the new conditions, Pietrzykowski's conditions, and Kuhn-Tucker constraint qualifications are explored.

New Conditions for Exactness of a Simple Penalty Function

Stephen Howe*

1. **Introduction.** In 1969 Pietrzykowski [1], [2] presented conditions under which a local maximum x_0 of the function f on the set

$$S_0 = \{x \in R^n: g_i(x) \le 0,\ i=1,\dots,m \text{ and } h_j(x)=0,\ j=1,\dots,p\}$$

is also an unconstrained isolated (strong) local maximum, for μ sufficiently small, of the penalty function

$$p(x,\mu) = \mu f(x) - \sum_{i=1}^{m} U(g_i(x)) - \sum_{j=1}^{p} |h_j(x)|$$

where $U: R \to R$ is defined by

$$U(\alpha) = \begin{cases} \alpha & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha \le 0 \end{cases}$$

and where $f, g_1, \dots, g_m, h_1, \dots, h_p$ are continuously differentiable real-valued functions on R^n. Specifically, Pietrzykowski's condition is that x_0 be an isolated local maximum of f on S_0 and that the set of "active gradients"

$$\{\nabla g_i(x_0) \text{ each } i \in I, \text{ and } \nabla h_j(x_0) \text{ each } j=1,\dots,p\}$$

be linearly independent, where $I = \{i: g_i(x_0) = 0\}$. It is worth noting that the linear independence of the gradients, as specified, constitutes a constraint qualification which validates the Kuhn-Tucker necessary optimality

---

* *Curriculum in Operations Research and Systems Analysis, University of North Carolina at Chapel Hill, 27514.* *This work was supported in part by the Office of Naval Research under grant number N000-14-67-A-0321-0003 (NR047-096).*

criterion at x_0. This suggests that any constraint qualification which validates the Kuhn-Tucker criterion might also guarantee the above result at an isolated local maximum of f on S_0.

In this note a new condition for the above result is presented. The new condition, unlike that of Pietrzykowski, depends upon the objective function f, and furthermore it neither implies nor is implied by Pietrzykowski's condition. It is also shown that the weak constraint qualification [3] does not guarantee the result at an isolated local maximum of f on S_0, and hence that the Kuhn-Tucker constraint qualifications are not sufficient for this result in general.

2. Result.

Let $f, g_1, \ldots, g_m, h_1, \ldots, h_p, S_0, x_0, I$ and $p(x,\mu)$ be defined as above.

THEOREM. *Suppose the following condition is satisfied.*

G: *For every nonzero* $y \in R^n$ *such that* $\nabla g_i(x_0)^T y \le 0$ *each* $i \in I$ *and* $\nabla h_j(x_0)^T y = 0$ *each* $j = 1,\ldots,p$ *it follows that* $\nabla f(x_0)^T y < 0$.

Then there is a $\mu_0 > 0$ *such that if* $0 < \mu \le \mu_0$ *then* $p(x,\mu)$ *has an unconstrained isolated local maximum at* x_0.

Remark: It can be shown that G is a sufficient, but not necessary, condition that

$$x_0 \text{ be an isolated local maximum of } f \text{ on } S_0.$$

Proof: Since $g_1,\ldots,g_m$ are continuous, there is a $\delta > 0$ such that $g_i(x) < 0$ for every $x \in B(x_0,\delta)$ and $i \notin I$. Here $B(x_0,\delta)$ is the open ball of radius δ about x_0. Now suppose, to get a contradiction, that there is a sequence $\{\mu_n\}$ of positive scalars such that $\mu_n \to 0$ as $n \to \infty$ and where $p(x,\mu_n)$ does not have an unconstrained isolated local maximum at x_0 for each n. Then for each n there is a point x_n such that $0 < ||x_n - x_0|| < \min\{\delta,\mu_n\}$ but for which $p(x_n,\mu_n) \geq p(x_0,\mu_n)$. Denote $y_n = x_n - x_0$ each n, noting that $y_n \to 0$ as $n \to \infty$. Consider the sequence $\{y_n/||y_n||\}$ of vectors in the compact set $\{z \in R^n: ||z||=1\}$. This sequence must have a subsequential limit $y_0 \in R^n$ such that $||y_0|| = 1$. Assume without loss of generality that $y_n/||y_n|| \to y_0$ as $n \to \infty$.

Now suppose that for some $r \in I$, $\nabla g_r(x_0)^T y_0 > 0$. Then for each n $p(x_n,\mu_n) - p(x_0,\mu_n) \leq \mu_n[f(x_n)-f(x_0)] - U(g_r(x_n))$.

