TECHNICAL REPORT NO. CS 74024-R

SEARCHING ONE MULTIPLIER IN GLM

by

Harvey J. Greenberg

Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

December 1974

ABSTRACT

A unified approach is developed for one-dimensional GLM. The major result is a convergence theorem for interval reduction. Comparative analysis of bisection, linear interpolation and tangential approximation reveals the relative advantages of tangential approximation.

Computing Reviews Number 5.41

Keywords: Optimization, Lagrange Multipliers, Mathematical Programming, Resource Allocation.

INTRODUCTION

The purpose of this paper is to present a unified theorem for interval reduction applicable to searching one multiplier in GLM. After establishing the main result, attention is focused on bisection, linear interpolation and tangential approximation. The relative merit of tangential approximation is presented.

We are given a family of mathematical programs of the following form:

$$\text{Max } f(x): \quad g(x) \leq b \text{ and } x \text{ in } S,$$

where S is an n-dimensional, compact set and f and g are real-valued, continuous, scalar functions on S. Each member of the family is distinguished by the "resource level," b, taken from the set of scalars that renders the associated program feasible, which we denote 'B'. It should be noted that this one-constraint problem may in fact be a subproblem embedded in a relaxation scheme for a more general case. For example, a multiplier vector may be searched cyclically, or we may be solving a surrogate program.

The Generalized Lagrange Multiplier Method (GLM) considers the Lagrangian dual,

$$\text{Min } L^*(y) + yb: \; y \geq 0,$$

where

$$L^*(y) = \text{Max } \{f(x) - yg(x): x \text{ in } S\}$$

To avoid lengthy review of GLM, we shall assume knowledge of the underlying concepts. (See, for example, references 2,3,4 and 7 for an introduction and references 1, 5, and 6 for further results.) Terms such as "gap" and "PR Space" are defined in Everett's (3) pioneering paper.

Fox and Landi (4) describe a parallel method (sweeping multiplier space). This analysis complements theirs since their use of bisection is not critical to their main idea. We shall show that the linear programming method (2) reduces to tangential approximation for the one dimensional case (6). Quadratic approximation of the Lagrangian dual leads to linear interpolation, and one may consider variations of this scheme. We shall further consider each of these methods after establishing our main result in the next section.

We define

$$f^*(b) = \text{Max}\,\{f(x) : g(x) \leq b \text{ and } x \text{ in } S\},$$

and we note the weak duality relation (1,6):

$$f^*(b) \leq L^*(y) + yb \quad \text{for all } y \geq 0 \text{ and } b \text{ in } B.$$

INTERVAL REDUCTION

Let us suppose, at a general iteration, we have two triples, namely

(Y0, B0, F0) and (Y1, B1, F1),

where

(1) $B0 < \bar{b} < B1$

(2) $F0 = f^*(B0)$ and $F1 = f^*(B1)$

(3) $L^*(Y0) = F0 - B0*Y0$ and $L^*(Y1) = F1 - B1*Y1$.

Collectively, these conditions mean that we have computed L* for y= Y0 and Y1 and obtained a Lagrange-maximal policy yielding the PR points (B0, F0) and (B1, F1), respectively. The value of $\bar{b}$ distinguishes the problem we wish to solve.

If we compute L*(y) and obtain a PR point, (b, f*(b)), then Everett's Monotonicity Theorem (3) implies:

if $b > \bar{b}$, then $y^* < y$ and if $b < \bar{b}$, then $y^* > y$,

where y* is a minimal multiplier. (Of course, if $b = \bar{b}$, then we terminate with no gap.)

We shall suppose we start with

$Y0 > y^* > Y1$.

If we select any multiplier in the open interval (YI, YO), then the generated PR point (b, f*(b)) satisfies

$$BO < b < B1.$$

This means we can reduce the interval of search on y at each iteration, and we are led to the following

INTERVAL REDUCTION ALGORITHM MODEL

Given (YO, BO, FO) and (Y1, B1, F1) satisfying (1)-(3) above and D=YO-Y1, an update is made such that D is decreased until $D \leq \varepsilon$.

I1. (select new multiplier)

Set y= M(YO, BO, FO, Y1, B1, F1)

I2. (maximize Lagrangian)

Compute (b,z) = (g(x), f(x)): L*(y) = f(x) - yg(x).

