Production optimization for a monopolist who does not set prices

L.J. Armour, D.J. Gates, and J.A. Rickard

Consider a monopolistic firm selling each business period's
production at the end of that period at a price determined by the
buyers, and wishing to determine the production for which
business period profit is maximal. In this paper we announce
the results of our investigations into what happens when the firm
employs a certain algorithm (based on linear approximations to
its average contribution profit function) in an attempt to
determine this optimal production. Our results are as follows:
firstly, for many apparently feasible average contribution profit
functions the algorithm generates production sequences globally
convergent to the optimal production, the convergence being
linear with convergence ratio dependent on the average
contribution profit function; secondly, in certain cases a
lower bound for the initial rate of convergence of the algorithm
can be obtained. Proofs are for the most part given only in
outline, and will be published in full later.

1. Introduction

In references [2] and [3], an algorithm is developed which a firm in
competition with a number of other similar firms might use in an attempt to
determine how much it should produce in a business period in order to
maximize its business period revenue. In the development of this algorithm
it is assumed that all of the firm's production for a business period is
sold at the end of that period at a price determined by the buyers, and

_____
Received 17 June 1977.

that the firm *does not know the functional relationship between production and revenue*.   The algorithm uses the production and revenue figures for the current business period and the one preceding it to generate a production figure for the next period, and appears to be novel in that it is based on the use of linear estimates of the firm's *average* revenue function rather than on estimates of the revenue function itself.

Gates and Rickard [2] showed that if the algorithm is employed simultaneously by all of the competing firms, and if it converges to a limiting production for each of them, then this set of limiting productions will be Pareto optimal for suitable revenue functions; the existence of a special class of revenue functions for which convergence actually does occur was also demonstrated. Now the condition for a set of productions to be Pareto optimal in the case of n competing firms reduces simply to the condition for revenue to be maximal when there is just one firm supplying the market, and this has led us to investigate the consequences of the algorithm being employed by a monopolistic firm producing only one type of goods, and selling each business period's production at the end of that period for whatever it will fetch. As a preliminary to this investigation, we have made some modifications to the model that would have resulted had we considered the firm to be operating under the conditions described in [2] and [3]; these have been aimed at producing a more realistic model, and are embodied in the following three assumptions:-

A(I)   The firm aims to maximize *contribution profit* (that is revenue less variable costs) rather than just revenue; thus the functions $J(\sigma)$ and $m(\sigma)$ of [1], where σ represents production, will be taken in the sequel as being the firm's contribution profit and average contribution profit functions respectively.

A(II)   A combination of endogenous and exogenous constraints limits the firm's production for a business period to a maximum of σ_{max} . Taking σ to be a continuous variable, we thus have that σ can be set at any value lying in $[0, \sigma_{max}]$.

A(III)   Market research carried out prior to the firm deciding to go into business indicated that its contribution profit was most likely to be maximal when the production was σ_{RM} , and that its operations would be

profitable only when the production lay in the interval $\left(\sigma_{RB1}, \sigma_{RB2}\right)$, that is, σ_{RB2} and σ_{RB1} are the market research estimates of the firm's upper and lower break even points respectively.

With regard to the first of these assumptions, it should be pointed out that in practice a firm would be more likely to aim at maximizing a discounted sum of the profits over all business periods since it commenced its operations rather than at determining the production for which contribution profit (and hence total profit) for a business period is maximal. However, the problem of maximizing a discounted sum of values of an unknown function does not appear to be at all tractable, and we have, to this stage, confined ourselves to the consideration of the simpler (but slightly less realistic) problem.

Gates and Rickard [2] implicitly assumed that if the firm produces the same amount in each of two business periods then the profit realized in each of those periods will be the same; this amounts to assuming that economic conditions are stable. We now make this assumption explicitly, but in so doing point out that the algorithm could also be used under inflationary conditions if suitably discounted values of costs and revenue were used.

We have now set the scene so far as the economic environment, the form of the market, and the nature and major managerial goal of the firm are concerned. (Before proceeding further, we should perhaps make the fairly obvious point that the form of the market indicates that we are probably not considering a manufacturing firm.) We now make a final assumption, which is concerned with the firm's belief as to the general form of $m(\sigma)$:

A(IV) The preliminary market research indicated that $m(\sigma)$ belongs to the class of functions (which we will refer to as the class C) that are continuous and monotonic decreasing in $\left(0, \sigma_{max}\right]$, and positive in some neighbourhood of the origin.

