Clearly, the convexity result of Theorem 3 rules out two disconnected subsets for Z^I as indicated in Fig. 3. We assert that the continuity of $\omega(z)$ (Theorem 4) forbids the partitioning subsets Z^{12}, Z^{23} and Z^{31} from having common interset boundaries. As shown in Fig. 3, suppose that Z^{23} and Z^{12} have a common interset boundary which intersects $[\bar{z},z]$ at $\hat{z}\equiv z(\hat{\theta})$. At $\hat{z}$, since $\omega(z(\theta))$ is continuous from the left, we must have $\omega_1(\hat{z})=0$, so that $\hat{z}\not\in z^{12}$. Hence the assertion made above. We conclude that Z^{123} is the trapezoidal strip separating z^{23} from z^{12} and z^{31}.

The tabulated results from Fig. 2 are plotted to obtain a complete partitioning of the parameter set in Fig. 4 where the following are worth noting.

1) The subsets Z^1, Z^2, and Z^3 include their respective boundaries; that is, they are closed.

2) The subsets Z^{12}, Z^{23}, and Z^{31} exclude their boundaries except for the intersect boundary with Z^{123}.

3) The subset Z^{123} is a completely open trapezoidal area which does not include any of its boundary points. To exhibit Z^{123}, which has very narrow width (on the order of 0.002), portions of Fig. 4 are shown on an enlarged horizontal scale.

V. Conclusions

Our theoretical results on convexity and continuity are capable of detecting the error in an example in the literature and our algorithm provides the correct and complete solution for the same. We have developed an algorithm for computing the parameterized solution to certain types of minmax problems.

References

[1] N. O. Da Cunha and E. Polak, "Constrained minimization under vector-valued criteria in linear topological spaces," in *Mathematical Theory of Control*, A. V. Balakrishnan and L. W. Neustadt, Eds. London, England: Academic, 1967, p. 96.
[2] A. Klinger, "Vector-valued performance criteria," *IEEE Trans. Automat. Contr.* (Corresp.), vol. AC-9, pp. 117–118, Jan. 1964.
[3] A. M. Geoffrion, "Proper efficiency and the theory of vector maximization," *J. Math. Anal. Appl.*, vol. 22, p. 618, 1968.
[4] Y. C. Ho, "Differential games, dynamic optimization and generalized control theory," *J. Optimiz. Theory Appl.*, vol. 6, p. 179, Sept. 1970.
[5] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proc. 2nd Berkeley Symp. Mathematical Statistics and Probability*, 1951, p. 481.
[6] S. Karlin, *Mathematical Methods and Theory in Games, Programming and Economics*, vol. 1. Reading, Mass.: Addison-Wesley, 1959.
[7] J. Medanic and M. Andjelic, "On a class of differential games without saddle point solutions," *J. Optimiz. Theory Appl.*, vol. 8, no. 6, 1971.
[8] R. Muralidharan, "Parameterized solution of a family of minimax problems," Ph.D. dissertation, Harvard Univ., Cambridge, Mass., 1973.
[9] J. Medanic and M. Andjelic, "Minmax solution of the multiple-target problem," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 597–604, Oct. 1972.
[10] R. Muralidharan and Y. C. Ho, "A piecewise closed form algorithm for a family of minmax and vector criteria problems," presented at the IEEE Decision and Control Conf., Phoenix, Ariz., Paper THA-2.3, Nov. 1974.
[11] W. I. Zangwill, *Nonlinear Programming: A Unified Approach*. Englewood Cliffs, N. J.: Prentice-Hall, 1969.
[12] R. Muralidharan, "Algorithm for computing the parameterized solution of a family of minmax problems," Div. Engineering and Applied Physics, Harvard Univ., Cambridge, Mass., Tech. Rep. 650, Mar. 1974.