I3. (update state)

If $b > \bar{b}$, set (Y1, B1, F1) = (y, b, z) and go to I4; else,
if $b < \bar{b}$, set (YO, BO, FO) = (y, b, z) and go to I4; else,
set (YO, BO, FO) = (Y1, B1, F1) = (y, b, z) and STOP.

I4. (test length)

set D = YO - Y1. If $D \leq \varepsilon$, STOP; else, go to I1.

An Interval Reduction Algorithm (IRA) consists of specifying the multiplier selection rule, M. Since M is a functional, continuity, and other properties, may be described for it. We now present the main result.

IRA Convergence Theorem:

The Interval Reduction Algorithm will converge for any continuous multiplier selection rule which maintains

$$YO > M(YO, BO, FO, Y1, B1, F1) > Y1.$$

PROOF:

The state sequence $(YO^k, BO^k, FO^k, Y1^k, B1^k, F1^k)$ is monotone, satisfying $(YO^k, B1^k, F1^k) > (y^*, b, f^*(b)) > (Y1^k, BO^k, FO^k)$.

Therefore, the

K be a subsequence for which the update (step I3) satisfies

$Y0^{k+1} = M(Y0^k, B0^k, F0^k, Y1^k, B1^k, F1^k)$ for k in K;

and let K^1 be the complementary sequence where

$Y1^{k+1} = M(Y0^k, B0^k, F0^k, Y1^k, B1^k, F1^k)$ for k in K^1.

Since M is continuous, it follows that $Y0^{\infty} = M(Y0^{\infty}, B0^{\infty}, F0^{\infty}, Y1^{\infty}, B1^{\infty}, F1^{\infty}) = Y1^{\infty}$, so the common limit is y^*. This completes the proof.

A word of caution is in order. Note that we need not have $B0^{\infty} = B1^{\infty}$. In fact, if $\bar{b}$ is in a duality gap, then, we know $B1^{\infty} > B0^{\infty}$. Actually, we can have $B1^{\infty} > B0^{\infty}$ even when $\bar{b}$ is not in a duality gap. This is the case if f* is concave, affine over (b0, b1), where b0 < b1, and f* is differentiable at b0 and at b1. In this case the multiplier converges to its correct value, namely

$$y^* = \frac{f^*(b1) - f^*(b0)}{b1 - b0}.$$

However, a rule which would not finitely terminate for $\varepsilon = 0$ results in $B0^{\infty} = b0$ and $B1^{\infty} = b1$. To see this, note that setting $y = y^*$ would generate all PR points of the form $(b, f^*(b))$, where $b0 \leq b \leq b1$. Therefore, $y^k \; y^*$ would generate $b^k \geq b1$ and $y^k > y^*$ would generate $b^k \leq b0$. Since f^* is differentiable, b0 and b1 have one multiplier apiece to generate them. Therefore, if y^* generates b0 and b1, it must be the case that $b^k > b1$ for $y^k < y^*$ and $b^k < b0$ for $y^k > y^*$.

BISECTION, INTERPOLATION AND TANGENTIAL APPROXIMATION

Let us now consider three particular multiplier selection rules:

BISECTION: $M = (Y0 + Y1)/2$

LINEAR INTERPOLATION: $M = aY0 + (1-a)\,Y1$, where $aB0 + (1-a)\,B1 = b$ (note: $0<a<1$)

TANGENTIAL APPROXIMATION: $M = (F1 - F0)/(B1 - B0)$

BISECTION conservatively produces a constant rate of interval reduction, namely $D^{k+1} = D^k/2$, no matter where y* is located. The only state values used are the two multipliers, Y0 and Y1.

LINEAR INTERPOLATION is motivated by supposing that L* is quadratic over the interval. Its rate structure is then linear affine, so we pretend

$$b(y) \simeq p\,y + q,$$

where b(y) is the resource level generated in step I2. Estimates of p and q may be deduced from knowing the associate subgradients, B0 and B1. This produces the specified value of a.

TANGENTIAL APPROXIMATION exploits generated functional information in PR space. The selection rule is an estimate of the slope, y*, at $\bar{b}$ by the ratio of change in payoff to change in resource level. It is important to note that the PR values, (B0, F0) and (B1, F1), are constructed to lie on the same Lagrangian contour for the selected multiplier - i.e., F0 - yB0 = F1 - yB1. Therefore, either L*(y) > F0 - yB0, in which case the interval, (B0, B1), is also reduced, or else y = y* and we stop with the gap interval, (B0, B1).

