Working Papers in
AGRICULTURAL ECONOMICS
No. 88-5

In Search of Optimal Control Models for Generic Commodity Promotion

by

Donald J. Liu
and
Olan D. Forker

Department of Agricultural Economics
New York State College of Agriculture and Life Sciences
A Statutory College of the State University
Cornell University, Ithaca, New York, 14853-7801

<u>TABLE OF CONTENTS</u>

TABLE OF FIGURES

ACKNOWLEDGEMENTS

Donald J. Liu is a Research Associate and Olan D. Forker is a Professor, Department of Agricultural Economics, Cornell University. Useful comments by Jon Conrad, Harry Kaiser, Timothy Mount, and Deborah Streeter are gratefully acknowledged. The authors also wish to acknowledge Shirley Arcangeli for her assistance in preparing this report and Joe Baldwin for drawing the figures.

The research is supported in part by funds provided by the dairy farmers of New York State under the authority of the New York State Milk Promotion Order.

EXECUTIVE SUMMARY

The lingering effect of advertising has led analysts interested in generic commodity promotion to place the issue in a dynamic setting. Thus far, attention has been focused on quantifying the sales-advertising relationship within the context of distributed-lag type econometric models. An important application of sales-advertising models is to simulate future sales under various levels of advertising expenditures in order to identify an optimal spending policy for promotion agencies. However, a serious drawback of this approach is that the truly optimal solution may be missed since it is impractical to exhaust all possible policy scenarios in the simulation. Additional complications are introduced if the interest is in long-term policy, in which a time pattern for the spending must also be selected.

In light of these problems, identifying a more comprehensive optimization framework is highly desirable. In the search for an alternative approach to advertising research, several dynamic optimization techniques are available, including mathematical programming and dynamic programming. In addition, an approach which has had increasing use is optimal control theory. The existing body of literature on applications of control theory to firm advertising could be adapted to generic commodity promotion. To facilitate the adaptation, it would be useful to have a synthesis of the diverse advertising control literature, which has been developed over the past thirty years by economists, operations researchers, and management scientists.

This paper presents a technical survey of three classes of advertising models in the context of control theory: capital theoretic models, sales response models, and diffusion models. Nerlove and Arrow's capital theoretic approach treats advertising as an investment of the firm on goodwill which affects current and future sales. In Vidale and Wolfe's sales response model, advertising is viewed as a means to capture up to some saturation point the uncaptured portion of the market. Gould's diffusion approach to advertising explicitly admits

the interaction between the uncaptured and the captured portions of the market through word-of-mouth.

Not surprisingly, the different classes of models describe different promotional environments and, hence, prescribe quite different promotion strategies. To simplify the issue, two groups of models, each with a corresponding advertising policy prescription, can be identified. In the context of generic commodity promotion, the first group has to do with media advertising while the second group involves word-of-mouth type promotion such as nutritional education programs. In the latter case, promotion is viewed as a scheme that encourages contact between the user and the nonuser, which enhances information spread about the good aspects of the promoted product.

For discussion purposes, it is assumed that there is a desired level of sales which is larger than the current sales level. The task of the promotional unit is to drive sales to the desired level by adopting an optimal advertising spending pattern over time. What is the optimal spending pattern?

In the case of media advertising, the policy prescribed by the model is the following. If there are diminishing sales returns to advertising, it is preferable to reach the goal gradually rather than achieving it immediately because the expense of additional sales increases at an increasing rate as sales expand. Further, the spending pattern is to advertise most heavily in the early periods in order to gain momentum and then to decrease the level of advertising efforts gradually as the sales level approaches its desired state. On the other hand, if sales returns to advertising are believed to be constant, the optimal policy is to advertise the maximum allowable amount right away and achieve the desired sales level as rapidly as possible. In either case, once the desired sales level is attained, the optimal policy is to spend just enough to maintain that level of sales.

In the case of word-of-mouth type promotion, the optimal policy is to begin with a low level of expenditures and build sales up to a "critical level" which is less than the equilibrium level. Once that critical level of sales is reached, the expenditure pattern is to advertise heavily initially but cut back spending gradually until sales approach the desired level. Similar to the case of media advertising, once the desired sales level is attained, the policy is to spend just enough to maintain that level of sales. The reason that the prescribed policy does not begin with a high level of promotion is because it makes little sense to spend a lot of money to enhance contact between the user and nonuser unless the number of users is large enough to effectively spread the information. Thus, it is only optimal to increase the contact rate gradually as the number of users increases over time. Once a certain "critical mass" is reached, the optimal policy is similar to the case of media advertising; spend heavily to gain momentum initially and reduce effort gradually as sales approach the desired level.

The underlying models on which the above policy prescriptions are based are by nature firm oriented, and therefore the strategies outlined must be adapted to generic commodity promotion. To reflect the economic environment in which the commodity promotion program operates, the model must be extended to account for such issues as supply response, government reaction in support policy, and the allocation of funds between primary and secondary markets. As the model becomes more realistic and thus more complex, it is less likely that an analytic solution will be obtainable, in which case obtaining an empirical solution may be the only alternative. Empirical research requires empirical observations, suggesting the need for a continuing effort on quality data collection.

In Search of Optimal Control Models for Generic Commodity Promotion

Donald J. Liu and Olan D. Forker

INTRODUCTION

Generic commodity promotion is not necessarily a zero-sum game in which producers gain and consumers pay. Promotion programs involve high stakes and, if not well conducted, can result in substantial losses to producers regardless of its impact on consumers. The size of the potential losses emphasizes the importance of improving our understanding of the economics of advertising in order to increase the efficiency of promotion efforts.

The continuing effects of advertising on sales after the original period of expenditure is a well recognized phenomenon in the literature (Kinnucan, 1985). The lingering effect of advertising has led analysts interested in generic commodity promotion to place the issue in a dynamic setting. Thus far, attention has been focused on quantifying the sales-advertising relationship within the context of distributed-lag type econometric models (e.g. Kinnucan, 1982; Liu and Forker; Nerlove and Waugh). In such models, an ex-post evaluation of the costs and benefits of promotion program is made by comparing actual sales during a given period with a sales level simulated under the assumption of no advertising effort. The dynamic simulation is based on econometric estimates of the relationship between promotion expenditures and sales.

Another important application of sales-advertising models is to simulate future sales under various levels of advertising expenditures in order to identify an optimal spending policy for promotion agencies. However, a serious drawback of this approach is that the truly optimal solution may be missed since it is impractical to exhaust all possible policy scenarios in the simulation. Additional complications are introduced if the interest is in long-term policy, in which a time path must also be selected. In light of these problems, identifying a more comprehensive optimization framework is highly desirable.

In searching for an alternative approach to advertising research, several dynamic optimization techniques are available, including mathematical programming and dynamic programming. In addition, an approach which has had increasing use is the state-space optimal control theory (Pontryagin, et al.). The existing body of literature on applications of control theory to firm advertising could be adapted to generic commodity promotion. To facilitate the adaptation, it would be useful to have a synthesis of the diverse advertising control literature, which has been developed over the past thirty years by economists, operations researchers, and management scientists.

This paper surveys three classes of advertising models in the context of control theory: capital theoretic models, sales response models, and diffusion models. Both the linear and nonlinear versions of the models are discussed. Nerlove and Arrow's capital theoretic approach treats advertising as an investment of the firm on goodwill which affects current and future demand. In Vidale and Wolfe's sales response model, advertising is viewed as a means to capture up to a saturation point the uncaptured portion of the market potential. Gould's diffusion approach to advertising explicitly admits the interaction between the uncaptured and the captured portions of the market either through inanimate media advertising (Stigler-Gould) or through word-of-mouth (Ozga-Gould).

The overall goal of the paper is to compare the different policy prescriptions which are yielded by each class of models. Not surprisingly, the different classes of models indicate quite different strategies. To understand the source of the differences in policy recommendations and hence their empirical relevancy, various models in each class are presented, along with comments on linkages among models. The discussion also includes observations about the empirical feasibility of the models, as well as differences in data requirements. This research is intended to provide researchers with a review of how control theory has been used to study advertising and to suggest the potential for extending the methodology to generic promotion analysis.

