Discussion Paper No. 22

A CLASS OF EXPONENTIAL PENALTY FUNCTIONS*

by

Frederic H. Murphy

November 7, 1972

Graduate School of Business
Northwestern University
Evanston, Illinois

I wish to thank Professor Robert Mifflin of Yale University for his helpful comments.

*This work was supported in part by National Science Foundation Grant GS-3032.

Abstract

A Class of penalty functions where the trial solutions may be interior or exterior to the feasible region of a nonlinear program are developed. Conditions under which the trial solutions become feasible are presented and a convergence rate is established. Also, trial values for the Lagrange multipliers where convergent subsequences of the trial multipliers converge to optimal Lagrange multipliers can be constructed from the trial solutions to the nonlinear program.

Consider the nonlinear programming problem (NLP)

$$\max_{x \in E^n} \text{imize } f(x) \tag{1}$$

subject to

$$g_i(x) \leq 0 \text{ for } i = 1,\ldots,m, \tag{2}$$

where f and g_i for $i = 1,2,\ldots,m$ are real valued functions defined on E^n.

Penalty function algorithms for solving NLP translate this difficult constrained problem into a sequence of easier unconstrained maximizations, where with each iteration the penalty for infeasibility increases in the exterior algorithms [3], or the penalty for being near the boundary while feasible decreases in the interior algorithms [3]. The penalty functions are constructed so that all convergent subsequences of solutions to the unconstrained problem converge to optimal solutions of NLP, either finitely or in the limit; and the value of the objective function either increases or decreases to the value of an optimal solution depending on the choice of penalty function.

We propose below a class of differentiable penalty functions in which with each iteration the advantage of being interior to the feasible region improves and the penalty for being exterior increases. The trial solutions, unlike with other penalty function algorithms, can be interior or exterior to the feasible region, and the value of the objective function is not necessarily monotonically increasing or decreasing at each iteration.

Separately and independently similar approaches were developed by Evans and Gould [2] and Allran and Johnsen [1]. Evans and Gould have made the most general statement of this new class of algorithms and therefore have the least detailed results.

Allran and Johnsen develop the most restrictive form of the penalty function, and their conditions for convergence must be clarified. The more general results herein reduce to statements of their results under appropriate restrictions.

Our class of functions is

$$F_k(x) = f(x) - \sum_{i=1}^{m} \frac{1}{s(k)} e^{r(k) g_i(x)}, \tag{3}$$

where $r(k) \geq s(k) \geq 1$ and $r(k) \to \infty$.

Members of this class of penalty functions are

$$f(x) - \sum_{i=1}^{m} e^{k g_i(x)} \tag{4}$$

$$f(x) - \sum_{i=1}^{m} \frac{1}{a^k} e^{a^k g_i(x)} \quad \text{where } a > 1 \tag{5}$$

$$f(x) - \sum_{i=1}^{m} \frac{1}{a^k} e^{b^k g_i(x)} \quad \text{where } 1 < a < b. \tag{6}$$

The class of functions treated by Allran and Johnsen [1] is (3) with $s(k) = 1$, and (3) is a specific example of the class developed by Evans and Gould [2].

A disadvantage with interior penalty functions is that it can take as long to find an interior starting solution as it takes to solve the nonlinear program [4;p.213]. Because the trial solution to an exponential penalty function can be feasible or infeasible in NLP, we need not look for an initial feasible solution. Also, by an appropriate choice of parameters, in the convex case we can guarantee that all trial solutions are feasible in NLP.

Exponential penalty functions have the property that they are uniformly bounded over the feasible region, and again by an appropriate choice of parameters the gradient remains uniformly bounded over the feasible region. In the limit as $k \to \infty$, the penalty function (3) has the value $f(x)$ if $x \in T$. If $s(k) \to \infty$, this is true even for the points at the boundary of the feasible region. As with other penalty functions, it can be shown that if $f(\mathbf{x})$ is concave and the constraints are convex, (3) forms a concave function, ensuring that a local maximum is a global maximum. A convergence rate, trial values for the Lagrange multipliers and upper and lower bounds on the value of an optimal solution are provided.