It is clear that bisection and linear interpolation need not converge finitely for $\varepsilon = o$. (See example below.) However, tangential approximation must converge finitely if the decision space is finite. This follows from the fact that continuation occurs only when the interval (B0, B1) is reduced. If the PR space is finite, we must eventually arrive at the gap region (b0, b1) or generate $\bar{b}$ (if there is no gap). Further, let us show a case where the decision space is not finite, yet tangential approximation must still converge finitely. Specifically, consider $\bar{b}$ in the gap region (b0, b1), and suppose f^* is not differentiable at the end points, b0 and b1. Then, since we have dual converence, we must eventually generate b0 (with Y0 sufficiently close to y^*) and b1 (with Y1 sufficiently close to y^*). Once the end points are generated, tangential approximation chooses y^* at the next iteration. Once this happens, b0 (or b1) is re-generated, and the process terminates. A case where f^* is not differentiable at the end points of a gap region is where there are adjacent gap regions. In this sense, tangential approximation is "typically" finitely convergent for $\varepsilon=o$, while the other two methods are not.

Let us now show that using linear programming to search Lagrange multipliers (2,6) reduces to tangential approximation in the case of one constraint.

Let us suppose the PR points, $\{(b^t, f^t)\}_{t=1}^{k}$, have been generated. Re-index, if necessary, so that $(b^t, f^t) < (b^{t+1}, f^{t+1})$. Note that $b^1 < \bar{b} < b^k$. The linear programming method chooses y^{k+1} to be a dual solution of:

$$\text{Maximize } \sum_{t=1}^{k} w_t f^t :$$

$$(1) \quad w \geqq 0 \text{ and } \sum_{t=1}^{k} w_t = 1$$

$$(2) \quad \sum_{t=1}^{k} w_t b^t \leqq b.$$

The dual is:

$$\text{Minimize } z + yb:$$

$$(1) \quad y \geqq 0$$

$$(2) \quad z + yb^t \geqq f^t \text{ for } t = 1, \ldots, k.$$

Define r such that

$$b^r < \bar{b} < b^{r+1}.$$

Then, we shall show the dual solution is

$$y = \frac{f^{r+1} - f^r}{b^{r+1} - b^r},$$

which is the tangential approximation rule.

Choosing

$$z = \text{Max } \{f^t - yb^t\}_1^k$$

makes (y,z) feasible in the dual.

Define

$$w_r = \frac{b^{r+1} - \bar{b}}{b^{r+1} - b^r}$$

$$w_{r+1} = \frac{\bar{b} - b^r}{b^{r+1} - b^r}$$

$$w_t = 0 \quad \text{for } t \neq r, r+1.$$

Clearly, w is feasible in the primal. Now we shall show that the primal and dual objective values are equal.

We have

$$\sum_{t=1}^{k} w_t f^t = w_r f^r + w_{r+1} f^{r+1}$$

and

$$z + y\bar{b} = z + \frac{f^{r+1} - f^r}{b^{r+1} - b^r} \bar{b}.$$

Therefore, our task is to show

$$(b^{r+1} - b^r)\, z = (b^{r+1} - \bar{b}) f^r + (\bar{b} - b^r) f^{r+1} - (f^{r+1} - f^r)\bar{b}$$

Equivalently, we must show

$$z\,(b^{r+1} - b^r) = f^r b^{r+1} - f^{r+1} b^r.$$

Suppose

$$z = f^p - yb^p.$$

Then, we must show

$$(b^{r+1} - b^r)\, f^p - (f^{r+1} - f^r) b^p = f^r b^{r+1} - f^{r+1} b^r.$$

It suffices to show p = r or r + 1, since then the above equation holds. (Note that z represents the Lagrangian contour value.)