Note that

$$\begin{aligned}
\frac{1}{||y_n||} g_r(x_n) &= \frac{1}{||y_n||}[g_r(x_n)-g_r(x_0)] \\
&= \nabla g_r(x_0)^T \frac{y_n}{||y_n||} + \frac{o(||y_n||)}{||y_n||} \\
&\to \nabla g_r(x_0)^T y_0 > 0 \quad \text{as} \quad n \to \infty.
\end{aligned}$$

Clearly for n sufficiently large $g_r(x_n) > 0$ so that $U(g_r(x_n)) = g_r(x_n)$, whence for such n

$$\begin{aligned}
\frac{1}{||y_n||}[p(x_n,\mu_n)-p(x_0,\mu_n)] &\leq \frac{\mu_n}{||y_n||}[f(x_n)-f(x_0)] - \frac{1}{||y_n||} g_r(x_n) \\
&= \mu_n \nabla f(x_0)^T \frac{y_n}{||y_n||} - \frac{1}{||y_n||} g_r(x_n) + \frac{o(||y_n||)}{||y_n||}
\end{aligned}$$

but the latter converges to $-\nabla g_r(x_0)^T y_0 < 0$ as $n \to \infty$, implying that for large n $p(x_n,\mu_n) < p(x_0,\mu_n)$ which is a contradiction. Thus it follows that $\nabla g_i(x_0)^T y_0 \leq 0$ for each $i \in I$. Similarly, it can be shown that $\nabla h_j(x_0)^T y_0 = 0$ for $j = 1,\ldots,p$, and hence from the assumption that G holds it follows that $\nabla f(x_0)^T y_0 < 0$. But then

$$\frac{1}{||y_n||}[f(x_n)-f(x_0)] = \nabla f(x_0)^T \frac{y_n}{||y_n||} + \frac{o(||y_n||)}{||y_n||}$$

$$\to \nabla f(x_0)^T y_0 < 0 \quad \text{as} \quad n \to \infty$$

implying that for n sufficiently large $f(x_n) < f(x_0)$ and hence $p(x_n,\mu_n) < p(x_0,\mu_n)$. This final contradiction establishes the theorem. □

3. Examples.

Example 1. Let $f,g_1,g_2: R \to R$ be defined by $f(x) = x$, $g_1(x) = x$, $g_2(x) = -x$. Both constraints are active at $x_0 = 0$, which is the only feasible point and hence the constrained maximum. Since there is no nonzero $y \in R$ such that $yg_i'(0) \leq 0$ for $i = 1,2$, G holds trivially, but the gradients $g_1'(0) = 1$ and $g_2'(0) = -1$ are not linearly independent. Therefore, the new condition G does not imply Pietrzykowski's condition.

Example 2. Let $f,g: R \to R$ be defined by $f(x) = x^3$, $g(x) = x$. The constraint is active at the constrained maximum $x_0 = 0$. Note that $g'(0) = 1$, and that for $y = -1$ $yg'(0) = -1 \leq 0$ but $yf'(0) = 0$. Therefore, Pietrzykowski's condition does not imply G.

Example 3. Let $f,g_1,g_2: R^2 \to R$ be defined by

$$f(x,y) = y + x^4$$
$$g_1(x,y) = y$$
$$g_2(x,y) = y^3 + x^6$$

(see Figure 1 below). It can be shown that the weak constraint qualification [3] is satisfied at the isolated local maximum $(0,0)$. It is not true, however, that for μ sufficiently small the point $(0,0)$ is a local maximum of

$p((x,y),\mu)$. To see this, define $h: R \to R$ by $h(x) = p((x,0),\mu) = \mu x^4 - x^6$ for any $\mu > 0$, noting that h has a local *minimum* at $x = 0$ for any $\mu > 0$. Hence the condition that x_0 be an isolated local constrained maximum at which the weak constraint qualification holds does not guarantee that the penalty function will be exact.

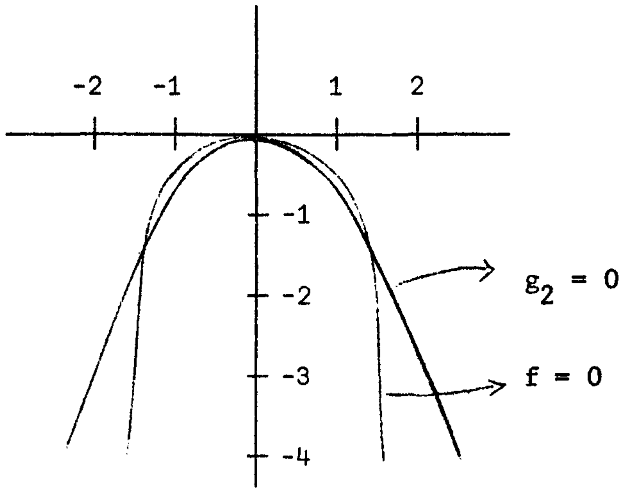

Figure 1

REFERENCES

[1] T. Pietrzykowski, "The potential method for conditional maxima in the locally compact metric spaces," *Numer. Math.* 14, 1970, 325-329.

[2] ______, "An exact potential method for constrained maxima," *SIAM Journal of Numerical Analysis*, 6, No. 2, June 1969.

[3] F.J. Gould and Jon W. Tolle, "A necessary and sufficient constraint qualification for constrained optimization," *SIAM J. Appl. Math.*, 20, No. 2, March 1971.