CAPITAL THEORETIC MODELS

The hypothesis that the effect of advertising persists for some period of time after the initial expenditures occur has led a number of economists to treat advertising in a manner analogous to investment in durable goods. An investment in a durable good results in a stock of "production capital" that affects the present and future character of output and, hence, the present and future net revenue of the investing firm. An investment in advertising results in a stock of "demand-generating capital" that affects the present and future demand for the product and, in turn, the present and future net revenue of the firm whose product has been advertised.

Nerlove-Arrow Linear Cost Model

Nerlove and Arrow call this stock of demand-generating capital "goodwill" and view it as a continuous variable which summarizes the effects of current and past advertising outlays on demand. Since consumers tend to forget the advertising messages over time, a dollar spent yesterday should contribute less than a dollar spent today. To capture this phenomenon, Nerlove and Arrow postulate that goodwill, like capital, depreciates over time at a constant proportional rate δ ($0 \leq \delta \leq 1$). Denote A(t) as the level of goodwill stock at time t, A'(t) as the time derivative of A(t), and u(t) as advertising expenditures at time t. Also, let w be the unit price of goodwill (which is assumed to be $1 in Nerlove and Arrow) and $\phi \equiv 1/w$. The evolution of the goodwill for a given level of A at the initial time (t=0) can be characterized by:[1]

(2) $A'(t) = \phi\, u(t) - \delta\, A(t)$, with $A(0)=A_0$.

Equation (2) says that net investment in goodwill (A'(t)) is equal to gross investment in the stock (ϕ u(t)) minus depreciation (δ A(t)). Though goodwill is not observable, consumer disposition such as attitudes toward the advertised product can be surveyed and used as a

1 The argument t of the variables will frequently be suppressed for conciseness. In this paper, the following notational rule will be used throughout. If Q(x,t) is a function of x and t, then Q_x and Q' will be used to denote the derivative of Q with respect to x and time, respectively.

proxy. Accordingly, (2) can be subjected to econometric estimation, providing that the expenditures data {u(t)} are also available. The objective functional is the present value of net revenue streams discounted at a fixed interest rate r:

(1) $$\max_{\{u(t)\}} \int_0^\infty e^{-rt} [\Pi(A(t),Z(t)) - u(t)]\, dt.$$

In the above, $\Pi(A,Z)$ is the profit before advertising expenditures and it follows the usual concavity assumption ($\Pi_A > 0$, $\Pi_{AA} < 0$, for $A \geq 0$), while Z is a vector of exogenous variables. More specifically, $\Pi(A,Z)$ equals $\{P(A,Z)\ Q(A,Z) - C(Q)\}$, where P(A,Z) and Q(A,Z) are the market equilibrium price and quantity for the commodity being promoted and C(Q) is the aggregated costs of production. Note that P(A,Z) and Q(A,Z) can be regarded as the reduced-form equations derived from a corresponding market equilibrium model.[2]

Though not imposed by the original authors, in order to reflect budgetary considerations, it is desirable to impose a maximum level of allowable expenditures u^{max} on gross investment of goodwill:

(3) $$0 \leq u(t) \leq u^{max}.$$

Subject to the evolution of "state variable" A in (2) and the constraints on "control variable" u in (3), the objective is to choose a piecewise continuous advertising policy $\{u(t),\ t\in(0,\infty)\}$ in such a way that (1) is maximized.[3] Since the assumption of a constant unit

---

2 In Nerlove and Arrow's original formulation, the control u is the gross investment on goodwill, rather than advertising expenditures. Thus, their model takes the following form:

(a) $$\max_{\{u(t)\}} \int_0^\infty e^{-rt} [\Pi(A(t),Z(t)) - w\, u(t)]\, dt,$$

(b) $$\text{s.t. } A'(t) = u(t) - \delta A(t), \text{ with } A(0)=A_0.$$

The two formulations are identical in the sense that they both describe the same economic process. However, equation (b) cannot be estimated directly even if an observable proxy is used for goodwill because the gross investment of goodwill u is not observable. Also, Nerlove and Arrow define profit before advertising costs as $R(P,A,Z) = P\ Q(P,A,Z) - C(Q)$, as they are mainly concerned with the case of monopoly. Since the commodity promotion groups are not likely to have the ability to set the price, Nerlove-Arrow's monopolistic formulation is modified here.

3 A characterization of the optimal goodwill investment policy for a corresponding finite planning horizon problem is in Appendix A.

goodwill price makes it a very specialized case, the model will be termed the Nerlove-Arrow Linear Cost Model. The formulation belongs to the class called "linear control problems", as both the objective functional and the constraints are linear in the control variable u. While Nerlove and Arrow use calculus of variations to characterize the optimal goodwill investment policy, a more unified approach based on optimal control theory will be used in this paper. As will be shown, the class of linear control problems has a special characterization of the optimal policy.

For convenience, assume Z(t) is a constant and can be ignored. To analyze the optimal advertising policy, the current-value Hamiltonian can be formed from (1) and (2):

(4) $H = \Pi(A(t)) - u(t) + \psi(t)\ [\phi\ u(t) - \delta\ A(t)].$

In the above, ψ is the current-value adjoint variable and it is assumed to be continuous. The adjoint variable can be interpreted as the shadow price of the state A. Thus, the Hamiltonian can be interpreted as the instantaneous profit which includes the cash flow Π - u and the value $\psi\ A'$ of the new goodwill created by the net investment ϕ u - δ A.

The next step is to account for the control constraints in (3), by forming the current-value Lagrangian with λ_1 and λ_2 as the Lagrangian multipliers associated with the constraints:

(5) $L = H + \lambda_1(t)\ u(t) + \lambda_2(t)\ [u^{max} - u(t)].$

From (5), one obtains the usual "Kuhn-Tucker condition":[4]

(6) $L_u = -1 + \psi\ \phi + \lambda_1 - \lambda_2 = 0,$

$\lambda_1 \geq 0, \qquad \lambda_1\ u = 0,$

$\lambda_2 \geq 0, \qquad \lambda_2\ (u^{max} - u) = 0.$

In accordance with Pontryagin's maximum principle (e.g. Clark; Kamien and Schwartz), additional necessary conditions are: (i) the "adjoint equation", $\psi' = r\psi - H_A$, (ii) the "state equation" in (2), (iii) the "transversality condition" $\psi(t)\ e^{-rt} = 0$ as $t\to\infty$ and, (iv) $H_{uu} \leq 0$. In view of (4), condition (i) says that capital gain (ψ') of the state and its marginal contribution to the cash flow (net of the opportunity

4 In the case where the budget constraint (3) does not appear in the model, the Kuhn-Tucker condition will be replaced by the "optimality condition" $H_u = 0$.

costs on depreciation) must equal the time costs ($r\psi$) of holding the capital. Condition (iii) specifies the endgame value of the state evaluated at the present-value form. Finally, condition (iv) is to ensure maximization, rather than minimization, and it is automatically satisfied because of the concavity assumption on $\Pi(A)$.[5]

By inspection, the Kuhn-Tucker is equivalent to:

$$(7) \quad u^*(t) = \begin{cases} 0 \\ u^s \in [0, u^{max}] \\ u^{max} \end{cases} \quad \text{if} \quad \begin{cases} \sigma(t) & < 0 \\ \sigma(t) \equiv -1 + \psi(t)\ \phi & = 0 \\ \sigma(t) & > 0. \end{cases}$$

In (7), $\sigma(t)$ is the coefficient of u in H (i.e. $\sigma = H_u$) and for the obvious reason it is called the switching function. The control scheme of (7) is called "bang bang"; u is at its minimum level (i.e. zero) while the shadow price of goodwill (ψ) is less than its cost ($w \equiv 1/\phi$) and is at its maximum level (i.e. u^{max}) when the shadow price is above the cost. However, interior control u^s could be feasible on a path along which the marginal condition $\sigma(t)=0$ holds. Such a path is known as a "singular" path which earns its name from the condition $H_{uu}=0$ and which is signified by the superscript s.