Since (b^r, f^r) and (b^{r+1}, f^{r+1}) were generated PR points, there are multiplier values, say y^r and y^{r+1}, such that

$$f^r - y^r b^r \geqq f^t - y^r b^t \text{ and } f^{r+1} - y^{r+1} b^r \geqq f^t - y^{r+1} b^t$$

for all $t=1,\ldots,k$. Therefore,

$$f^t - yb^t \leqq f^r - yb^r - (y^r - y)(b^r - b^t).$$

Since $y > y^r$, and since $b^t < b^r$ for $t < r$, it follows that

$$f^t - yb^t < f^r - yb^r \text{ for } t < r.$$

Similarly, for $t > r+1$, we have

$$f^t - yb^t \leqq f^{r+1} - yb^{r+1} - (y^{r+1} - y)(b^{r+1} - b^t)$$
$$< f^{r+1} - yb^{r+1}.$$

Hence, $z = f^r - yb^r = f^{r+1} - yb^{r+1}$, which completes the proof.

Now let us compare these three methods, each of which satisfies the conditions in the IRA Convergence Theorem. First, let us analyze the following

<u>example</u>: Maximize $2x_1 + x_2^2 + x_3^3$:

(1) $x_1 + x_2 + x_3 \leqq 4$

(2) $x_1 = 0, 1$

$x_2 = 0, 1, 2$

$x_3 = 0, 1, 2, 3$

Tables I, II, and III below list the state sequences generated for each of the three methods.

TABLE I: State Sequence for BISECTION

Multipliers Y1	Multipliers Y0	Resources B1	Resources B0	Payoffs F1	Payoffs F0	GLM Bound
0	10	6	0	33	0	33
0	5	6	3	33	27	32
0	2.5	6	3	33	27	29.5
1.25	2.5	6	3	33	27	29.5
1.875	2.5	6	3	33	27	29.5
⋮		⋮		⋮		⋮

TABLE II: State Sequence for LINEAR INTERPOLATION

Multipliers Y1	Multipliers Y0	Resources B1	Resources B0	Payoffs F1	Payoffs F0	GLM Bound
0	10	6	0	33	0	33
0	6.667	6	3	33	27	33
0	4.444	6	3	33	27	33
0	2.96	6	3	33	27	32.92
1.9734	2.96	6	3	33	27	29.0532
⋮		⋮		⋮		⋮

TABLE III: State Sequence for TANGENTIAL APPROXIMATION

Multipliers Y1	Multipliers Y0	Resources B1	Resources B0	Payoffs F1	Payoffs F0	GLM Bound
0	10	6	0	33	0	33
0	5.5	6	3	33	27	32.5
2	2	5,6	3,4	27,28	31,33	29

We note that Tangential Approximation converges finitely for $\varepsilon = 0$, while the other two do not. This affects the GLM bound $(L^*(y) + yb)$ since it is strongest at the minimal multiplier. (It happens, in this example, that there is no gap, and an optimal policy is $x^* = (1,0,3)$, corresponding to the PR point, (4,33).)

To further analyze the relative behavior of these three methods let us simplify the program to where

$$g(x) = \sum x_j$$

and $S = \{x: 0 \leq x \leq \bar{b},\ x \text{ integer}\}$. Then, we start with

$$Y0 = YMAX,\ B0 = 0,\ F0 = 0$$

$$Y1 = 0,\ B1 = n\bar{b},\ F1 = L^*(0) = FMAX\ ,$$

where YMAX is sufficiently large to produce $x^* = 0$ as the Lagrangian maximum.

At the first iteration we have

BISECTION: $y_B = YMAX/2$

LINEAR INTERPOLATION: $y_L = YMAX\ (1 - 1/n)$

TANGENTIAL APPROXIMATION: $y_T = FMAX/n\bar{b}$.

Note these satisfy

$$y_T < y_B < y_L\ ,$$

and Figure 1 illustrates these choices.