Convergence Results

Let

(7) $$S = \{x \mid g_i(x) \leq 0 \text{ for } i = 1,\ldots,m\},$$

(8) $$T = \{x \mid g_i(x) < 0 \text{ for } i = 1,\ldots,m\},$$

(9) $$x^k \in E^n \text{ maximize } F_k(x) \text{ over } E^n \text{ for } k = 1,2,\ldots,$$

and x^* be an optimal solution to NLP. We use $|y|$ to represent the Euclidean norm of y for y of any dimension, and $\overline{T}$ to represent the closure of the set T.

<u>Theorem 1</u> Assume that

(a) S is a nonempty compact set,

(b) $\overline{T} = S$,

(c) The functions $f(x), g_1(x), \ldots, g_m(x)$ are continuous,

(d) $x^k \in X$, a compact set.

(e) $r(k)$ $s(k)$ 1.

Then any convergent subsequence of x^k converges to an optimal solution of NLP.

Instead of condition d, Evans and Gould [2] provide a growth rate condition on the objective function and constraints to guarantee the existence of x^k and ensure that x^k is in some compact set after a finite number of iterations. The choice of condition d is motivated by the fact that nonlinear programming algorithms are designed to be used on a digital computer, which automatically restricts x^k to a compact set. Condition c along with condition d guarantees the existence of x^k since we are maximizing a continuous function (3) over a compact set. In the convex case, condition d is automatically satisfied.

Note that condition b implies $T \neq \emptyset$. Allran and Johnsen [1] use the weaker assumption that the interior of S is nonempty. This assumption is not sufficient in the non-convex case. To see the need for condition b, consider the example:

$$(10) \qquad \underset{x \in E^1}{\text{maximize}} \ \frac{1}{2} x$$

subject to

$$(11) \qquad g(x) \geq 0 \text{ where}$$

$$(12) \qquad g(x) = \begin{cases} 4|x| - 1 & \text{for } |x| \leq \frac{1}{4} \\ 0 & \text{for } \frac{1}{4} < |x| \leq 1 \\ |x| - 1 & \text{elsewhere} \end{cases}$$

Letting $s(k) = 1$ and $r(k) = k$, we can determine $F_k(x)$:

$$(13) \qquad F_k(x) = \begin{cases} \frac{1}{2} x - e^{k(4|x|-1)} & \text{for } |x| \leq \frac{1}{4} \\ \frac{1}{2} x - 1 & \text{for } \frac{1}{4} < |x| \leq 1 \\ \frac{1}{2} x - e^{k(|x|-1)} & \text{elsewhere} \end{cases}$$

Hence,

$$(14) \qquad F_\infty(x) = \begin{cases} \frac{1}{2} x & \text{for } |x| < \frac{1}{4} \\ \frac{1}{2} x - 1 & \text{for } \frac{1}{4} \leq |x| \leq 1 \\ -\infty & \text{elsewhere.} \end{cases}$$

In this case with $F_\infty(x)$ the discontinuous limit of a sequence of continuous functions, the maximum of $F_\infty(x)$ does not exist. The supremum is $\frac{1}{8}$, the limit of any sequence x_h, $h = 1,2,\ldots$ with

$x_h \uparrow \frac{1}{4}$, whereas the maximum of the original problem (10) and (11) is $\frac{1}{2}$ at $x = 1$. If $s(k) = k$, then

$$
(15) \qquad F_k(x) = \begin{cases} \frac{1}{2}x - \frac{1}{k}e^{k(4|x|-1)} & \text{for } |x| \leq \frac{1}{4} \\ \frac{1}{2}x - \frac{1}{k} & \text{for } \frac{1}{4} < |x| \leq 1 \\ \frac{1}{2}x - \frac{1}{k}e^{k(|x|-1)} & \text{elsewhere} \end{cases}
$$

with the result that

$$
(16) \qquad F_\infty(x) = \begin{cases} \frac{1}{2}x & \text{for } -1 \leq x \leq 1 \\ -\infty & \text{elsewhere.} \end{cases}
$$

From this we may infer that condition b is unnecessary if $s(k) \uparrow \infty$, which will be stated in Theorem 2 below.

<u>Lemma 1</u> Under conditions a, c, d, and e, there exists a convergent subsequence of x^k; and for any such subsequence indexed, say, by k_u, we have $x^{k_u} \to \bar{x} \in S$.