On the Method of Multipliers for Convex Programming

DIMITRI P. BERTSEKAS

***Abstract*—It is known that the method of multipliers for constrained minimization can be viewed as a fixed stepsize gradient method for solving a certain dual problem. In this short paper it is shown that for convex programming problems the method converges globally for a wide range of possible stepsizes. This fact is proved for both cases where unconstrained minimization is exact and approximate. The results provide the basis for considering modifications of the basic stepsize of the method of multipliers which are aimed at acceleration of its speed of convergence. A few such modifications are discussed and some computational results are presented relating to a problem in optimal control.**

Manuscript received May 28, 1974; revised January 15, 1975. Paper recommended by E. Barnes, Past Chairman of the IEEE S-CS Computational Methods Committee. This work was carried out partly at the Department of Engineering-Economic Systems, Stanford University, Stanford, Calif. and supported in part by the National Science Foundation under Grant GK 29237 and partly at the Coordinated Science Laboratory, University of Illinois, Urbana, and supported in part by the Joint Services Electronics Program under Contract DAAB-07-72-C-0259 and in part by the U. S. Air Force under Grant AFOSR 73-2570.

The author is with the Department of Electrical Engineering and the Coordinated Science Laboratory, University of Illinois, Urbana, Ill. 61801.

I. Introduction

A sequential unconstrained minimization technique originally proposed by Hestenes [4] and Powell [7] and known as the method of multipliers has rapidly become a focal point of attention in the area of constrained minimization. The properties of the method have been investigated by a number of authors (see [1] for a more complete account) and it has been demonstrated that multiplier methods offer distinct advantages over standard penalty methods.

One way to view multiplier methods is to consider them as fixed stepsize gradient methods for solving a certain dual problem. This viewpoint has been adopted by the author in [1] where local convergence and rate of convergence results were given for general constrained minimization problems. Furthermore there was given in [1, Sect. 5] an analysis of the possibility for altering the basic stepsize of the method. It was pointed out that for problems with inherently convex structure there is a potential for acceleration of convergence of the method by modification of its basic stepsize. One such possibility based on an extrapolation device was discussed, and some related convergence and rate of convergence results were given together with a computational example.

The purpose of this short paper is to complement the analysis of [1] by providing some new results for convex programming problems which establish convergence of the method of multipliers for a wide range of stepsizes. These results may be used to guarantee that certain modified stepsize rules based on extrapolation and aimed at accelerating convergence will not destroy the overall convergence of the algorithm. A brief discussion of such rules and some computational results which complement those given in [1] are provided in the last section of this short paper.

II. The Method of Multipliers for Convex Programming

Consider the following convex programming problem

$$\text{minimize } f_0(x) \text{ subject to } x\in X\subset R^n, f_i(x)\leqslant 0, \qquad i=1,\cdots,m. \tag{1}$$

The functions f_0, f_i, $i=1,\cdots,m$ are real valued convex functions on R^n and X is a closed convex subset of R^n. We make the following two standing assumptions.

Assumption 1: Problem (1) has a nonempty and compact solution set.

Assumption 2: Problem (1) has a nonempty and compact set of Lagrange multiplier vectors Y^*.

It is well known that Assumption 2 is satisfied if, for example, there exists a point $\bar{x}\in X$ such that $f_i(\bar{x})<0$, $i=1,\cdots,m$. Consider now the ordinary dual functional $g_0:R^m\to[-\infty,\infty)$

$$g_0(y)=\begin{cases}\inf\limits_{x\in X}\left\{f_0(x)+\sum\limits_{i=1}^{m}y^if_i(x)\right\} & \text{if } y^i\geqslant 0, i=1,\ldots,n\\ -\infty \text{ otherwise}\end{cases} \tag{2}$$

and for any $c>0$ the functional $g_c:R^m\to(-\infty,\infty)$

$$g_c(y)=\inf_{x\in X}\left\{f_0(x)+\frac{1}{2c}\sum_{i=1}^{m}\left[\{\max[0,y^i+cf_i(x)]\}^2-(y^i)^2\right]\right\} \tag{3}$$

where we denote by y^i the ith coordinate of the vector $y\in R^m$. The functional g_c plays a central role in the computational and analytical aspects of multiplier methods [8], [10], [5], [6]. We summarize below the properties of g_c which are of interest to us.

Property 1: The functionals $g_c, c>0$ and g_0 have a common set of maximizing points, the set of Lagrange multipliers Y^* of problem (1) [10]; [5].

