ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Murphy, Frederic H.

Working Paper

Penalty Function Algorithms with the Potential of Limit Convergence

Discussion Paper, No. 30

Provided in Cooperation with:
Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

Suggested Citation: Murphy, Frederic H. (1973) : Penalty Function Algorithms with the Potential of Limit Convergence, Discussion Paper, No. 30, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:
https://hdl.handle.net/10419/220390

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Discussion Paper No. 30

PENALTY FUNCTION ALGORITHMS WITH THE POTENTIAL OF LIMIT CONVERGENCE

by

Frederic H. Murphy

January 16, 1973

ABSTRACT

Variations on the traditional exterior penalty functions are presented to allow for the possibility of finite convergence.

In discussions about nonlinear programming algorithms, penalty function algorithms are considered inelegant in part because an infinite number of iterations are necessary to find an optimal solution to a nonlinear program. We should look for ways to take advantage of the parallels between Lagrangians and penalty functions to allow for the possibility of finitely convergent penalty functions.*/ Here we are concerned with differentiable exterior penalty function.

Consider the nonlinear program (NLP)

$$(1) \quad \text{maximize } f(x)$$

subject to

$$(2) \quad g_i(x) \leq 0,$$

where we assume for now $f(x)$ is strictly concave and differentiable and $g_i(x)$ is convex and differentiable for $i = 1,\ldots,m$. We also assume Stater's constraint qualification holds, which means there exists an x_o with $g_i(x_o) < 0$; and the feasible region defined by (2) is compact. Let $\pi_1,\ldots,\pi_m$ be optimal Lagrange multipliers for NLP. Then the unique maximum of x^* of

$$(3) \quad f(x) - \sum_{i=1}^{m} \pi_i g_i(x)$$

is an optimal solution to NLP [5].

---

*/ I am indebted to Professor Robert Mifflin of Yale University for suggesting this approach.

Let $A_k(\cdot)$ be a differentiable exterior penalty function at iteration k. That is, $A_k(g_i(x)) = 0$ for $g_i(x) \leq 0$ and $A_k(g_i(x)) \to \infty$ as $k \to \infty$ for $g_i(x) > 0$. The penalty function takes the form

$$(4) \quad f(x) - \sum_{i=1}^{m} A_k(g_i(x))$$

at iteration k. Letting x^k maximize (4) we have $x^k \to x^*$ as $k \to \infty$. If the unconstrained maximum of $f(x)$ is infeasible, $\nabla f(x^*) \neq 0$. Taking the gradient of (4) we have

$$(5) \quad \nabla f(x) - \sum_{\{i \mid g_i(x)>0\}} A_k(g_i(x)) \nabla g_i(x) = \nabla f(x) - \sum_{i=1}^{m} A_k'(g_i(x))\nabla g_i(x)$$

since $A_k'(g_i(x)) = 0$ for $g_i(x) \leq 0$. For x feasible in NLP (5) becomes $\nabla f(x)$ again because $A_k'(g_i(x))$ is zero for all feasible x. Since the gradient of (5) at any feasible x is never zero, x^k is always infeasible, one of the major complaints with exterior penalty functions, and convergence can never take place in a finite number of iterations.

The first step in constructing a penalty function with the potential of convergence within a finite number of iterations is to convert the exterior penalty function into an exponential penalty function [2] and [6], a penalty function where x^k can be either feasible or infeasible. Let $r_k \geq 0$ be a sequence of real numbers where $r_k \to 0$. Then clearly

$$(6) \quad f(x) - \sum_{i=1}^{m} A_k(g_i(x) + r_k)$$

is a convergent penalty function. In other words, letting x_k maximize (6), given the same conditions on $f(x)$ and $g_1(x),\ldots,g_m(x)$ as needed for convergent subsequences of x^k to converge to optimal solutions on NLP [4],

convergent subsequences of x_k converge to optimal solutions of NLP.