To determine the singular solution, note that the adjoint equation $\psi' = r\ \psi - H_A$ is:

$$(8) \quad \psi' = (r + \delta)\ \psi - \Pi_A.$$

Again, equation (8) says that the capital gain (ψ') and the marginal contribution of the capital (Π_A) must equal the opportunity costs of holding the capital which include time costs ($r\psi$) and depreciation costs ($\delta\psi$). From (7), $\psi(t) = 1/\phi$ and, hence, $\psi' = 0$ along the singular path. Given these, (8) becomes $(r + \delta)/\phi = \Pi_A(A^s)$, which implicitly specifies the singular stock level A^s.

By the maximum principle, the optimal control is that policy for which the singular goodwill stock A^s be attained as rapidly as possible and maintained thereafter (Clark, p.53). In other words, the optimal path is the one which lies as close as possible to the singular path.

5 Since the integrand in (1) is jointly concave in u and A and the state equation (2) is linear in u and A, the necessary conditions are also sufficient for optimality (Kamien and Schwartz, pp. 122-123). In the latter models to be discussed, the state equation is concave in u and A. In such a case, sufficiency of the necessary conditions also requires $\psi(t) \geq 0$ for all t.

Thus, as shown in Figure 1, if the initial stock of goodwill is $A_0{}^{I} < A^{s}$ (the case where $\sigma(t) > 0$), it is optimal to apply $u^{*}(t) = u^{max}$ until t^{I} at where A^{s} is attained. On the other hand, if the A(0) is $A_0{}^{II} > A^{s}$ (the case where $\sigma(t) < 0$), the optimal policy is $u^{*}(t) = 0$ until t^{II} at where goodwill depreciates to A^{s}. Once A^{s} is reached, apply the singular control $u^{s} = \delta A^{s}/\phi$ so as to cover the depreciation and ride the "golden path".[6]

Thus, the optimal advertising policy for the Nerlove-Arrow linear cost model is characterized by a bang-bang control followed by a singular control with the property that the singular path is to be approached as rapidly as possible. The attainment of the singular path is desirable because it describes the equilibrium goodwill while the most rapid approach strategy is due to the assumption that the cost of adding to goodwill is constant (w). This policy characterization is referred to as the "turnpike property" in which an automobile driver is to approach the turnpike as fast as possible while obeying the speed limit of the local highway.[7]

<u>Nerlove-Arrow Nonlinear Cost Model</u>

The assumption of constant cost of adding to goodwill in the previous section is likely to be unrealistic and should be regarded as a special case. This can be made clear by considering the cost function as the dual from a production function that utilizes media and agency services to produce goodwill. Then the cost function should be convex

---

6 The time of switching (t^{I} or t^{II}) can be determined from (2). For example, in the case where $A_0 < A^{s}$, the policy is $u^{*}(t) = u^{max}$ until t^{I} at where A^{s} is attained. Thus, (2) becomes $A' = \phi\, u^{max} - \delta\, A(t)$, for $t \in [0, t^{I}]$. Given A_0 and A^{s}, it holds that $t^{I} = (1/\delta)\ \ln\{(A_0\delta - u^{max}\phi)/(A^{s}\delta - u^{max}\phi)\}$. On the other hand, if $A_0 > A^{s}$, $u^{*}(t) = 0$ until t^{II} at where A^{s} is attained. Thus, (2) becomes $A' = -\,\delta\, A(t)$, for $t \in [0, t^{II}]$, implying $t^{II} = (1/\delta)\ln(A_0/A^{s})$.

7 Note that in the special case where $u^{max} = \infty$ and $A_0 < A^{s}$, the most rapid approach is to jump instantaneously to A^{s} by applying an "impulse control" at t=0 of an appropriate magnitude and then follow the turnpike for $t \geq 0$. This is the case considered in Nerlove and Arrow. On the other hand, the control constraint u^{max} may be so small that the singular path is not attainable. For discussion of the optimal policy on this latter case, see Appendix B.

if the production function has the usual concavity. In fact, this is the underlying motivation for Gould to replace the linear cost function of Nerlove-Arrow with a nonlinear cost function. The result is that the optimal goodwill investment policy is no longer of the bang-bang-singular type.

Denote the nonlinear cost function as W(g) where g is gross investment of goodwill and W is convex in g ($W_g > 0$, $W_{gg} > 0$, for $g \geq 0$). Thus, the gross investment g for a given level of advertising expenditures u is $W^{-1}(u)$ which is concave in u. Denote $W^{-1}(u)$ as $\Phi(u)$. The model becomes:

(1) $\max\limits_{\{u(t)\}} \int_0^\infty e^{-rt} [\Pi(A(t),Z) - u(t)]\, dt,$

(9) s.t. $A'(t) = \Phi(u(t)) - \delta\, A(t)$, with $A(0)=A_0$.

The control constraint (2) is ignored for simplicity.[8] Also, it is assumed that Z(t) is constant over time as it is convenient to consider the long-term equilibrium solution for a nonlinear model. With a specific assumption on the functional form of Φ, equation (9) can be estimated econometrically. The current-value Hamiltonian from (1) and (9) is

(10) $H = \Pi(A(t)) - u(t) + \psi(t)\, [\Phi(u(t)) - \delta\, A(t)].$

The optimality condition $H_u=0$ is

(11) $1 = \psi\, \Phi_u.$

In view of the duality between Φ and the cost function W, (11) is essentially the marginal condition found in (7), although the marginal cost of goodwill is not a constant here.[9] The adjoint equation $\psi' = r\,\psi - H_A$ is the same as before:

(8) $\psi' = (r + \delta)\, \psi - \Pi_A.$

Differentiating (11) with respect to time yields $\psi'\, \Phi_u + \psi\, \Phi_{uu}\, u' = 0$. Substituting into this expression (11) for ψ and (8) for ψ' yields

(12a) $u' = [-1/\Phi_{uu}]\, [(r + \delta)\, \Phi_u - \Pi_A\, \Phi_u^2].$

Also, the state equation (9) has to be satisfied and it is reproduced:

(12b) $A' = \Phi(u) - \delta\, A.$

---

8 Alternatively, one can regard the control as constrained by the convexity of the cost function.

9 From (11), one notes that $\psi(t)$ is positive for all t. In view of footnote 5, this means the necessary conditions are also sufficient.

Equations (12a) and (12b) consist of a plane-autonomous system of nonlinear ordinary differential equations. From the theory of ordinary differential equations, the system represented by (12) possesses a unique solution (u(t),A(t)) passing through any given initial point (u(0),A(0)). The solutions form a family of trajectories, one through each point of the (u,A) plane. The entire collection of these trajectories is referred to as the "phase diagram" of the system represented by (12). The geometrical analysis of these trajectories is facilitated by considering the isoclines in the phase diagram; the loci on which $u'=0$ and $A'=0$. The directional arrows in Figure 2 indicate the movements of the trajectories in each of the four isosectors determined by the two isoclines.[10]

The intersection of these two isoclines is the equilibrium point (u^*,A^*) for the system as both u and A are stationary at that point (Figure 3). Given the directional arrows appearing in the figure, it is apparent that the equilibrium point is a saddle point for the differential system in (12).[11] Associated with the saddle point are two

---

10 From (12a), the isocline for u is given by $(r + \delta) = \Pi_A \Phi_u$. Totally differentiating the isocline with respect to u and A yields $du/dA|_{u'=0} = - \Pi_{AA} \Phi_u /(\Phi_{uu} \Pi_A) < 0$, indicating that $u'=0$ locus has a negative slope. Moreover, from (12a) $\partial u'/\partial u = \Pi_A \Phi_u > 0$. Hence, at points to the right of the u-isocline, u' is positive (i.e. u is increasing). Similarly, for points to the left of the locus, u' is negative (i.e. u is decreasing). From (12b), the isocline for A is given by $A = \Phi(u) / \delta$. Hence, $\partial A/\partial u|_{A'=0} = \Phi_u / \delta > 0$ and $\partial^2 A/\partial u^2|_{A'=0} = \Phi_{uu} / \delta < 0$, indicating the locus is positively sloped and convex in u-A plane. Further, $\partial A'/\partial A = -\delta < 0$ indicates that at points to the right of $A'=0$ locus, $A' < 0$ (i.e. A is decreasing) and at points to the left of the locus, $A' > 0$ (i.e. A is increasing).