Property 2: The minimization problem indicated in (3) has a solution for every $y\in R^m$ [5], [6].

Property 3: The functional g_c is real valued, concave and continuously differentiable. Its gradient $\nabla g_c(y)$ has coordinates given by

$$\frac{\partial g_c(y)}{\partial y^i} = \max\left[-\frac{y^i}{c}, f_i[x(y)]\right] \qquad i=1,\cdots,m \tag{4}$$

where $x(y)$ is any point (not necessarily unique) attaining the infimum in (3) [9], [5].

Property 4: For any $y', y \in R^m$ we have [9]

$$g_c(y') \geq g_c(y) + \langle \nabla g_c(y), y'-y\rangle - \frac{1}{2c}\|y'-y\|^2. \tag{5}$$

The method of multipliers is simply the fixed stepsize gradient iteration

$$y_{k+1} = y_k + c\nabla g_c(y_k) \qquad k=0,1,\cdots \tag{6}$$

where the gradient $\nabla g_c(y_k)$ is obtained via (4) and the minimization indicated in (3).

We shall consider exact or inexact implementations of iterations of the following general form which includes as a special case the iteration (6)

$$y_{k+1} = y_k + a_k \nabla g_c(y_k) \qquad k=0,1,\cdots \tag{7}$$

$$\delta c \leq a_k \leq 2(1-\delta)c \qquad k=0,1,\cdots \tag{8}$$

where δ, c are any scalars satisfying $0<\delta\leq\frac{1}{2}$, $c>0$. The next section provides global convergence results for iterations of this type.

III. Convergence Results

We have the following proposition.

Proposition 1: The sequence $\{y_k\}$ generated by iteration (7), (8) is bounded and each of its limit points is a Lagrange multiplier of problem (1).

Proof: From (5), (7), (8) we obtain

$$g_c(y_{k+1}) \geq g_c(y_k) + a_k\|\nabla g_c(y_k)\|^2 - \frac{a_k^2}{2c}\|\nabla g_c(y_k)\|^2$$

$$= g_c(y_k) + \frac{a_k(2c-a_k)}{2c}\|\nabla g_c(y_k)\|^2 \geq g_c(y_k) + c\delta^2\|\nabla g_c(y_k)\|^2. \tag{9}$$

Hence $g_c(y_{k+1}) \geq g_c(y_k)$ for all k and $\{y_k\}$ belongs to the set $\{y\in R^m | g_c(y) \geq g_c(y_0)\}$. But this set is compact since it is a level set of the concave function g_c which has a compact set of maximizing points Y^* [11, Cor. 8.7.1]. Hence $\{y_k\}$ is bounded. Also from (9) we obtain that $\|\nabla g_c(y_k)\|\to 0$ from which, the result follows. Q.E.D.

Consider now approximate implementations of iteration (7), (8) of the form

$$y_{k+1} = y_k + a_k p(x_k,y_k), \quad \delta c \leq a_k \leq 2(1-\delta)c \tag{10}$$

where $p(x_k,y_k)$ is an approximation to $\nabla g_c(y_k)$ obtained by a procedure to be described in what follows. Assume, in addition to Assumptions 1 and 2, the following.

Assumption 3: $X=R^n$ and the functions $f_0, f_1, \cdots, f_m$ are differentiable.

Assumption 4: The function to be minimized in (3)

$$F(x,y) = f_0(x) + \frac{1}{2c}\sum_{i=1}^{m}\left[\{\max[0, y^i + cf_i(x)]\}^2 - (y^i)^2\right] \tag{11}$$

satisfies for some $q>0$

$$F(x',y) \geq F(x,y) + \langle \nabla_x F(x,y), x'-x\rangle + \frac{q}{2}\|x'-x\|^2, \forall x,x' \in R^n, y\in R^m.$$

(This assumption is satisfied in particular if f_0 is a uniformly convex function.)