We now show that the x_k's can be either feasible or infeasible in NLP. First, if $r_k = 0$, we have our original penalty function (4) again, $x_k = x^k$, and x_k is infeasible at every iteration. We can ensure a subsequence of feasible solutions by starting with an $r_1 > 0$ and setting $r_{k+1} = r_k$ if x_k is infeasible or setting $r_{k+1} = \frac{1}{2} r_k$ if x_k is feasible. The r_k's change infinitely often. To see this, assume the value of r_k is reduced only a finite number of times, and let a be the smallest value of r_k, which means $a > 0$. For k sufficiently large x_k is arbitrarily close to the x that maximizes $f(x)$ subject to the constraints $g_i(x) \leq -a$ for $i = 1,\ldots,m$. By the continuity of the constraints, for k sufficiently large,

$$(7) \quad g_i(x_k) \leq -\frac{a}{2} \quad \text{for } i = 1,\ldots,m$$

and x_k is feasible in (2). This means r_k is reduced infinitely often. The penalty function (6) retains the desired property of exterior penalty functions that after a finite number of iterations $A_k(g_i(x_k) + r_k) = 0$ for $g_i(x)$ a nonbinding constraint at any optimal solution to NLP.

For convenience choose r_k so that

$$(8) \quad A_k'(r_k) = 1.$$

As an example, with $A_k(g_i(x)) = k\{\max[0,g_i(x)]\}^2$, $A_k(r_k) = kr_k^2$, setting $r_k = \frac{1}{2k}$ we have $A_k'(r_k) = 1$.

We now state and prove the theorem that allows the possibility of finite convergence.

Theorem 1 If $f(x)$ is strictly concave and differentiable and $g_1(x),\ldots,g_m(x)$ are convex and differentiable, and if strict complementarity holds, (i.e., $g_i(x^*) = 0$ implies $\pi_i > 0$), then there exists a set of real numbers $p_1,\ldots,p_m$ and a $\delta > 0$ where $p_i \geq \delta$ for $i = 1,\ldots,m$ such that after a finite number of iterations x^* is the unique maximum of

$$(9) \quad f(x) - \sum_{i=1}^{m} p_i A_k(g_i(x) + r_k).$$

Proof: Let

$$(10) \quad \delta = \min \{\pi_i \mid \pi_i > 0 \text{ for } i = 1,\ldots,m\}.$$

Let

$$(11) \quad p_i = \begin{cases} \pi_i \text{ if } \pi_i > 0 \\ \delta \text{ if } \pi_i = 0 \end{cases} \qquad \text{for } i = 1,\ldots,m.$$

The gradient of (9) is

$$(12) \quad \nabla f(x) - \sum_{i=1}^{m} p_i A_k'(g_i(x) + r_k) \nabla g_i(x).$$

Let M index the set of constraints that are binding at x^*. After a finite number of iterations, $A_k(g_i(x^*) + r_k) = 0$ for constraints that are not indexed by M, that is, there exists a K where for $k \geq K$, $A_k(g_i(x^*) + r_k) = 0$ for $i \notin M$. Evaluating (12) at x^* for $k \geq K$ we have

$$(13) \quad \nabla f(x^*) - \sum_{i=1}^{m} p_i A_k'(g_i(x^*) + r_k) \nabla g_i(x^*)$$

$$= \nabla f(x^*) - \sum_{i=1}^{m} p_i A_k'(r_k) \nabla g_i(x^*)$$

$$= \nabla f(x^*) - \sum_{i \in M} p_i \nabla g_i(x^*).$$

The first equality holds because $g_i(x^*) = 0$ for $i \in M$ and $A_k(g_i(x^*) + r_k) = 0$ for $i \notin M$ implies $A_k'(g_i(x^*) + r_k) = 0$ since $A_k(\cdot)$ is a continuously differentiable penalty function. The second equality holds by our choice of r_k. But (13) is just

$$(14) \quad \nabla f(x^*) - \sum_{i=1}^{m} \pi_i \nabla g_i(x^*) = 0$$

because $x^*, \pi_1, \ldots, \pi_m$ form a saddle point of the Lagrangian. By (14) we know x^* maximizes (9).

If we were to set $p_i = \pi_i$ for $i = 1, \ldots, m$, on maximizing (9) we would find x^* in the first iteration. However, we never know the optimal Lagrange multipliers at the first iteration and using a set of trial multipliers with some equal to zero would mean that (9) is no longer a convergent penalty function. In other words if a guess is made of the optimal multipliers at every iteration and the multiplier for some constraint is set to zero infinitely often (i.e., no upper bound is placed on the number of iterations this is done), then the possibility of a convergent subsequence of x_k's with a limit infeasible in NLP exists. Although the assumptions of strict concavity and strict complimentarity are hard to test for, we can construct penalty function algorithms that converge without the necessity of these conditions, yet incorporate a search for the Lagrange multipliers.