<u>Proof</u> Assume $\bar{x} \notin S$. This means that for some constraint $g_h(x)$ and for some k_u sufficiently large, we have

$$
(17) \qquad g_h(x^{k_u}) \geq \delta > 0.
$$

This implies $F_{k_u}(x^{k_u}) \to -\infty$ as $k_u \to \infty$ since $\frac{1}{s(k_u)} e^{r(k_u)\delta} \to -\infty$ when $r(k) \geq s(k)$ (use L'Hospital's rule on $x^{-1}e^x$ as $x \to \infty$ to see this). However,

$$
(18) \qquad F_{k_u}(x) \geq f(x) - m
$$

for x feasible in NLP since e raised to a negative number is less than one and $s(x) \geq 1$. That is, we have a uniform lower bound on the maximum of $F_{k_u}(x)$ and a contradiction. Therefore, $\overline{x} \in S$.

Proof of Theorem 1

For any $x_o \in T$ we have

$$f(x^{k_u}) \geq F_{k_u}(x^{k_u}) \geq F_{k_u}(x_o). \tag{19}$$

Because $F_{k_u}(x_o) \rightarrow f(x_o)$, and $f(x^{k_u}) \rightarrow f(\overline{x})$ by the continuity of $f(x)$, taking limits in (21) we have

$$f(\overline{x}) \geq f(x_o). \tag{20}$$

Since we can choose x_o arbitrarily close to x^*,

$$f(\overline{x}) \geq f(x^*). \tag{21}$$

By lemma 1 $\overline{x}$ is feasible; therefore, $\overline{x}$ is optimal in NLP.

If we require $s(k) \uparrow \infty$, we may drop assumption b in Theorem 1; that is, we no longer require that $\overline{T} = S$.

Theorem 2 If conditions a, c, d, and e hold, and if $s(k) \uparrow \infty$, then any convergent subsequence of x^k converges to an optimal solution of NLP.

Proof Let $x^{k_u} \rightarrow \overline{x}$, then by Lemma 1 $\overline{x}$ is feasible.

Since $s(k) \rightarrow \infty$, $F_k(x^*) \rightarrow f(x^*)$ as $k \rightarrow \infty$.

Now

$$f(x^{k_u}) \geq F_{k_u}(x^{k_u}) \geq F_{k_u}(x^*). \tag{22}$$

By taking limits as $k_u \rightarrow \infty$ and noting that $f(x)$ is continuous we have

$$f(\overline{x}) \leq f(x^*). \tag{23}$$

That is, (2[illegible]) is an equality and $\bar{x}$ is optimal in NLP.

By noting (19) and (22) we can say

<u>Corollary 1</u> If the conditions of either Theorem 1 or Theorem 2 are satisfied, then

$$F_k(x^k) \to f(x^*) \text{ as } k \to \infty. \tag{24}$$

All exterior algorithms have the disadvantage of producing trial solutions that are infeasible at each iteration and are feasible only in the limit. Algorithms like the cutting-plane methods of Kelley [5] and Veinott [8] have this difficulty as well as does the differentiable exterior penalty function discussed in [9].

Allran and Johnsen [1] show that with their penalty function, after a finite number of iterations there exists an $x \in T$ that is a local maximum of $F_k(x)$. This is not true for the more general function (3). For example, let $f(x) = x$, $g_1(x) = (x-1)^3$ and $g_2(x) = -x$ in NLP, $x \in E^1$. Using (5)

$$\nabla F_k(x) = 1 - e^{a^k(x-1)^3}\, 3(x-1)^2 + \bar{e}^{a^k x}. \tag{25}$$

For k large, $\nabla F_k(x) \approx 1$, or greater than one, for $x \in S = [0,1]$, which means any local maximum of $F_k(x)$ is infeasible. Also, it is not true that any convergent subsequence of local maxima of $F_k(x)$ converge to a feasible point of NLP for all possible choices of the $g_i(x)$, even with the penalty function of Allran and Johnsen [1]. However, under certain conditions all of the trial solutions will be feasible after a finite number of iterations. It is <u>not</u> necessarily true, however, that once a trial point is feasible, all the subsequent x^k are feasible. The trial solutions may be feasible and then infeasible a <u>finite</u> number of times.