We define $p(x_k,y_k)$ in (10) to be any vector with coordinates

$$p^i(x_k,y_k) = \max\left[-\frac{y_k^i}{c}, f_i(x_k)\right] \qquad i=1,\cdots,m \tag{12}$$

with x_k any vector satisfying the criterion

$$\|\nabla F(x_k,y_k)\| \leq r_k\|y_{k+1}-y_k\| \tag{13}$$

where $\nabla F(x_k,y_k)$ denotes the gradient of F with respect to x, and y_{k+1} is given in terms of x_k, y_k by (10), (12). The scalar r_k is the kth element of an *a priori* fixed sequence $\{r_k\}$ with $r_k > r_{k+1} > 0, r_k \to 0$. The definition of $p(x_k,y_k)$ is motivated by the form of $\nabla g_c(y_k)$. Since $p(x_k,y_k) = \nabla g_c(y_k)$ whenever $\nabla F(x_k,y_k)=0$ one can rightfully view $p(x_k,y_k)$ as an approximation to $\nabla g_c(y_k)$. The inequality (13) may be viewed as a termination criterion for the minimization of $F(x_k,y_k)$. Similar termination criteria have been introduced and discussed in [1], [5], [6]. Notice that the inequality (13) is satisfied for any point x_k which minimizes $F(x,y_k)$. However one can easily show that if y_k is not a Lagrange multiplier then (13) is satisfied for all points in an appropriate neighborhood of the set of minimizing points of $F(x,y_k)$. Thus the criterion (13) will be satisfied within a finite number of iterations during the minimization of $F(x,y_k)$ if y_k is not a Lagrange multiplier. We have the following convergence result.

Proposition 2: Let $\{y_k\}$ be any sequence of points generated by the algorithm (10), (12) with x_k satisfying the criterion (13). Then if $\{y_k\}$ is a bounded sequence, every limit point of $\{y_k\}$ is a Lagrange multiplier for problem (1).

Proof: It has been shown in [5] that the fact that the sequence $\{y_k\}$ is bounded implies that the set

$$\bar{X} = \bigcup_{k=0}^{\infty}\{x \mid \|\nabla F(x,y_k)\| \leq r_k\|y_{k+1}-y_k\|\}$$

is also bounded. This implies by [11, Theorem 10.4] that $p(x,y_k)$ satisfies the following uniform Lipschitz condition:

$$\|p(x',y_k)-p(x,y_k)\| \leq \left\{\sum_{i=1}^{m}[f_i(x)-f_i(x')]^2\right\}^{\frac{1}{2}} \leq L\|x-x'\| \forall x',x\in\bar{X}$$
$$k=0,1,\cdots \tag{14}$$

where $L>0$ is the Lipschitz constant.

Also from Assumption 4 we have for any vector $x(y_k)$ such that $\nabla F[x(y_k),y_k]=0$

$$F(x_k,y_k) \geq F[x(y_k),y_k] + \frac{q}{2}\|x_k - x(y_k)\|^2$$

$$F[x(y_k),y_k] \geq F(x_k,y_k) + \langle \nabla F(x_k,y_k), x(y_k)-x_k\rangle + \frac{q}{2}\|x_k-x(y_k)\|^2$$

where $\{x_k\}\subset\bar{X}$ is the sequence generated by the algorithm. It follows that

$$q\|x_k-x(y_k)\|^2 \leq \langle \nabla F(x_k,y_k), x_k - x(y_k)\rangle \leq \|\nabla F(x_k,y_k)\|\,\|x_k-x(y_k)\|.$$

Thus we obtain

$$q\|x_k - x(y_k)\| \leq \|\nabla F(x_k,y_k)\| \tag{15}$$

for every vector $x(y_k)$ such that $\nabla F[x(y_k),y_k]=0$. In addition, we have by (12) and (4)

$$p[x(y_k),y_k] = \nabla g_c(y_k). \tag{16}$$

From (13)–(16) there follows

$$\|p(x_k,y_k)-\nabla g_c(y_k)\| \leq L\|x_k - x(y_k)\|$$

$$\leq \frac{L}{q}\|\nabla F(x_k,y_k)\| \leq \frac{r_kL}{q}\|y_{k+1}-y_k\|. \tag{17}$$