The most naive approach is to perform a Lagrange multiplier search as in Everett [3] while ensuring convergence as a penalty function algorithm. Let s_k be a sequence of real numbers where $s_k \to 0$ and $s_k A_k(\cdot)$ is a convergent penalty function. For example, if

$$(15)\quad A_k(g_i(x)) = k\{\max[0,g_i(x)]\}^2,$$

letting $s_k = k^{\frac{1}{2}}$, we have

$$(16)\quad s_k A_k = k^{\frac{1}{2}}\{\max[0,g_i(x)]\}^2$$

which is still a convergent penalty function.

Taking the Everett search technique presented in Fiacco and McCormick [4] we can construct the following algorithm.

Let $\delta_1^1 > 0,\ldots,\delta_m^1 > 0$ and $u_1^1 \geq s_1,\ldots,u_m^1 \geq s_1$ be fixed positive real numbers. Letting $p_1^k,\ldots,p_m^k$ be our trial values for Lagrange multipliers at iteration k, we set $p_1^1 = u_1^1,\ldots,p_m^1 = u_m^1$. We adjust our trial values for the Lagrange multipliers in the following manner. At iteration k, if

$$(17)\quad g_i(x_k) > 0 \geq g_i(x_{k-1}),\ \delta_i^{k+1} = \frac{\delta_i^k}{4},\ u_i^{k+1} = (1+\delta_i^{k+1})p_i^k$$

$$(18)\quad g_i(x_{k-1}) > g_i(x_k) > 0,\ \delta_i^{k+1} = 1.3\,\delta_i^k,\ u_i^k = (1+\delta_i^{k+1})p_i^k$$

$$(19)\quad g_i(x_k) \geq g_i(x_{k-1}) > 0,\ \delta_i^{k+1} = 2\delta_i^k,\ u_i^k = (1+\delta_i^{k+1})p_i^k$$

$$(20)\quad 0 > g_i(x_k) > g_i(x_{k-1}),\ \delta_i^{k+1} = 1.3\,\delta_i^k,\ u_i^{k+1} = (1-\delta_i^{k+1})p_i^k$$

$$(21)\quad 0 > g_i(x_{k-1}) > g_i(x_k),\ \delta_i^{k+1} = 2\,\delta_i^k,\ u_i^{k+1} = (1-\delta_i^{k+1})p_i^k$$

$$(22)\quad g_i(x_{k-1}) \geq 0 \geq g_i(x),\ \delta_i^{k+1} = \frac{\delta_i^k}{4},\ u_i^{k+1} = (1-\delta_i^{k+1})p_i^k$$

$$(23)\quad g_i(x_k) = 0,\qquad \delta_i^{k+1} = \frac{\delta_i^k}{4},\ u_i^{k+1} = p_i^k\ .$$

Let

$$(24)\quad p_i^{k+1} = \begin{cases} u_i^{k+1} & \text{if } u_i^{k+1} \geq s_k \\ s_k & \text{otherwise} \end{cases}$$

We then maximize

$$(25)\quad f(x) = \sum_{i=1}^{m} p_i^{k+1} A_{k+1}(g_i(x) + r_{k+1}).$$

Another approach is to use the trial Lagrange multipliers as they naturally appear in penalty function algorithms. Let $p_1,\ldots,p_m$ be real numbers greater than zero; let x_k maximize

$$(26)\quad f(x) - \sum_{i=1}^{m} p_i A_k(g_i(x) + r_k).$$

Then

$$(27)\quad \nabla f(x_k) - \sum_{i=1}^{m} p_i A_k' (g_i(x_k) + r_k) \nabla g_i(x_k) = 0,$$

and $p_i A_k'(g_i(x_k) + r_k)$ forms a trial Lagrange multiplier for constraint i at iteration k. We can now construct our algorithm. Let $p_1',\ldots,p_m'$ be fixed real numbers greater than zero. At iteration k we maximize

$$(28)\quad f(x) - \sum_{i=1}^{m} p_i^k A_k(g_i(x_k) + r_k).$$

With x_k the solution to (27), for $i = 1,\ldots,m$ set

$$(29)\quad p_i^{k+1} = \begin{cases} p_i^k A_k'(g_i(x)) & \text{if } p_i^k A_k'(g_i(x_k)) \geq s_k \\ s_k & \text{otherwise.} \end{cases}$$

Note that the above algorithms converge under the normal conditions for convergence of exterior penalty function algorithms [4] while having

the added feature of the potential finite convergence. We now present an algorithm where concavity of the objective function and convexity of the constraints is required. Also we assume there is an x_o with $g_i(x_o) < 0$ for $i = 1,\ldots,m$. The algorithm is a variant of the Generalized Programming algorithm of Dantzig and Wolfe [1].