<u>Theorem 3</u> If $f(x)$ is concave, $g_1(x),\ldots,g_m(x)$ are convex on E^n, T is nonempty, $\frac{r(k)}{s(k)} \to \infty$, and $\nabla f(x)$, $\nabla g_1(x),\ldots,\nabla g_m(x)$ are continuous, then there exists an integer K' such that every penalty function maximizer is feasible for all $k \geq K'$.

<u>Proof</u> We show that for a fixed point $x_0 \in T$ and any x_B on the boundary of the feasible region, the directional derivative of $F_k(x)$ at x_B in the direction (x_B-x_0) is negative for k sufficiently large. Although the value of k is dependent on the choice of x_B, the continuity of gradients allows us to find a K when the directional derivative is negative for all x_B for $k \geq K$. Hence $F_k(x)$ is decreasing as x is translated from x_B out of the feasible region in the direction (x_B-x_0). Now, for any $x \notin S$, there is an x_B on the line connecting x_0 and x. Because the directional derivative at x_B in the direction (x_B-x_0), which is the same as $(x-x_B)$, is negative, $F_k(x_B) > F_k(x)$ for k sufficiently large by the concavity of $F_k(x)$. Thus, there is a boundary point x_B with $F_k(x_B) > F_k(x)$ corresponding to each infeasible x for k sufficiently large, which means x^k is feasible for k sufficiently large.

Note that

$$\nabla F_k(x) = \nabla f(x) - \sum_{i=1}^{m} e^{r(k)g_i(x)} \frac{r(k)}{s(k)} \nabla g_i(x). \tag{26}$$

Let x_0 be a point in T and x_B be a boundary point of S, that is, for at least one $h \in \{1,\ldots,m\}$, $g_h(x_B) = 0$. Let

$$2\delta \equiv \max\{g_i(x_0): \ i = 1,\ldots,m\}. \tag{27}$$

Since each $g_i(x)$ is convex,

$$(28) \qquad 0 > 2\delta > g_i(x_0) \geq g_i(x) - \nabla g_i(x)(x_0-x).$$

If $g_i(x)$ [illegible], we have

$$(29) \qquad 0 > \delta > \nabla g_i(x)(x_0-x).$$

At iteration k,

$$(30) \qquad \nabla F_k(x_B) = \nabla f(x_B) - \sum_{\{i:\, g_i(x_B) > \delta\}} e^{r(k)g_i(x_B)} \frac{r(k)}{s(k)} \nabla g_i(x_B) - \sum_{\{i:\, g_i(x_B) \leq \delta\}} e^{r(k)g_i(x_B)} \frac{r(k)}{s(k)} \nabla g_i(x_B).$$

$$(31) \qquad \nabla F_k(x_B)\cdot(x_0-x_B) = \nabla f(x_B)\cdot(x_0-x_B) - \sum_{\{i:\, g_i(x_B) > \delta\}} e^{r(k)g_i(x_B)} \frac{r(k)}{s(k)} \nabla g_i(x_B)\cdot(x_0-x_B) - \sum_{\{i:\, g_i(x_B) \leq \delta\}} e^{r(k)g_i(x_B)} \frac{r(k)}{s(k)} \nabla g_i(x_B)\cdot(x_0-x_B).$$

Observe that for h such that $g_h(x) \leq \delta$, $x \in S$

$$(32) \qquad e^{r(k)g_h(x)} \frac{r(k)}{s(k)} |\nabla g_h(x)\cdot(x_0-x)| \leq e^{r(k)\delta} \frac{r(k)}{s(k)} |\nabla g_h(x)\cdot(x_0-x)| \to 0$$

as $k \to \infty$, since $e^{r(k)\delta} r(k) \to 0$ as $k \to \infty$ and $s(k) \geq 1$. Note that (32) is uniform for $x \in S$ with $g_h(x) \leq \delta$ since $\nabla g_h(x)\cdot(x_0-x)$ is uniformly bounded for $x \in S$ by our assumption on the continuity of the gradients of $g_1(x),\ldots,g_m(x)$.