Assumption A(IV) is crucial, since the whole rationale for the algorithm's development is based on the supposition that $m(\sigma)$ is monotonic decreasing in $\left(0, \sigma_{max}\right]$, so that if the firm believed that $m(\sigma) \in C$ this would provide the motivation for it to use the algorithm. Without going into details (these may be found in [1]) it appears to us

that under our assumed marketing and economic conditions, average contribution profit functions belonging to the class C would be likely to occur fairly frequently, so that in making assumption A(IV) we are, so far as we can see, modelling real world situations.

In the remainder of this paper we report the results of our investigations; in brief, we have shown that:

R(I) For many classes of likely average contribution profit functions, employment of the algorithm leads to the determination of σ^* , the production for which business period contribution profit is maximal.

R(II) When a production sequence generated by the algorithm converges, its asymptotic behaviour is to converge linearly with convergence ratio dependent on $m(\sigma)$.

R(III) In certain cases, a lower bound for the *initial* rate of convergence of the algorithm can be obtained; this initial rate is of course likely to be of more interest to the firm than the asymptotic rate.

The proof of the result R(I) is relatively straightforward, but tedious, and we will simply indicate the method used. The second result is proved in a fairly standard manner, so again the full proof will not be given; however, the proof of the result R(III) is quite short, and will be given in full. Before we proceed, it would perhaps be advisable to give a brief résumé of the development of the algorithm, and describe some auxiliary procedures that would be employed in certain singular situations.

2. The algorithm

Suppose $\sigma(t)$ and $\sigma(t-1)$ are the two most recent productions, with $J(t)$ and $J(t-1)$ the corresponding contribution profits, and $m(t)$ and $m(t-1)$ the respective average contribution profits (obtained by dividing the contribution profits by the corresponding productions). The firm approximates to the unknown function $m(\sigma)$ by the linear function $\tilde{m}(\sigma|t)$ whose graph passes through the points $\big(\sigma(t), m(t)\big)$ and $\big(\sigma(t-1), m(t-1)\big)$, and intersects the σ-axis at $\sigma_I(t)$. The firm now constructs the *approximate* contribution profit function $\tilde{J}(\sigma|t) = \sigma\tilde{m}(\sigma|t)$. If $m(\sigma) \in C$, as the firm believes, then $\tilde{m}(\sigma|t)$ has negative gradient, and elementary calculus gives that $\tilde{J}(\sigma|t)$ is maximal at $\sigma = \frac{1}{2}\sigma_I(t)$. This

value of σ is now taken as $\sigma(t+1)$, so that we have

(1) $\sigma(t+1) = \tfrac{1}{2}\sigma_I(t)$,

that is

(2) $\sigma(t+1) = \tfrac{1}{2}\{\sigma(t) - m(t)[\sigma(t) - \sigma(t-1)]/[m(t) - m(t-1)]\}$.

(We note here that (2) was obtained by Gates and Rickard [2], but by direct differentiation of $\tilde{J}(\sigma|t)$ with respect to σ rather than via (1); (1) and (2) are obviously equivalent, but one or the other will be more convenient to use, depending on the situation under discussion.)

In order to use this production generating procedure, the firm has to set values for the initial productions $\sigma(1)$ and $\sigma(2)$. It will be shown in the following section that the choice made for these values is immaterial so far as the ultimate attainment of the firm's goal is concerned; however, for the sake of definiteness we will suppose that $\sigma(1) = \sigma_{RM}$ and $\sigma(2) = \tfrac{1}{2}(\sigma_{RB1} + \sigma_{RB2})$ *unless* $J(1)$ is negative, in which case $\sigma(2)$ is taken equal to σ_{RB1} (for a discussion of the rationale for these choices see [1]).

When using the production generating equations above, three singular cases can arise; these cases, and the auxiliary procedures the firm uses to deal with them, are as follows:

(i) Due to production values being calculated only to a certain level of accuracy, it may happen that the calculated value of $\sigma(t+1)$ is equal to $\sigma(t)$, so that $\sigma(t+2)$ is indeterminate. The firm proceeds by choosing as $\sigma(t+1)$ a value infinitesimally smaller than $\sigma(t)$, that is, it chooses $\sigma(t+1) = \sigma(t) - \varepsilon$, $\varepsilon > 0$. If subsequently two other successive productions are calculated as having the value $\sigma(t)$, it chooses instead of the second of these the value $\sigma(t) - \varepsilon/2$, and so on.