11 The point (u^*,A^*) will be a saddle if the eigenvalues of the characteristic equation of the system (12) are real and opposite in sign (e.g. Kamien and Schwartz, pp. 306-311). In verifying this, linearize (12) around (u^*,A^*) to obtain

(13a) $u' = \beta_{11} (u - u^*) + \beta_{12} (A - A^*)$

(13b) $A' = \beta_{21} (u - u^*) + \beta_{22} (A - A^*)$,

where:

$$\beta_{11} \equiv \partial u'/\partial u = \Pi_A(A^*) \Phi_u(u^*) > 0$$

$$\beta_{12} \equiv \partial u'/\partial A = \Pi_{AA}(A^*) \Phi_u(u^*)^2 / \Phi_{uu}(u^*) > 0$$

unique trajectories (sketched in heavy lines), called separatrices, that converge toward the saddle point as $t \to \infty$. All other paths ultimately lead either to an infinitely large amount of (u,A) or to a zero level. The case where the (u,A) pair diverges to infinity is obviously undesirable, and the case of zero goodwill stock is also nonoptimal (as shown in Gould).

Accordingly, the separatrices define an unique optimal control policy for goodwill investment. In principle, the separatrices can be approximated by numerical computation. In addition, qualitative insights into the optimal policy can be gained by inspecting the separatrices in the phase diagram. To see this, note that the left separatrix depicts the optimal advertising time path as goodwill stock approaches A^* from below (i.e. from the left) while the right separatrix depicts the time path of the policy as the stock approaches A^* from above (i.e. from the right).

Thus, in the case where the initial goodwill stock level is below the long-term equilibrium level, the optimal policy is to advertise heavily in the early periods (in order to catch up) and then to decrease the level of advertising efforts as the stock approaches its long-term equilibrium A^* from below. On the other hand, if the initial stock is above the equilibrium level, the optimal policy is characterized by a low level of advertising effort in the initial periods (in order to ease up) and increasing efforts over time as the stock approaches A^* from above. In any case, the optimal approach path is not the most rapid approach, but rather a more gradual approach engendered by the convexity of goodwill cost function.

$$\beta_{21} \equiv \partial A'/\partial u = \Phi_u(u^*) \qquad > 0$$

$$\beta_{22} \equiv \partial A'/\partial A = -\delta \qquad < 0.$$

The characteristic equation of (13) is

$$k^2 - k\,(\beta_{11} + \beta_{22}) + (\beta_{11}\,\beta_{22} - \beta_{12}\,\beta_{21}) = 0,$$

which has roots

$$k_1, k_2 = 1/2\,(\beta_{11}+\beta_{22}) \pm 1/2\,[(\beta_{11}+\beta_{22})^2 - 4\,(\beta_{11}\beta_{22}-\beta_{12}\beta_{21})]^{1/2},$$

and they are real and opposite in sign if

(14) $\beta_{11}\,\beta_{22} - \beta_{12}\,\beta_{21} < 0.$

In view of the signs of β_{ij} given above, condition (14) is satisfied. Hence, the roots are real and opposite in sign verifying (u^*,A^*) is a saddle point.