Therefore, there is a K such that, for $k \geq K$ and $x \in S$ with $g_i(x) \leq \delta$

$$(33) \qquad \left| e^{r(k)g_i(x)} \frac{r(k)}{s(k)} \nabla g_i(x)\cdot(x_0-x) \right| < \epsilon.$$

For $x \in S$ with $0 > g_h(x) > \delta$ for some $h \in \{1,\ldots,m\}$ by (29),

$$(34) \qquad \frac{r(k)}{s(k)} e^{r(k) g_h(x)} \nabla g_h(x) \cdot (x_0 - x) \leq \frac{r(k)}{s(k)} e^{r(k) g_h(x)} \delta < 0.$$

For $x \in S$ with $g_h(x) = 0$ for some $h \in \{1,\ldots,m\}$ by (29),

$$(35) \qquad \frac{r(k)}{s(k)} \nabla g_h(x) \cdot (x_0 - x) \leq \frac{r(k)}{s(k)} \delta \downarrow -\infty .$$

Note that (35) like (33) is uniform for $x \in S$.

By our assumptions that $\nabla f(x)$ is continuous and S is compact, we know that there is an $M > 0$ with

$$(36) \qquad |\nabla f(x) \cdot (x_0 - x)| \leq M.$$

Using (29), (33), and (34) for $k \geq K$ with $x = x_B$,

$$(37) \quad \nabla F_k(x_B) \cdot (x_0 - x_B) \geq -M - \frac{r(k)}{s(k)} \sum_{\{i:\ g_i(x_B)=0\}} \nabla g_i(x_B) \cdot (x_0 - x_B) - \sum_{\{i:\ g_i(x_B) \leq \delta\}} \epsilon$$

$$\geq -M - \frac{r(k)}{s(k)} \sum_{\{i:\ g_i(x_B)=0\}} g_i(x_0) - m\epsilon$$

$$\geq -M - \frac{r(k)}{s(k)} \delta - m\epsilon .$$

Consequently, by (35) there is a K' such that $\nabla F_k(x_B) \cdot (x_0 - x_B) > 0$ for all x_B on the boundary of S for $k \geq K'$.

Since the directional derivative of $F_k(x)$ at x_B in the direction $(x_0 - x_B)$ (towards x_0) is positive, it is negative in the direction $(x_B - x_0)$ (from x_B away from the feasible region). Therefore, $F_k(x) < F_k(x_B)$ for all $x = x_B + q(x_0 - x_B)$, $q < 0$, because $F_k(x)$ is concave. We know that for each $x \notin S$, there is a corresponding boundary point x_B such that

(38) $$x = x_B + q(x_0 - x_B) \text{ with } q < 0.$$

Therefore, for each $x \notin S$, there is an associated point $x_B \in S$ where

(39) $$F_k(x) < F_k(x_B) \text{ for } k \geq K';$$

that is, $x^k \in S$ for $k \geq K'$.

We next establish convergence rates for our class of penalty functions $F_k(x)$. Since we have shown that the trial solutions are feasible after a finite number of iterations, we need only establish rates of convergence for feasible trial solutions.

<u>Theorem 4</u> If $x^k \in S$, then

(40) $$\frac{m}{s(k)} \geq f(x^*) - f(x^k) \geq 0$$

<u>Proof</u> Since $x^k \in S$, $f(x^*) - f(x^k) \geq 0$. Also, $g_i(x^*) \leq 0$, that is, $e^{r(k)g_i(x^*)} \leq 1$, and

(41) $$\begin{aligned} f(x^*) - \frac{m}{s(k)} &\leq F_k(x^*) \\ &\leq F_k(x^k) \\ &\leq f(x^k), \end{aligned}$$

or

(42) $$0 \leq f(x^*) - f(x^k) \leq \frac{m}{s(k)}.$$

We see that from the same set of inequalities (41), we can express the convergence rate in terms of the value of the penalty function at the trial solution. That is, from (41),

(43) $$\frac{m}{s(k)} \geq f(x^*) - F_k(x^k) \geq 0.$$

Note that the upper bound of (42) and (43) does not depend on x^k being feasible.

Under the assumption of uniform concavity of $f(x)$, we may establish the rate of convergence of x^k to an optimal solution.