(ii) $m(t-1)$ and $m(t)$ are negative, and $\sigma_I(t)$ is non-positive. If $t = 2$, the firm proceeds by choosing as $\sigma(3)$ an infinitesimally small positive value. If $t = \tau > 2$, $\sigma(\tau+1)$ is chosen as σ_{MP} , where σ_{MP} is the production for which $J(\sigma)$ has to that stage been maximal.

(iii) $\sigma(t+1)$ as calculated exceeds σ_{max} . The firm proceeds by

choosing σ_{RB2} as $\sigma(t+1)$. If the subsequently calculated value of $\sigma(t+2)$ exceeds σ_{max} then σ_{max} is chosen as $\sigma(t+2)$, and if $\sigma(t+3)$ as calculated also exceeds σ_{max} , $\sigma(t+3)$ is chosen as infinitesimally smaller than σ_{max} ; that is, $\sigma(t+3) = \sigma_{max} - \varepsilon$, $\varepsilon > 0$. If subsequently a similar situation arises, then σ_{max} , $\sigma_{max} - \varepsilon/2$, and so on, are used instead of $\sigma_{RB2}, \sigma_{max}$, and so on.

Again the reader will find the rationale for these auxiliary procedures in [1]. Of course, other procedures could be devised to deal with the singular situations; however, it turns out that in the long run it is of no consequence what procedures are adopted.

From now on we will refer to (1) together with the auxiliary procedures as Algorithm A (or simply as "the algorithm").

3. Convergence of Algorithm A to the optimal production

We will denote by C_A the class of functions belonging to C and possessing the properties

(A) $m(\sigma)$ is strictly convex in $\left(0, \sigma_{max}\right]$,

(B) $m\left(\sigma_{max}\right) \leq 0$.

Also, we will denote by C_B the class of functions belonging to C_A and possessing the properties

(C) $m(\sigma)$ is differentiable in $\left(0, \sigma_{max}\right)$,

(D) $\sigma m(\sigma)$ has only one stationary value in $\left(0, \sigma_A\right)$, where σ_A is the σ-intercept of the graph of $m(\sigma)$.

We now outline the proofs of the two basic results used to arrive at the result R(I) of Section 1. These are:

(C1) *If* $m(\sigma) \in C_A$ *, then any sequence of productions generated by the algorithm converges to a limit lying in* $\left(0, \sigma_A\right)$ *.*

(C2) *If* $m(\sigma) \in C_B$ *, then the limit referred to in* (C1) *is* σ^* *.*

As a preliminary, we observe that in view of the auxiliary procedures adopted, $\sigma(t) \in (0, \sigma_{max}/2]$ for all t , and that in particular $\sigma(t+1) \in (0, \sigma_A/2]$ if $\sigma(t-1)$ and $\sigma(t)$ both lie in $(0, \sigma_A]$.

Proof of (C1). The proof is by exhaustion; using $\{\sigma(t)\}$ to denote the sequence of productions $\sigma(3), \sigma(4), \ldots$ generated by the algorithm, we consider the evolution of $\{\sigma(t)\}$ for every admissible ordering of $\sigma(1), \sigma(2), \sigma(3)$, and σ_A . It turns out that the basic results are those concerning the situations where $\sigma(1)$ and $\sigma(2)$ both lie in $(0, \sigma_A]$, since in each of the other cases it can be shown that eventually two consecutive members of $\{\sigma(t)\}$ will lie in $(0, \sigma_A]$. As an example of the type of reasoning used, consider the case where $\sigma(2) < \sigma(1) \leq \sigma_A$, and $\sigma(3) < \sigma(2)$. (This is case A.1 of [1].)

By property (A), $\sigma_I(3) < \sigma_I(2)$, so that by (1), $\sigma(4) < \sigma(3)$. By (A) again, $\sigma_I(4) < \sigma_I(3)$, so that $\sigma(5) < \sigma(4)$, and so on; that is, $\{\sigma(t)\}$ is monotonic decreasing, and since it is bounded below (by zero) we thus have that the sequence of productions converges to a limit lying in $[0, \sigma_A)$.

Whilst the proofs in a number of the other cases are slightly less straightforward, we feel that this example is sufficiently typical to render the inclusion of proofs of the less straightforward cases unnecessary.