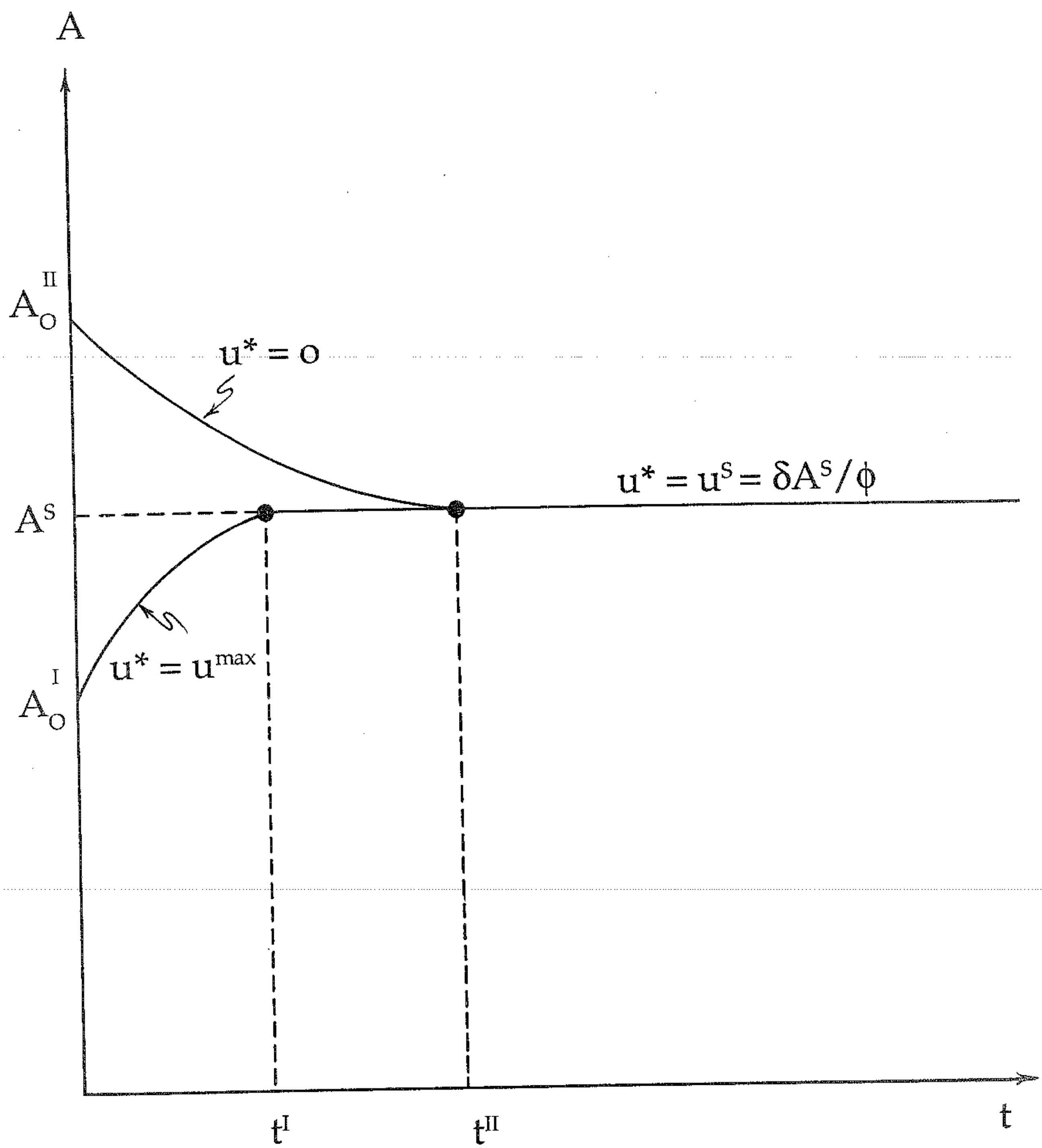

Figure 1: Bang-Bang-Singular Control

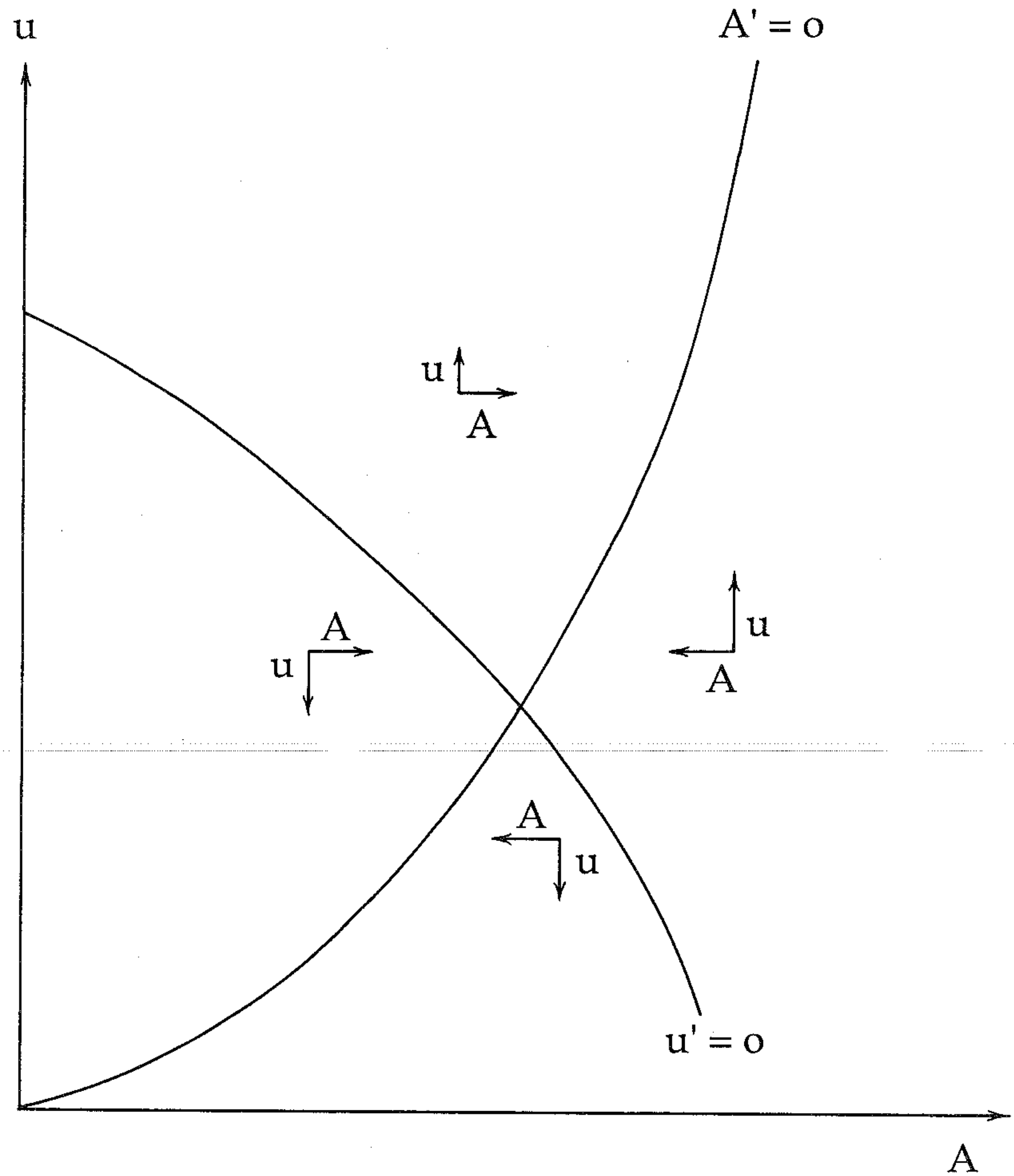

Figure 2: Isoclines (Nerlove-Arrow Nonlinear Cost Model)

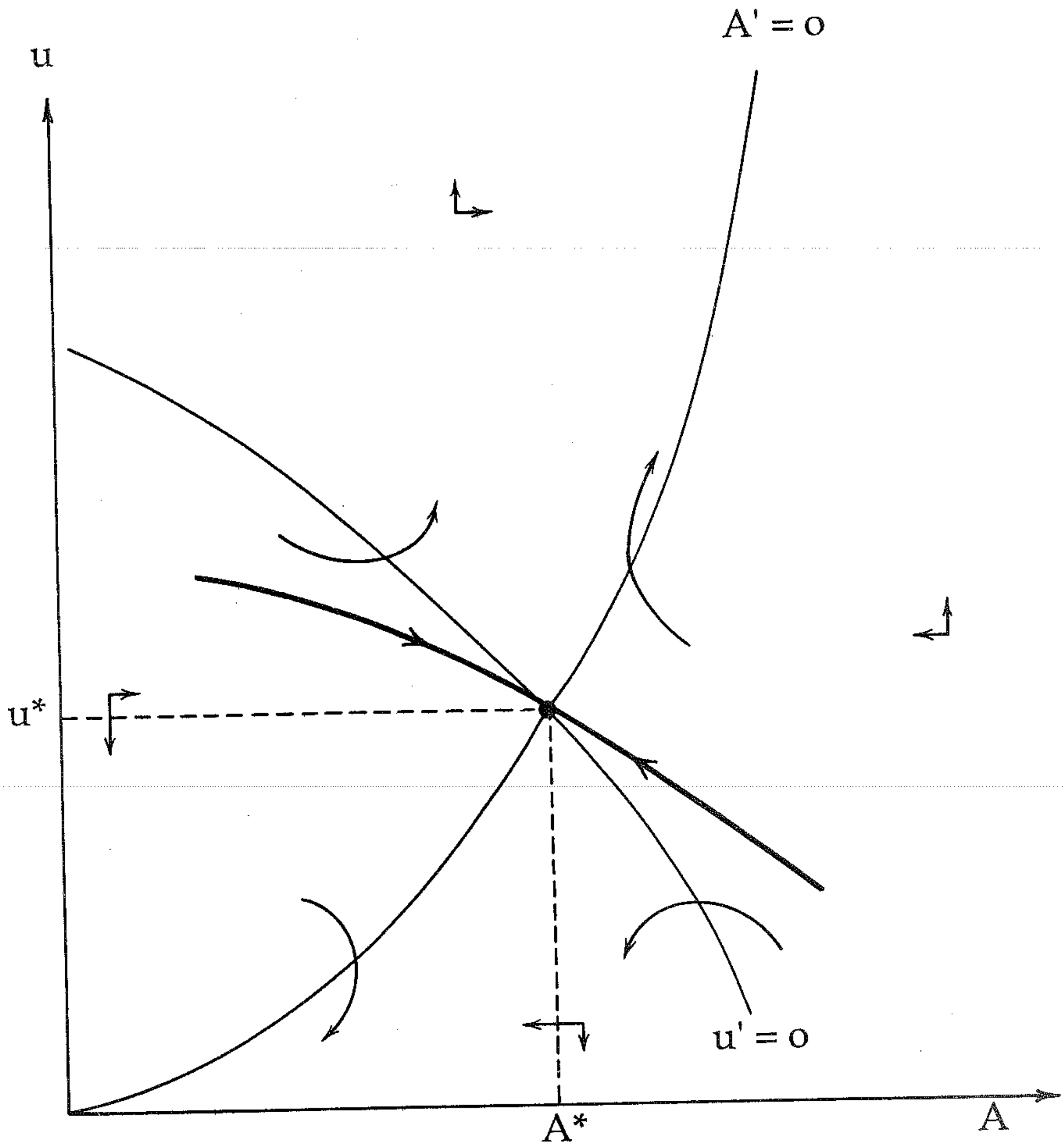

Figure 3: Phase Diagram (Nerlove-Arrow Nonlinear Cost Model)

SALES RESPONSE MODELS

One major empirical problem associated with the capital theoretic models discussed in the previous section is the difficulty in observing or measuring goodwill. Even if consumer disposition measures such as attitude toward the advertised product can be used as proxies, those measures are expensive to obtain, and questionable in terms of precision. As an alternative, the state variable A might be interpreted as sales, rather than the fuzzy concept of goodwill. In this case, ϕ in (2) should be interpreted as the sales response coefficient and Φ in (9) as the sales response function of advertising expenditures. Thus, the linear model assumes a constant sales response while the nonlinear model admits a diminishing sales response arising from the increasing marginal cost of promoting additional sales.