At iteration k we have a linear program (RM) known as the restricted master,

$$(30) \quad \text{maximize} \quad f(x_o)w_o + \ldots + f(x_{\ell(k)})w_{\ell(k)}$$

subject to

$$(31) \quad g_i(x_o)w_o + \ldots + g_i(x_{\ell(k)})w_{\ell(k)} \leq 0 \quad \text{for } i = 1,\ldots,m$$

$$(32) \quad w_o + \ldots + w_{\ell(k)} = 1$$

$$(33) \quad w_j \geq 0 \quad \text{for } j = 0,\ldots,\ell(k),$$

where $\ell(k) + 1$ is the number of columns at iteration k.

Let $W_o^k,\ldots,W_{\ell(k)}^k$ be an optimal solution to RM, and let

$$(34) \quad x^k = x_o W_o^k + \ldots + x_{\ell(k)} W_{\ell(k)}^k.$$

Let

$$(35) \quad I_{k+1} = \{i \mid g_i(x_o)W_o^k + \ldots + g_i(x_{\ell(k)})\, W_{\ell(k)}^k > -\epsilon\},$$

where $\epsilon > 0$ is fixed for all iterations. For $i \in I_{k+1}$ set $p_i^{k+1} = 0$, for $i \in I_{k+1}$ set

$$(36)\qquad p_i^{k+1} = \begin{cases} u_i^k & \text{if } u_i^k > s_k \\ s_k & \text{otherwise,} \end{cases}$$

where u_i^k is the optimal solution to the dual of RM corresponding to $W_1^k,\ldots,W_{\ell(k)}^k$. We then,

$$(37)\qquad \text{maximize } f(x) - \sum_{i\in I_{k+1}} p_i^{k+1} A_{k+1}(g_i(x))$$

for $x \in X$, a compact set containing the feasible region. In RM we consider basic solutions [8] only. We drop all nonbasic columns, except the column associated with x_o. We then add the column

$$(38)\qquad \begin{bmatrix} f(x_{\ell(k+1)}) \\ g_1(x_{\ell(k+1)}) \\ \vdots \\ g_m(x_{\ell(k+1)}) \\ 1 \end{bmatrix}$$

to form a new RM, with at most $m+3$ nonslack columns, and continue. Convergence of this algorithm is proved in a more general context in [7].

It has been known for a while that trial values for the Lagrange multipliers can be generated from the penalty function at each iteration. Yet this information was never used to aid in determining the penalty function for the next iteration. With the three algorithms above, we are in a position to take advantage of this information in the next iteration. At the moment, there are no computational results to compare the effectiveness of these algorithms and the original exterior penalty functions. This work is in the planning stages.

REFERENCES

[1] Dantzig, G., Linear Programming and Extensions, Princeton University Press, Princeton, New Jersey, 1963.

[2] Evans, J. P., and F. J. Gould, "Stability and Exponential Penalty Function Techniques in Nonlinear Programming," Institute of Statistics Mimeo Series No. 723, University of North Carolina, Chapel Hill, 1970.

[3] Everett, H., "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," Operations Research, 11, 399-47 (1963).

[4] Fiacco, A. V. and G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, New York, 1968.

[5] Mangasarian, O. L., Nonlinear Programming, McGraw-Hill, New York, 1969.

[6] Murphy, F. H., "A Class of Exponential Penalty Functions," Discussion Paper No. 22, The Center for Mathematical Studies in Economics and Management Science, Northwestern University, 1972.

[7] Murphy, F. H., "A Column Generation Algorithm for Nonlinear Programming," Discussion Paper No. 20, The Center for Mathematical Studies in Economics and Management Science, Northwestern University, 1972.

[8] Wagner, H. M., Principles of Operations Research with Applications to Managerial Decisions, Prentice-Hall, 1969.

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