Proof of (C2). Suppose $m(\sigma) \in C_B$. By (C1), $\{\sigma(t)\}$ will converge to a limit, $\overline{\sigma}$ say, where $\overline{\sigma} \in (0, \sigma_A)$. In view of (C), (2) then becomes

$$\overline{\sigma} = \tfrac{1}{2}\{\overline{\sigma} - m(\overline{\sigma})[m'(\overline{\sigma})]^{-1}\}$$

as $t \to \infty$, which reduces to

$$(3) \qquad\qquad \overline{\sigma}m'(\overline{\sigma}) + m(\overline{\sigma}) = 0 .$$

(Here, and in the sequel, $m'(\sigma) \equiv dm(\sigma)/\sigma$, and so on.) Now in view of (D), σ^* , the optimal production, will be the one solution for σ of

$$(4) \qquad\qquad \sigma m'(\sigma) + m(\sigma) = 0$$

which lies in $(0, \sigma_A)$. Comparing (3) and (4), we conclude that $\bar{\sigma}$ is a stationary value of $J(\sigma)$; since σ^* is the only stationary value of $J(\sigma)$ in $(0, \sigma_A)$, and $\bar{\sigma} \in (0, \sigma_A)$, we thus have that $\bar{\sigma} = \sigma^*$.

From the results (C1) and (C2) it is relatively easy to show that employment of the algorithm leads to the determination of σ^* in each of the following situations:

1. $m(\sigma) \in C$ and possesses properties (A) and (C), and in addition possesses the properties:

(E) $\sigma m(\sigma)$ is unimodal at a maximum in $(0, \sigma_{max}]$.

(F) The tangent to the graph of $m(\sigma)$ at $\sigma = \sigma_{max}$ has σ-intercept lying in $[\sigma_{max}, 2\sigma_{max}]$.

2. $m(\sigma)$ is restricted to a domain $(0, \sigma_{max}]$ of a function belonging to C_B , and possesses the property:

(G) $\sigma m(\sigma)$ has no stationary value in $(0, \sigma_{max}]$.

3. $m(\sigma)$ is linear in $(0, \sigma_{max}]$.

By means of a procedure based essentially on the mean value theorem, we have also been able to show that convergence to σ^* occurs when $m(\sigma) \in C$, and possesses property (E) together with the properties:

(H) The tangent to the graph of $m(\sigma)$ at any point in $(0, \sigma_{max}]$ intersects the σ-axis in $(0, 2\sigma_{max}]$.

(I) $m(\sigma)$ is twice differentiable in $(0, \sigma_{max})$, and in this interval

$$K = m(0) |m''(\sigma)|_{max} / (m'(\sigma))^2_{min} < 1 ,$$

where $|m''(\sigma)|_{max}$ and $(m'(\sigma))^2_{min}$ are respectively the maximum value of $|m''(\sigma)|$ and the minimum value of $(m'(\sigma))^2$ in $(0, \sigma_{max})$, and $m(0) = \sup\{m(\sigma) : \sigma \in (0, \sigma_{max}]\}$. $\big($Since $m(\sigma) \in C$, $K \geq 0$ in

$\left(0, \sigma_{max}\right]$.)

Now in practice there is likely to be a production, σ_{min} , such that
the firm must either produce an amount greater than or equal to σ_{min} , or
produce nothing at all. For this situation it is easy to show from the
preceding results that production sequences generated by the algorithm
converge to σ^* *whether* $m(\sigma)$ *belongs to* C *or not*, providing that
$m\left(\sigma_{min}\right) > 0$ and possesses properties analogous to (A), (C), (E), (H), and
(I) in $\left[\sigma_{min}, \sigma_{max}\right]$.

Since in every one of the above cases convergence to σ^* occurs
regardless of the values of $\sigma(1)$ and $\sigma(2)$, it follows that the choice of
initial productions and of auxiliary procedures to deal with the singular
situations is immaterial.

It should be noted that if $m(\sigma)$ is strictly concave in $\left(0, \sigma_{max}\right]$,
then production sequences generated by Algorithm A may diverge by
oscillation.

4.   Rate of convergence of Algorithm A

In this section we indicate the method of proof for result R(II) of
Section 1, and give the proof of result R(III) in full.

Suppose that $m(\sigma)$ is twice differentiable at $\sigma = \sigma^*$, and belongs
to one of those classes of functions for which production sequences
generated by the algorithm converge to σ^* . Then by expanding $m(t)$ in a
Taylor series about $\sigma = \sigma^*$, using the fact that $-m(\sigma^*)/m'(\sigma^*) = \sigma^*$, and
neglecting second and higher order terms in the series, we obtain, for
sufficiently large t , the approximate equation

(5) $4\varepsilon(t+1) = M[\varepsilon(t)+\varepsilon(t-1)]$

where $\varepsilon(t) = \sigma(t) - \sigma^*$ and $M = m(\sigma^*)m''(\sigma^*)/[m'(\sigma^*)]^2$.