The phenomenon of diminishing marginal sales returns to advertising can also arise if there is an upper bound on the market potential or saturation.

Vidale-Wolfe Sales Response Model

Based on several experimental studies of advertising sales effectiveness, Vidale and Wolfe argue that changes in the rate of sales of a product depend not only on whether new users are attracted by advertising, but also on how much erosion of the advertising effect there is among current users. Specifically, in their model, the response coefficient ρ measures the effect of advertising expenditures on the uncaptured portion of the market, while the decay coefficient δ reflects the rate at which the captured markets "forgets" the previous advertising effort. In addition, they found that the impact of advertising expenditures on nonusers diminishes as the sales rate approaches a market saturation point K. Based on their observations, Vidale and Wolfe propose a sales equation of the following:

(15) $A'(t) = \rho\, u(t)\, (1 - A(t)/K) - \delta\, A(t)$, with $A(0)=A_0$.

The introduction of a finite saturation level, which is absent in previous models, results in diminishing returns to advertising in the Vidale-Wolfe model. That is, as larger and larger numbers of new users are captured, the uncaptured portion of the market shrinks, leading to a diminishing effect of advertising expenditures on sales. Note that as $K \to \infty$, the saturation phenomenon disappears, reducing (15) to (2). Defining $\phi \equiv \rho/K$, the state equation (15) can be expressed more concisely as

(16) $A'(t) = \phi\, u(t)\, (K - A(t)) - \delta\, A(t)$,

with $A(0)=A_0$, and $0 \leq \phi u \leq 1$. In estimating (16), K can either be specified based on *a priori* knowledge or be estimated simultaneously with other coefficients (ϕ and δ) in the model. The optimal advertising policy can be analyzed by maximizing the objective functional (1), subject to (16) and the control constraints (3). Since both the objective functional and the state equation are linear in control u, the problem can be treated as a linear control problem which yields a bang-bang-singular optimal policy.[12]

Extended Vidale-Wolfe Sales Response Model

The diminishing marginal sales returns to advertising aspect is an important advantage of the Vidale-Wolfe model over the Nerlove-Arrow linear cost model. Similar to that in Nerlove-Arrow, the response coefficient ϕ [or ρ if one uses (15)] is a constant for any level of advertising expenditures in Vidale-Wolfe. The diminishing returns is purely market-size oriented and it arises from the specification of a market saturation. The response coefficient, however, need not be a constant. In fact, it should decrease as the level of expenditures increases if the response is regarded as a result of a concave production process that utilizes media and agency services. The Vidale-Wolfe model can be extended to allow for both the market-size and production oriented diminishing sales returns to advertising:

(17) $A'(t) = \Phi(u(t))\, (K - A(t)) - \delta\, A(t)$, with $A(0)=A_0$,

where $\Phi(u)$ is the response function $(0 \leq \Phi(u) \leq 1)$ and is concave ($\Phi_u > 0$, $\Phi_{uu} < 0$, for $u \geq 0$). With a functional specification of $\Phi(u)$, (17) can be estimated econometrically. The objective is to maximize (1) subject to (17). The current-value Hamiltonian for the problem is:

(18) $H = \Pi(A(t)) - u(t) + \psi(t)\, [\Phi(u(t))\, (K - A(t)) - \delta\, A(t)]$.

12 With a more detailed specification of $\Pi(A)$, however, one can obtain a more detailed solution and, hence, better insight into the problem. In Appendix C, $\Pi(A)$ is assumed to be a linear function of A and a closed-form optimal policy is derived.

From (18), one obtains the optimality condition and the adjoint equation:

(19) $1 = \psi \Phi_u (K - A)$,

(20) $\psi' = (r + \delta + \Phi) \psi - \Pi_A$.

The interpretation for (19) is the same as before. The optimal advertising policy is selected so that the last dollar expended is exactly equal to the shadow price of the sales times the additional sales due to that last dollar.[13] Equation (20) indicates that the change in the imputed value of the state (i.e. capital gain) and its marginal contribution to the cash flow must equal the marginal opportunity costs of having the state at that level. Here the opportunity costs include not only the time costs and the depreciation costs but also costs of the diminished marginal advertising sales effectiveness ($\Phi\psi$) which arises as the saturation point is approached. Obviously, there is a tradeoff; advertising increases sales, but the added sales depress the effectiveness of future advertising.

Differentiating (19) with respect to time yields $\psi' \Phi_u (K-A) + \psi \Phi_{uu} u' (K-A) - \psi \Phi_u A' = 0$. Substituting A' from (17), ψ from (19) and ψ' from (20) into the above expression to obtain

(21a) $u' = [-1/\Phi_{uu}] \{[r + \delta K / (K - A)] \Phi_u - \Pi_A \Phi_u^2 (K - A)\}$.

Also, reproduce the state equation (17) here:

(21b) $A' = \Phi(u) (K - A) - \delta A$.

Equations (21a) and (21b) consist of a plane autonomous system. The directional arrows in Figure 4 indicate the movements of the trajectories in each of the four isosectors determined by the u-isocline and the A-isocline.[14] The intersection of these two isoclines is the

---

13 From (19), one notes that $\psi(t)$ is positive for all t. In view of footnote 5, this means the necessary conditions are also sufficient.

14 From (21a), the u-isocline is determined by those (u,A) for which:

(22a) $[r + \delta K / (K - A)] = \Pi_A \Phi_u (K - A)$.

Total differentiate (22a) yields:

$$du/dA|_{u'(t)=0} = \frac{[\delta K/(K-A)^2] - \Pi_{AA} \Phi_u (K-A) + \Pi_A \Phi_u}{\Pi_A \Phi_{uu} (K-A)} < 0.$$

Thus, the $u'=0$ locus is negatively sloped. Moreover, from (21a), $\partial u'/\partial u|_{u'=0} = \Pi_A \Phi_u (K-A) > 0$. Thus, $u' > 0$ for points above the u-isocline and $u' < 0$ for points below the locus.

equilibrium point (u^*, A^*) for the system (Figure 5). Again, given the directional arrows, it appears that the equilibrium point is a saddle point for the differential system in (21).[15]

As indicated by the separatrices in the phase diagram, the general qualitative property of the optimal advertising policy is similar to that found for the Nerlove-Arrow nonlinear model (c.f. Figure 3). Specifically, for $A_0 < A^*$ the optimal policy is to advertise most heavily at the start of the campaign and continually decrease advertising expenditures as A approaches A^* (see left separatrix). On the other hand, if the initial state is above the equilibrium level, the optimal policy is characterized by a low level of advertising efforts in the initial periods and increasing efforts over time as the stock approaches A^* from above (see right separatrix).

From (21b), the isocline for A is given by:

(22b) $A = K\,\Phi(u) \,/\, (\delta + \Phi(u))$.

The curve is upwardly sloped and convex in (u-A) plane because $\partial A/\partial u|_{A'=0} = \delta K\Phi_u \,/\, [\delta+\Phi]^2 > 0$ and $\partial^2 A/\partial u^2|_{A'=0} = [\delta K\Phi_{uu}\ (\delta+\Phi) - 2\delta K\Phi_u{}^2] \,/\, [\delta+\Phi]^3 < 0$. Note from (22b) that u increases without bound along the A-isocline as A approaches K. Further, from (21b), $\partial A'/\partial u = \Phi_u\ (K - A) > 0$ indicates that $A' > 0$ for those points above the A-isocline and $A' < 0$ for points below the locus.

15 To verify the saddle point conjecture, linearizing (21) around (u^*, A^*) and expressing the result in the format of (13), one finds

$$\beta_{11} \equiv \partial u'/\partial u = \Pi_A(A^*)\ \Phi_u(u^*)\ (K - A^*) \qquad > 0$$

$$\beta_{12} \equiv \partial u'/\partial A = [-1/\Phi_{uu}(u^*)]\ [\delta K\Phi_u(u^*)\ /\ (K-A^*)^2 - \Pi_{AA}(A^*)\ \Phi_u(u^*)^2\ (K-A^*) + \Pi_A(A^*)\ \Phi_u(u^*)^2] \qquad > 0$$

$$\beta_{21} \equiv \partial A'/\partial u = \Phi_u(u^*)\ (K - A^*) \qquad > 0$$

$$\beta_{22} \equiv \partial A'/\partial A = -\ (\delta + \Phi(u^*)) \qquad < 0.$$

In view of the signs for β_{ij}, the saddle point condition (14) is obviously satisfied.