Apply now (5) with $y'=y_{k+1}$ and $y=y_k$. In view of (10), (17) we have for every k

$$g_c(y_{k+1}) \geq g_c(y_k)+\langle \nabla g_c(y_k), y_{k+1}-y_k\rangle - \frac{1}{2c}\|y_{k+1}-y_k\|^2$$

$$= g_c(y_k)+\left(\frac{1}{a_k}-\frac{1}{2c}\right)\|y_{k+1}-y_k\|^2+\langle \nabla g_c(y_k)-p(x_k,y_k), y_{k+1}-y_k\rangle$$

$$\geq g_c(y_k)+\left[\frac{\delta}{2(1-\delta)c}-\frac{r_kL}{q}\right]\|y_{k+1}-y_k\|^2. \tag{18}$$

Since $r_k \to 0$ it follows that $\|y_{k+1}-y_k\|\to 0$. Hence by (10), (17) $p(x_k,y_k)\to 0$ and $\|p(x_k,y_k)-\nabla g_c(y_k)\|\to 0$ implying $\|\nabla g_c(y_k)\|\to 0$ and the result follows. Q.E.D.

We mention that the boundedness assumption on $\{y_k\}$ does not appear to be very restrictive. Also a closer examination of the proof reveals that if either all the functions f_i, $i=1,\cdots,m$ are Lipschitz continuous, or the sequence $\{r_k\}$ is bounded above by a sufficiently small positive number then the boundedness of $\{y_k\}$ is guaranteed and need not be assumed. In any case it is possible to eliminate the boundedness assumption if one specifies upper and lower bounds on the iterates y_k, say, $0\leq y_k^i \leq A^i$, and modifies the iteration (10) to take the form

$$y_{k+1}^i=\begin{cases} A^i & \text{if } y_k^i+a_kp^i(x_k,y_k)\geq A^i \\ y_k^i+a_kp^i(x_k,y_k) & \text{if } 0\leq y_k^i+a_kp^i(x_k,y_k)\leq A^i \\ 0 & \text{if } y_k^i+a_kp^i(x_k,y_k)\leq 0 \end{cases} \tag{19}$$

$$\delta c \leq a_k \leq 2(1-\delta)c. \tag{20}$$

Proposition 3: Let $\{y_k\}$ be any sequence of points generated by the algorithm (19), (20) with x_k satisfying the criterion (13). Assume also that the scalars A^i satisfy $A^i \geq \bar{y}^i$, $i=1,\cdots,m$ for some Lagrange multiplier $\bar{y}=(\bar{y}^1,\cdots,\bar{y}^m)$ of problem (1). Then every limit point of the sequence $\{y_k\}$ is a Lagrange multiplier for problem (1).

Proof: Proceeding similarly as in the proof of Proposition 2 one obtains for all k [c.f. (17)]

$$g_c(y_{k+1})\geq g_c(y_k)+\left[\frac{\delta}{2(1-\delta)c}-\frac{r_kL}{q}\right]\|y_{k+1}-y_k\|^2.$$

Hence $\|y_{k+1}-y_k\|\to 0$. Now if a subsequence $\{y_k\}_{k\in K}$ of $\{y_k\}$ converges to a point $\bar{y}$ we have for any z with $0\leq z^i\leq A^i$, $i=1,\cdots,m$ and every $k\in K$

$$\langle \nabla g_c(y_k), z-y_k\rangle = \langle p(x_k,y_k), z-y_{k+1}\rangle + \langle p(x_k,y_k), y_{k+1}-y_k\rangle$$

$$+\langle \nabla g_c(y_k)-p(x_k,y_k), z-y_k\rangle$$

$$\leq \frac{1}{a_k}\langle y_{k+1}-y_k, z-y_{k+1}\rangle + \langle p(x_k,y_k), y_{k+1}-y_k\rangle$$

$$+\|\nabla g_c(y_k)-p(x_k,y_k)\|\,\|z-y_k\|$$

$$\leq \|y_{k+1}-y_k\|\,\left\|\frac{z-y_{k+1}}{a_k}+p(x_k,y_k)\right\|+\frac{r_kL}{q}\|y_{k+1}-y_k\|\,\|z-y_k\|.$$

Taking limits as $k\to\infty$ the right hand side tends to zero. It follows that

$$\langle \nabla g_c(\bar{y}), z-\bar{y}\rangle \leq 0 \quad \forall z,\ 0\leq z^i\leq A^i, \qquad i=1,\cdots,m.$$

Hence $\bar{y}$ maximizes g_c over the set $\{z|0\leq z^i\leq A^i, i=1,\cdots,m\}$. Since this set contains by assumption a Lagrange multiplier it follows that $\bar{y}$ is a Lagrange multiplier for problem (1). Q.E.D.