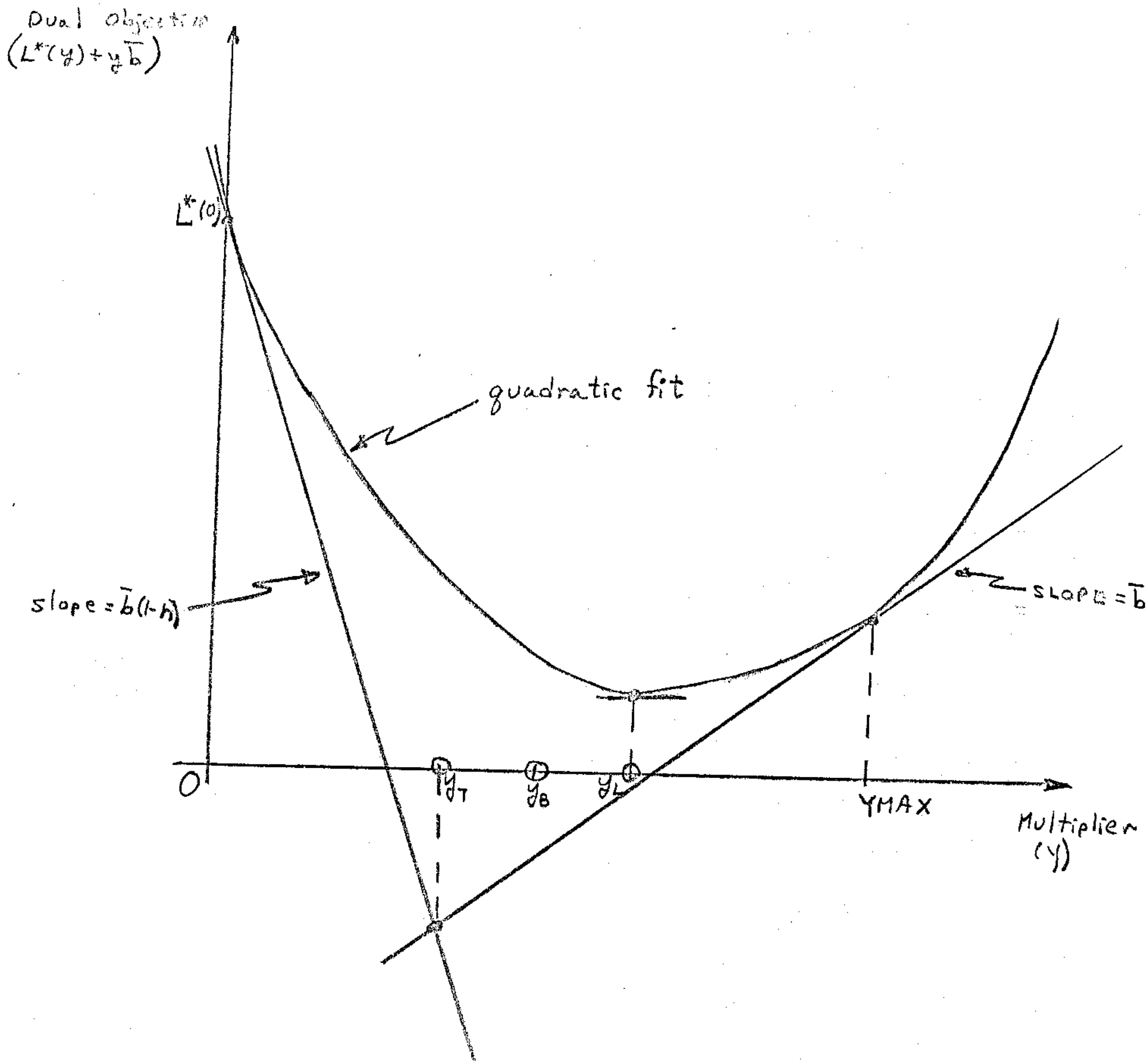

FIGURE 1: GRAPHICAL SKETCH OF FIRST ITERATION

It seems clear that linear interpolation is slow, at least in the beginning. Moreover, finite problems produce a piece-wise linear Lagrangian, so quadratic approximation does seem inappropriate.

Normally, tangentential approximation captures the structure to produce rapid convergence, which is finite for finite decision spaces, and it is typically finite for nonfinite decision spaces when gaps are present. (To see this last point refer to the previous discussion when f* is not differentiable at the end points of the gap region.) However, if f* has a "long tail," then bisection can be helpful, particularly to overcome computed errors due to small multiplier values, i.e., small changes in payoff to large changes in resource values.

To further justify the relative merit of tangential approximation, and to indicate speed of convergence, a collection of problems were run (on an IBM 370/158), using the number of variables as a parameter, and problems of the following form:

$$\text{Maximize} \quad \sum_{j=1}^{n} f_j(x_j) :$$

$$\sum_{j=1}^{n} x_j \leq 10 * n$$

$$x_j \in \{0, 1, \ldots, 20\} \quad \text{for } j=1, \ldots, n,$$

where $f_j(0) = 0$ and $f_j(x+1) > f_j(x)$ for all j, x.