It is easily shown that $0 \le M < 2$, and that $-\frac{1}{2} < \lambda_2 \le 0 \le \lambda_1 < 1$,
where λ_1 and λ_2 are the roots of the auxiliary equation of (5). Since
λ_1^t dominates λ_2^t for t large, we thus have

$$\lim_{t \to \infty} \left| \varepsilon(t+1)/\varepsilon(t) \right| = \lambda_1 \ ,$$

so that production sequences generated by Algorithm A converge linearly with convergence ratio λ_1 . (Our definitions of linear convergence and convergence ratio are those of Luenberger [4].) We note that the convergence ratio in this case is not an intrinsic property of the algorithm, being determined by properties of $m(\sigma)$ at $\sigma = \sigma^*$.

Whilst the asymptotic convergence behaviour of an algorithm is important in many situations, a firm trying to determine its optimal production would probably be more interested in the *initial* rate of convergence. In this regard, we have been able to show that if $\{\sigma(t)\}$ exhibits either of two certain types of behaviour then an upper bound can be obtained for $\left| \sigma(t+1)-\sigma^* \right| / \left| \sigma(t)-\sigma^* \right|$ for $t = 2, 3, 4, \ldots$.

TYPE I. $\{\sigma(t)\}$ is monotonic decreasing and $\{\sigma_I(t)-\sigma(t)\}$ is monotonic increasing. Since at σ^* the tangent to the graph of $m(\sigma)$ has σ-intercept $2\sigma^*$, $\{\sigma_I(t)-\sigma(t)\} \to \sigma^*$, and since $\{\sigma_I(t)-\sigma(t)\}$ is monotonic increasing we have

$$\sigma_I(t) - \sigma(t) < \sigma^* \ \ \text{for} \ \ t \geq 2 \ ,$$

that is,

$$2\sigma(t+1) - 2\sigma^* < \sigma(t) - \sigma^* \ ,$$

so that

$$[\sigma(t+1)-\sigma^*]/[\sigma(t)-\sigma^*] < \tfrac{1}{2} \ \ \text{for} \ \ t \geq 2$$

$\Big($since $\sigma(t) > \sigma^*$ for $t \geq 2\Big)$.

TYPE II. $\{\sigma(t)\}$ is monotonic increasing and $\{\sigma_I(t)-\sigma(t)\}$ is monotonic decreasing. Again, $\{\sigma_I(t)-\sigma(t)\} \to \sigma^*$, and since $\{\sigma_I(t)-\sigma(t)\}$ is monotonic decreasing we have

$$\sigma_I(t) - \sigma(t) > \sigma^* \ \ \text{for} \ \ t \geq 2 \ ,$$

that is

$$2\sigma^* - 2\sigma(t+1) < \sigma^* - \sigma(t) \ ,$$

so that

$$[\sigma^*-\sigma(t+1)]/[\sigma^*-\sigma(t)] < \tfrac{1}{2} \quad \text{for} \quad t \geq 2$$

$\left(\text{since} \quad \sigma^* > \sigma(t) \quad \text{for} \quad t \geq 2\right).$

Thus whenever $\{c(t)\}$ exhibits either Type I or Type II behaviour we have

$$|\sigma(t+1)-\sigma^*|/|\sigma(t)-\sigma^*| < \tfrac{1}{2}$$

for $t \geq 2$. Since $|\sigma(2)-\sigma^*|$ must be less than σ_{max} , this gives

$$|\sigma(t)-\sigma^*| < 2^{2-t}\sigma_{max} \quad \text{for} \quad t \geq 3 .$$

References

[1] L.J. Armour, D.J. Gates and J.A. Rickard, "A convergent optimization algorithm with application to the monopolistic firm", submitted.

[2] D.J. Gates and J.A. Rickard, "Pareto optimum by independent trials", *Bull. Austral. Math. Soc.* 12 (1975), 259-265.

[3] D.J. Gates, J.A. Rickard, and D.J. Wilson, "A convergent adjustment process for firms in competition", *Econometrica* (to appear).

[4] David G. Luenberger, *Introduction to linear and nonlinear programming* (Addison-Wesley, Reading, Massachusetts; Menlo Park, California; London; Don Mills, Ontario; 1973).

Department of Mathematics,
Footscray Institute of Technology,
Footscray, Victoria;

Commonwealth Scientific and Industrial Research Organization,
Division of Mathematics and Statistics,
Canberra, ACT;

Department of Mathematics,
University of Melbourne,
Parkville, Victoria.