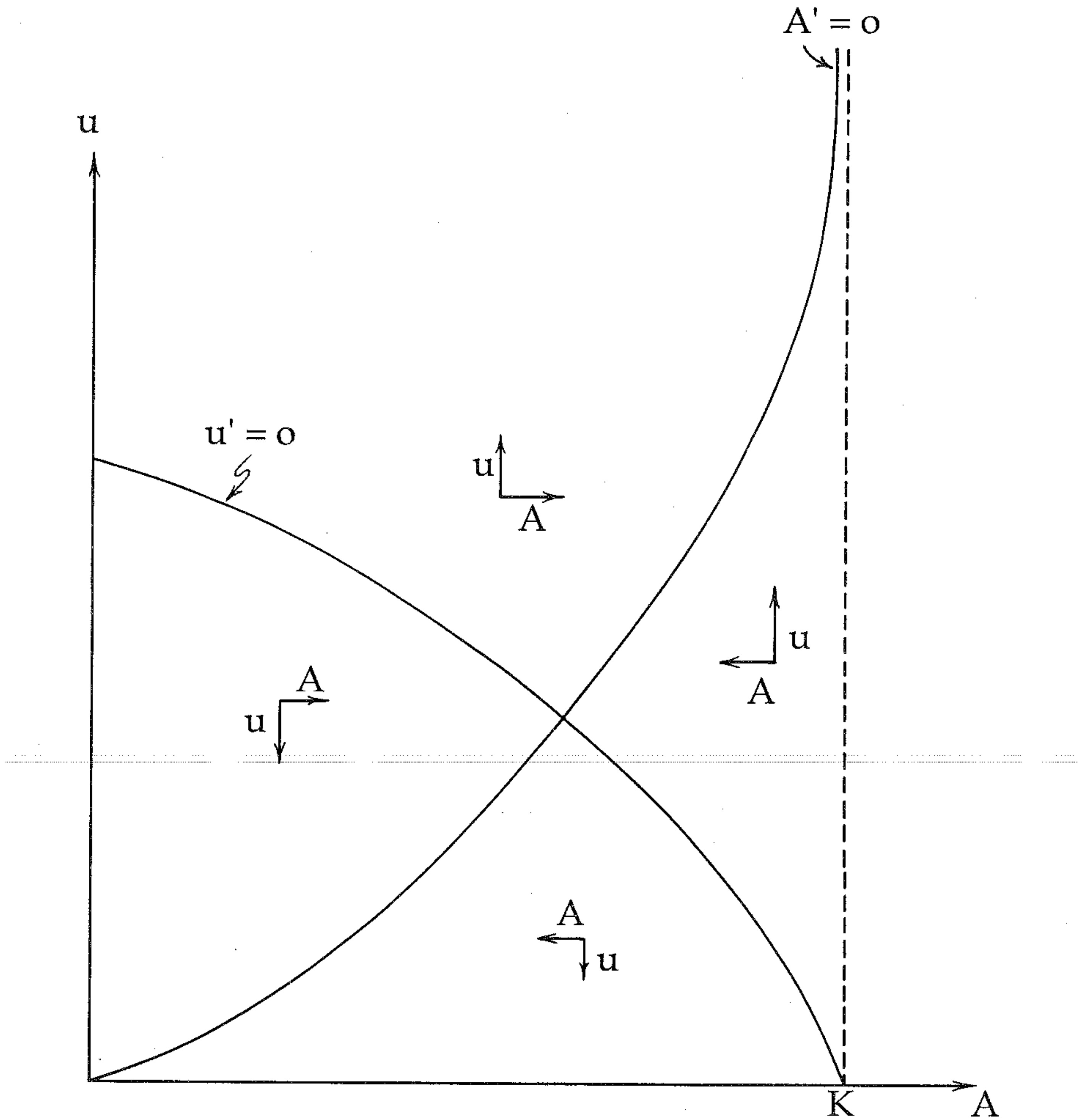

Figure 4: Isoclines (Extended Vidale-Wolfe Sales Response Model)

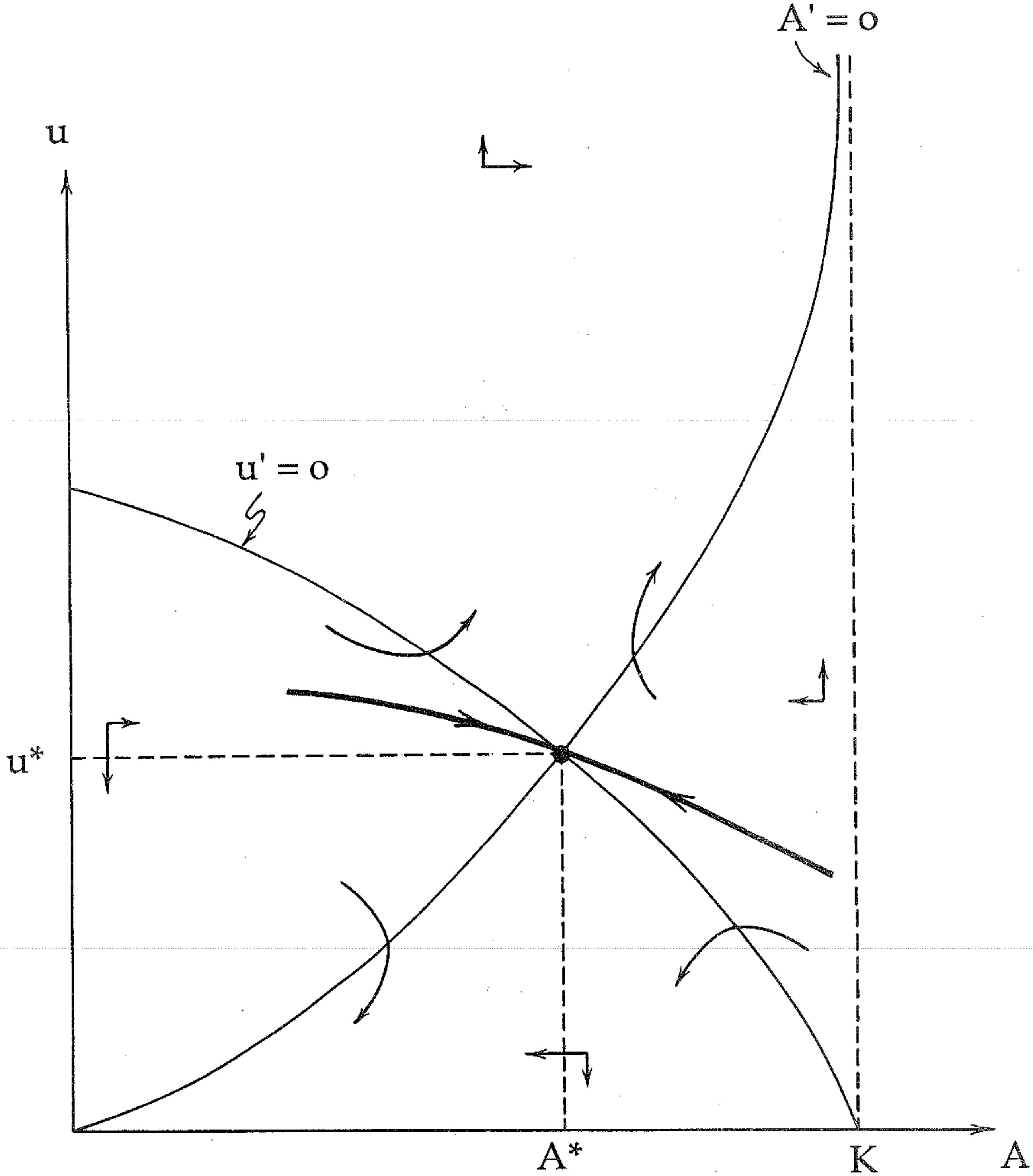

Figure 5: Phase Diagram (Extended Vidale-Wolfe Sales Response Model)

DIFFUSION MODELS

In the preceding models, advertising is assumed to have an impact on sales over time but the specific mechanisms leading to the dynamics are left unspecified. In a diffusion context, Gould argues that individuals learn of the advertisement by coming into contact with either the advertising medium or other people who are aware of the advertisement. Since different types of informational contact may lead to different dynamics between the state and the control, it appears important to investigate various diffusion processes.