Running 20 monte carlos per case, the objective function was generated as follows. Given $f_j(x)$, (starting with $f_j(0) = 0$), $f_j(x+1)$ is computed by

$$f_j(x+1) = f_j(x) + 100{*}R,$$

where R is a uniform random number (between 0 and 1).

Each of the three multiplier rules were run simultaneously until tangential approximation terminated with the final state values. The average and maximum number of iterations for the 20 monte carlo runs are given in the table below for the four cases, n=100,200,300,400. To measure relative effectiveness of tangential approximation two averages are tabulated, namely

(1) gap size, B1-B0, where

$$B1 = \text{Min}\left\{b^t:\ b^t \geq \bar{b}\right\}$$
$$B0 = \text{Max}\left\{b^t:\ b^t \leq \bar{b}\right\}$$

where t is iteration number, and b^t is the associated generated resource level;

(2) relative multiplier error, (Y0-Y1)/y*, where Y0 corresponds to B0, Y1 corresponds to B1 and y* is the minimal multiplier (found by tangential approximation).

It should be noted that tangential approximation computes the minimal multiplier, y*, and the least gap range, B1-B0 (as well as the strongest GLM bound).

We see that for linear interpolation and bisection the multiplier value tends to be within 10% of the optimum (within 1% for larger problems). Therefore, all three procedures appear nearly equal in effectiveness if only the multiplier value is needed. However, the gap size is markedly greater for linear interpolation and bisection compared to the actual gap size found by tangential approximation! This would affect the ability to obtain a "good" feasible solution, as for example by starting at the feasible policy associated with B0 and heuristically increasing some of the activity levels to increase the objective value towards the GLM bound. (See (5) or (6) for more on gap closing.)

In summary, a general interval reduction algorithm has been presented with a convergence proof that only assumes continuity and a bound on the nonexpansive range of the selection rule. Three procedures which satisfy these assumptions are bisection, linear interpolation and tangential approximation. For finite decision spaces tangential approximation converges finitely, while the other

RESULTS OF 20 MONTE CARLO RUNS (IBM 370/158)

NUMBER OF VARIABLES	AVERAGE NUMBER OF ITERATIONS	MAXIMUM NUMBER OF ITERATIONS	AVERAGE GAP SIZE (B1-B0)			AVERAGE MULTIPLIER ERROR |Y1-Y0| /y*	
			TANGENTIAL APPROXIMATION	BISECTION	LINEAR INTERPOLATION	BISECTION	LINEAR INTERPOLATION
100	9.1	11	10.40	367.05	248.20	.097	.072
200	10.0	11	13.60	359.70	153.20	.018	.043
300	9.5	10	15.20	320.50	137.30	.009	.023
400	9.0	9	20.40	284.70	127.20	.003	.011

two procedures do not (in general). Further, an empirical study revealed that, although the multiplier value may be sufficiently close to its optimal value, the generated resource levels do not converge as quickly to bounding points (for example, end points of the gap region). This results in a less effective policy from which branch and search may be used to resolve the gap, or at least obtain a "good" solution.

REFERENCES

1. M. Bellmore, H. J. Greenberg and J. J. Jarvis, "Generalized Penalty Function Concepts in Mathematical Optimization," Opns. Res., 18 (1970) 229-252.
2. R. B. S. Brooks and A. M. Geoffrion, "Finding Everett's Lagrange Multipliers by Linear Programming," Opns. Res., 14 (1966) 1149-1153.
3. H. Everett, "Generalized Lagrange Multiplier Method for Solving Problems of Optimal Allocation of Resources," Opns. Res., 11 (1963) 399-417.
4. B. L. Fox and D. M. Landi, "Searching for the Multiplier in One-Constraint Optimization Problems," Opns. Res., 18 (1970) 253-262.
5. G. G. Furman and H. J. Greenberg, "Optimal Weapon Allocation with Overlapping Area Defenses," Opns. Res., 21 (1973) 1291-1308.
6. H. J. Greenberg and T. C. Robbins, "Finding Everett's Lagrange Multipliers by Generalized Linear Programming, Part I: Theoretical Foundations," CS/OR Tech. Rept. No. CP-70008, Southern Methodist University, Dallas, Texas, 1970.
7. L. S. Lasdon, Optimization Theory for Large Systems, Macmillan Co., New York, 1970.