Notice that upper bounds A^i to a Lagrange multiplier are readily available if a feasible point interior to the constraints and a lower bound to the optimal value of the problem are known [3, p. 647].

We finally note that the convergence results of this section can be trivially generalized to the practically important case of the iteration

$$y_{k+1}=y_k+a_k\nabla g_{c_k}(y_k)$$

$$\delta c_k \leq a_k \leq 2(1-\delta)c_k$$

with $c_k=\min\{s_k,\bar{c}\}$ where $\{s_k\}$ is a sequence of positive numbers tending to infinity and $\bar{c}$ is an arbitrarily large positive constant.

IV. Modified Stepsize Rules for the Method of Multipliers

This section considers various rules for choosing the stepsize a_k in the iteration (7), (8) with the purpose of accelerating convergence. These rules are based on extrapolation of one form or another and therefore they are meaningful only for the case where the ordinary dual functional g_0 is sufficiently smooth. We shall assume that the dual functional has a unique maximizing point $\bar{y}$ and is twice continuously differentiable in a relative neighborhood of the form $\{y|y^i=0\ \forall i \text{ s.t. } \bar{y}^i=0\}\cap\{y|\,\|y-\bar{y}\|<\epsilon\}$, where $\epsilon>0$ is some scalar. This assumption is satisfied for example if $X=R^n$ and problem (1) has a unique solution $\bar{x}$ satisfying the standard second order sufficiency conditions for a minimum. We note that one can easily prove that, under these assumptions, the approximate Lagrange multipliers generated by iteration (7), (8) which correspond to inactive constraints will converge to zero within a finite number of iterations. This fact suggests that extrapolation will eventually involve only the Lagrange multipliers corresponding to active constraints.

The extrapolation devices that we consider are based on the fact that every time the function $F(x,y_k)$ of (11) is minimized yielding a point x_k then one obtains both the value and the gradient of the ordinary dual functional g_0 at the point

$$\tilde{y}_{k+1}=y_k+c\nabla g_c(y_k).$$

This value and gradient are given by

$$g_0(\tilde{y}_{k+1})=f_0(x_k)+\sum_{i=1}^{m}\tilde{y}_{k+1}^if_i(x_k) \tag{21}$$

$$\frac{\partial g_0(\tilde{y}_{k+1})}{\partial y^i}=\max\left[-\frac{y_k^i}{c}, f_i(x_k)\right] \qquad i=1,\cdots,m. \tag{22}$$

One possibility for extrapolation based on the relations above has been described in [1]. Given y_0 one finds $y_1=y_0+c\nabla g_c(y_0)$ and $\tilde{y}_2=y_1+c\nabla g_c(y_1)$ by means of unconstrained minimization. Then a quadratic or cubic approximation of $g_0(y)$ is made along the line passing through y_1, $\tilde{y}_2$ based on the knowledge of $g_0(y_1)$, $\nabla g_0(y_1)$, $g_0(\tilde{y}_2)$, $\nabla g_0(\tilde{y}_2)$ given by (21), (22). Then one sets