Definition 1 A real valued function $f(x)$ is uniformly concave [6] on a convex set T if there exists a nondecreasing function $\delta(v) > 0$ on $(0,\infty)$ such that for $x, y \in T$

$$f(\tfrac{1}{2}(x+y)) \geq \tfrac{1}{2} f(x) + \tfrac{1}{2} f(y) + \delta(|x-y|), \tag{44}$$

where $|x-y|$ is the Euclidean norm.

An example of a uniformly concave function is any strictly concave function over T with T compact [6]. Since uniformly concave is stronger than strictly concave, x^* is the unique solution to NLP and $x^k \to x^*$.

Theorem 5 If $f(x)$ is uniformly concave on S and $\delta(v)$ is strictly increasing in v, then for $x^k \in S$,

$$|x^k - x^*| \leq \delta^{-1}\left(\frac{m}{2s(k)}\right). \tag{45}$$

Proof Since x^k is feasible, by Theorem 4

$$|f(x^*) - f(x^k)| \leq \frac{m}{s(k)}. \tag{46}$$

And since x^* is optimal and $\frac{1}{2}(x^*+x^k)$ is feasible

$$\begin{aligned} f(x^*) &\geq f(\tfrac{1}{2}(x^*+x^k)) \\ &\geq \tfrac{1}{2} f(x^*) + \tfrac{1}{2} f(x^k) + \delta(|x^*-x^k|) \end{aligned} \tag{47}$$

by uniform concavity. Thus using (47)

$$\tfrac{1}{2} f(x^*) - \tfrac{1}{2} f(x^k) \geq \delta(|x^*-x^k|). \tag{48}$$

Using (46),

(49) $$\frac{1}{2}\frac{m}{s(k)} \geq \delta(|x^*-x^k|).$$

Which means

(50) $$\delta^{-1}\left(\frac{m}{2s(k)}\right) \geq |x^*-x^k|.$$

As an example, if $\delta(r) = r^2$, then

(51) $$\sqrt{\frac{m}{2s(k)}} \geq |x^*-x^k|.$$

As with the penalty functions, we can generate trial Lagrange multipliers where the limit of any convergent subsequence is an optimal set of Lagrange multipliers. Here setting

(52) $$u_i^k = \frac{r(k)}{s(k)}\, e^{r(k)g_i(x^k)} \qquad \text{for } i = 1,\ldots,m$$

we have trial Lagrange multipliers. With $s(k) = 1$, these are the trial multipliers of Allran and Johnsen [1]. The proof that convergent subsequences of u_i^k converge to the Lagrange multipliers of NLP is omitted as it is routine. It can be found in [7].

References

(1) Allran, R.R. and Johnsen, S.E.V., "An Algorithm for Solving Nonlinear Programming Problems Subject to Nonlinear Inequality Constraints," The Computer Journal, Vol. 13 (1970), pp. 171-177.

(2) Evans, V.R. and Gould F.J., "Stability and Exponential Penalty Function Techniques in Nonlinear Programming," Institute of Statistics Memeo Series #723, University of North Carolina, Chapel Hill, 1970.

(3) Fiacco, A. and McCormick, G., Nonlinear Programming, Sequential Unconstrained Minimization Techniques, Wiley, New York, 1968.

(4) Fletcher, R., ed., Optimization, Academic Press, London, 1969.

(5) Kelley, J.E., Jr., "The Cutting-Plane Method for Solving Convex Programs," Journal of the Society for Industrial and Applied Mathematics, Vol. 8 (1960), pp. 703-712.

(6) Levitin, E.S. and Polyak, B.T., "Constrained Minimization Methods," U.S.S.R. Computational Mathematics and Mathematical Physics, Vol. 6 (1966), pp. 1-50.

(7) Murphy, F. H., Topics in Nonlinear Programming: Penalty Function and Column Generation Algorithms, Ph. D. thesis, Yale University, 1971.

(8) Veinott, A.F., Jr., "The Supporting Hyperplane Method for Unimodal Programming, "Operations Research, Vol. 15 (1967), pp. 147-152.

(9) Zangwill, W., "Nonlinear Programming via Penalty Functions," Management Science, Vol. 13 (1967), pp. 344-358.