$$y_2=y_1+a_2\nabla g_c(y_1)$$

where a_2 is the stepsize which maximizes the value of the quadratic or cubic approximation to $g_0(y)$ over all a in an interval $[\delta c, 2(1-\delta)c]$. If the approximation is quite accurate (as one expects it to be near the solution $\bar{y}$) then the point y_2 should be closer to $\bar{y}$ than the point $\tilde{y}_2=y_1+c\nabla g_c(y_1)$ which would be obtained by the ordinary stepsize of the method of multipliers. At the same time the convergence of the method is not destroyed since the stepsize a_2 will be in the interval of convergence $[\delta c, 2(1-\delta)c]$. Once y_2 has been obtained as above, one proceeds to determine $y_3=y_2+c\nabla g_c(y_2)$, $\tilde{y}_4=y_3+c\nabla g_c(y_3)$, $g_0(y_3)$, $\nabla g_0(y_3)$, $g_0(\tilde{y}_4)$, $\nabla g_0(\tilde{y}_4)$. Based on this information a quadratic or cubic approximation of $g_0(y)$ along the line through y_3, $\tilde{y}_4$ is made to determine y_4 as earlier. Thus by performing extrapolation at every second iteration and keeping the stepsize within the interval $[\delta c, 2(1-\delta)c]$ convergence is maintained while hopefully acceleration of convergence is achieved. Notice that the approximation is performed only at every second iteration since at the points $y_2, y_4,\cdots,y_{2k},\cdots$, neither the value of g_0 nor its gradient are known. However it is still possible to perform

TABLE I

	Multiplier Method		Multiplier Method with Extrapolation	
k	y_k	$\nabla g_c(y_k)$	y_k	$\nabla g_c(y_k)$
0	0	0.47010	0	0.47010
1	0.47010	0.14590	0.47010	0.14590
2	0.61600	0.05080	0.69914	−0.00222
3	0.66680	0.01830	0.69563	0.00000
4	0.68511	0.00669		
5	0.69181	0.00242		
6	0.69423	0.00089		

the approximation at every step when there is only one constraint and the dual problem is one-dimensional. For this case it appears that the extrapolation device is very efficient.

Another possibility for approximating at every step is based on the fact that even though $g_0(y_2)$ and $\nabla g_0(y_2)$ are unavailable, the approximation performed along the line through y_1, $\tilde{y}_2$ yields an approximate value for $g_0(y_2)$. This value together with $g_0(\tilde{y}_3)$, $\nabla g_0(\tilde{y}_3)$ may be used to construct a quadratic approximation for g_0 along the line through y_2, $\tilde{y}_3$ and determine y_3 by maximizing the approximation of g_0 within the interval of convergence. The process is similarly repeated at every step. In this way convergence may be even faster. However since the approximation errors are accumulated in this procedure it is perhaps wise to use it only after one is fairly close to the solution.

The modified stepsize rule which involves quadratic extrapolation at every second iteration was tested in [1] both for the case of exact and inexact minimization on the Rosen–Suzuki problem and resulted in significant computatonal savings for the case where the penalty parameter c was kept constant or was increased at a relatively low rate. This may be explained by the fact that when the penalty parameter is relatively small, more unconstrained minimization cycles are typically required and hence the beneficial effect of extrapolation is utilized more often. We provide below an example where only one constraint is handled by means of a penalty function and hence one may use extrapolation at every iteration. The computational results support the natural conjecture that for cases involving only one constraint the extrapolation device is extremely effective, much more so than in typical cases where several constraints are handled by means of a penalty. Consider an optimal control problem involving the system

$$x_{k+1}=x_k+u_k \qquad k=0,1,\cdots,N-1$$

and the cost functional

$$J(u_0,u_1,\cdots,u_{N-1})=\sum_{k=0}^{N-1}(e^{-\beta_k u_k}-1)$$

subject to $x_N \leq 1$, $x_0=0$, $0 \leq u_k$, $k=0, 1,\cdots,N-1$ and where $\beta_0,\cdots,\beta_{N-1}$ are given positive scalars. This problem may be viewed as a problem of optimal allocation of a finite amount of resource into N different activities. The method of multipliers with and without cubic extrapolation was used to eliminate the terminal state constraint $x_N \leq 1$. The results for the case $N=10$, $c=1$, and a sequence $\{\beta_k\} = \{10^{-5},10^{-4},10^{-3},10^{-2},10^{-1},1,10,20,40,40\}$, are shown in Table I.

References

[1] D. P. Bertsekas, "Combined primal-dual and penalty methods for constrained minimization," Dep. Eng.-Econ. Syst., Stanford Univ., Stanford, Calif., working paper, Jan. 1973; also in *SIAM J. Contr.*, Aug. 1975, to be published.
[2] ——, "Convergence rate of penalty and multiplier methods," in *Proc. 1973 IEEE Conf. Decision and Control*, San Diego, Calif., Dec. 1973.
[3] D. P. Bertsekas and S. K. Mitter, "A descent numerical method for optimization problems with nondifferentiable cost functionals," *SIAM J. Contr.*, vol. 11, no. 4, pp. 637–652, 1973.
[4] M. R. Hestenes, "Multiplier and gradient methods," *J. Optimiz. Theory Appl.*, vol. 4, no. 5, pp. 303–320, 1969.
[5] B. W. Kort and D. P. Bertsekas, "Combined primal-dual and penalty methods for convex programming," Dep. Eng.-Econ. Syst., Stanford Univ., Stanford, Calif., working paper, Aug. 1973, submitted for publication.
[6] ——, "Multiplier methods for convex programming," in *Proc. 1973 IEEE Conf. Decision and Control*, San Diego, Calif., Dec. 1973.
[7] M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, Ed. New York: Academic, 1969, pp. 283–298.
[8] R. T. Rockafellar, "New applications of duality in convex programming," presented at the 7th Int. Symp. Math. Programming, The Hague, The Netherlands, 1970; also in *Proc. 4th Conf. Probability*, Brasov, Romania, 1971.
[9] ——, "A dual approach to solving nonlinear programming problems by unconstrained optimization," *Math. Prog.*, vol. 5, no. 3, pp. 354–373, 1973.
[10] ——, "The multiplier method of Hestenes and Powell applied to convex programming," *J. Optimiz. Theory Appl.*, vol. 12, no. 6, pp. 555–562, 1973.
[11] ——, *Convex Analysis*. Princeton, N. J.: Princeton Univ. Press, 1970.

Control of Linear Discrete-Time Stochastic Dynamic Systems with Multiplicative Disturbances

MASANAO AOKI, SENIOR MEMBER, IEEE

Abstract—**Multiplicative random disturbances frequently occur in economic modeling. The money multiplier in a simple monetary macroeconomic model is treated as a random variable in this paper. The optimal control law is derived, and some consequences of erroneous modeling of the random disturbance are exhibited by simulation.**

I. Introduction

There seems to be considerable interest in applying, or in assessing the applicability of stochastic control techniques to econometric models, ranging from a single equation model to large scale econometric planning or forecasting models of national economies. In some of these models random coefficients are important and are likely to be used increasingly in econometrics. In other words, in models of economic origin, random disturbances are often modeled as multiplicative disturbances. Much of the stochastic control literature is concerned with models in which random disturbances are modeled as additive disturbances in the coefficients.

Control rules resulting from these two different specifications of stochastic disturbances are usually, and sometimes substantially, different. Caution is necessary in choosing stochastic specifications, since control rules based on models which incorrectly specify stochastic disturbances may destabilize systems rather than stabilize them. Although it is not often clear *a priori* whether additive random disturbances are more plausible than multiplicative ones, we must understand implications of multiplicative random disturbances in stochastic control problems.

The purpose of the paper is to call attention to the importance of correct stochastic specification by considering a stochastic control model in which a random coefficient arises naturally from the économics being modeled. We consider short-run control of the money stock. This problem gives rise to a control problem in which the control "gain" is a random variable because there are "slippages" from the source base which the monetary authority actually controls.

The model is described in Section II and the institutional reason for the randomness of the money multiplier is described in the Appendix. Section III discusses control of a simple macroeconomic model with the random money multiplier. We derive control rules when the randomness in the money multiplier is modeled as multiplicative and as additive disturbances. Some aspects of the stochastic misspecification are discussed using simulation in Section IV.

II. Random Money Multiplier

Consider a control system described by a linear difference equation

$$x_{t+1}=A_t x_t+m_t u_t \tag{1}$$

where u_t is the scalar control variable and where m_t is a control gain vector. (This assumption of the scalar control variable is made for the simplicity of presentation.)

Manuscript received June 13, 1974; revised February 3, 1975. Paper recommended by D. Pekelman, Past Chairman of the IEEE S-CS Economic Systems Committee.
The author was with the Department of System Science, University of California, Los Angeles, Calif. 90024. He is now with the Departments of Economics and Electrical Engineering, University of Illinois, Urbana, Ill. 61801.