Stigler-Gould Diffusion Process

Gould's first diffusion model involves the spread of information as a result of contact with media advertising. Based on Stigler's price diffusion process, Gould proposes a state equation which is identical to (17).

(17) $A'(t) = \Phi(u(t))\ (K - A(t)) - \delta\ A(t)$, with $A(0)=A_0$.

Alternatively, within some data range, a linear version of the model may be preferred:

(16) $A'(t) = \phi\ u(t)\ (K - A(t)) - \delta\ A(t)$, with $A(0)=A_0$.

The state variable A in this case is defined as the number of individuals who are aware of the information contained in the advertisement and K is the total population. Accordingly, $\Phi(u)$ (or ϕ), which can be called the reach function (or reach coefficient), characterizes the extent of the targeted population (K-A) influenced by a given level of advertising expenditure u. In practice, data on product awareness can be obtained through market research and both (16) and (17) can be estimated econometrically. The objective is to maximize (1) subject to (16) or (17), and thus is essentially the same problem discussed in the previous section.

Ozga-Gould Diffusion Process

In the Stigler-Gould diffusion model, the spread of information occurs strictly as a result of contact with media advertising. An alternative formulation is that potential consumers, who are not aware of the product or its characteristics, acquire information from those who are informed. The existing informed population acts as agents of the seller in disseminating information about the product either through demonstration or through word of mouth (similar to the spread of disease!). Such processes for information dissemination are particularly important in studying commodity promotion, since a significant portion of the available promotion funds are spent on education efforts such as the nutritional education program.

Following Ozga, Gould's second diffusion approach to advertising is specified as:

(23) $A'(t) = \phi\ u(t)\ A(t)\ (K - A(t)) - \delta\ A(t)$, with $A(0)=A_0$,

where A is the number of individuals who are aware of the product, and K is the total population. The model differs from Stigler-Gould diffusion in that (23) allows a logistic type interaction between the informed (A) and the uninformed (K-A). The extent of this interaction is captured in the contact coefficient ϕ $(0 \leq \phi u \leq 1)$. Subject to (23), the objective is to maximize (1) by selecting the optimal expenditures (u) on such contact-enhancing activities as nutrition education which are conducted through doctors' offices, school superintendents, and household heads. This is a linear control problem as both the objective functional and the state equations are linear u.

Alternatively, consider a nonlinear version of (23):

(24) $A'(t) = \Phi(u(t))\ A(t)\ (K - A(t)) - \delta\ A(t)$, with $A(0)=A_0$.

In the above, $\Phi(u)$ is the contact function $(0 \leq \Phi(u) \leq 1)$ which characterizes the extent of contact between the informed and the uninformed. Since social and informational contacts occur with greater frequency within social networks than across them, expenditures on contact enhancing activity is assumed to be a convex function of the contact rate. Accordingly, taking the inverse of the expenditures function, the contact function $\Phi(u)$ is concave. The objective is to maximize (1) subject to (24). The current-value Hamiltonian for the problem is:

(25) $H = \Pi(A(t)) - u(t) + \psi(t)\ [\Phi(u(t))\ A(t)\ (K - A(t)) - \delta\ A(t)]$.

From (25), one obtains the optimality condition and the adjoint equation:

(26) $1 = \psi\ \Phi_u\ A\ (K - A)$,

(27) $\psi' = [r + \delta + \Phi\ A - \Phi\ (K - A)]\ \psi - \Pi_A$.

Equation (26) indicates that the optimal policy should be such that the last dollar expended is exactly equal to the shadow price of the stock of informed population times the addition to the stock due to that last dollar.[16] Equation (27) says that the change in the imputed value of the informed population and its marginal contribution to the cash flow must equal the marginal opportunity costs of having the informed population at that level. Here the opportunity costs are associated with time, depreciation, and the diminishing returns which occur as saturation is approached (i.e. $\Phi A \psi$), net of the benefits arising from increased word-of-mouth effects as the stock of informed population expands (i.e. $-\Phi(K-A)\psi$). Thus, the tradeoff becomes more complicated in the case where the informed population is the source of information. On the one hand, it is desirable to increase the informed population because it enhances both sales and the spread of information. On the other hand, the effectiveness of the promotion expenditures is diminished as the informed population is enlarged.

Differentiating (26) with respect to time yields $\psi' \Phi_u A(K-A) + \psi \Phi_{uu} u' A(K-A) + \psi \Phi_u A'(K-A) - \psi \Phi_u A A' = 0$. Substituting into the above expression A' from (24), ψ from (26), and ψ' from (27), yields:

(28a) $u' = [-1/\Phi_{uu}] \{[r + \delta A / (K - A)] \Phi_u - \Pi_A \Phi_u^2 A (K - A) \}$

Also, reproduce the state equation (24) here:

(28b) $A' = \Phi(u) A (K - A) - \delta A.$

Equations (28a) and (28b) consist of a plane autonomous system. As derived in Appendix D and shown in Figure 6, the u-isocline has a quadratic shape and there are two A-isoclines including the u-axis. The directional arrows in the figure indicate the movements of the trajectories in each of the five isosectors determined by the u-isocline and the A-isocline. Since the u axis is also an isocline for A, there is no way of achieving any positive level of A regardless of the expenditure level if A_0 is zero. That is, since social contact between the informed and the uninformed is the only media for information spread, a zero level of the informed population means a lacking of the media. In this case, as pointed out by Glaister in a related paper, high-pressure promotion campaigns such as "free sample", "introductory

16 From (26), one notes that $\psi(t)$ is positive for all t. In view of footnote 5, this means the necessary conditions are also sufficient.

offer", and "enlisting the help of retailers" must be conducted to establish a base level of users.

There are two stationary points, L and H, in the phase diagram (Figure 7). The high equilibrium H is a saddle point while the low equilibrium L is unstable. If the initial informed population is large enough, the optimal advertising (contact enhancing) policy is the one characterized by the separatrices in the figure. From the left separatrix, the optimal path differs significantly from that found in previous nonlinear models. In the earlier models, the optimum path of advertising expenditures always requires the heaviest outlays in the early periods with continuous reductions in expenditures as the state approaches its equilibrium from below. In contrast, the optimum path for the present model begins with a low level of expenditures, building up to a maximum of $u^{\#}$, which is greater than the equilibrium level u^*, and then cut back toward u^* as A approaches A^*. However, similar to the previous models, in the case where the initial state A_0 is greater than A^*, the optimum path for u is monotonically increasing and does not reverse as A approaches A^* from above.

The rationale for the policy where $A_0 > A^*$ is as before while for $A_0 < A^*$ the reasoning is as follows. It makes little sense to spend a lot of funds to enhance the contact rate if there is not an existing informed population large enough to serve as the agent of information spread. Thus, it is only optimal to increase the contact rate gradually as the informed population increases over time. Once the "critical mass" $A^{\#}$ is reached, the optimal policy is to spend heavily initially and reduce effort gradually as A approaches A^* from below; a case similar to the previous two nonlinear models. In other words, since the left separatrix must be extended into the isosector where u is decreasing and A is increasing in order to approach the saddle point, u must over-shoot first. Once the separatrix reaches the isosector just mentioned, the control u starts to decrease as A approaches A^* from below. One implication of this result is that it may be profitable to conduct some high pressure promotion campaigns at the initial phase so as to enlarge the informed population and, hence, hasten the attainment of the steady state.

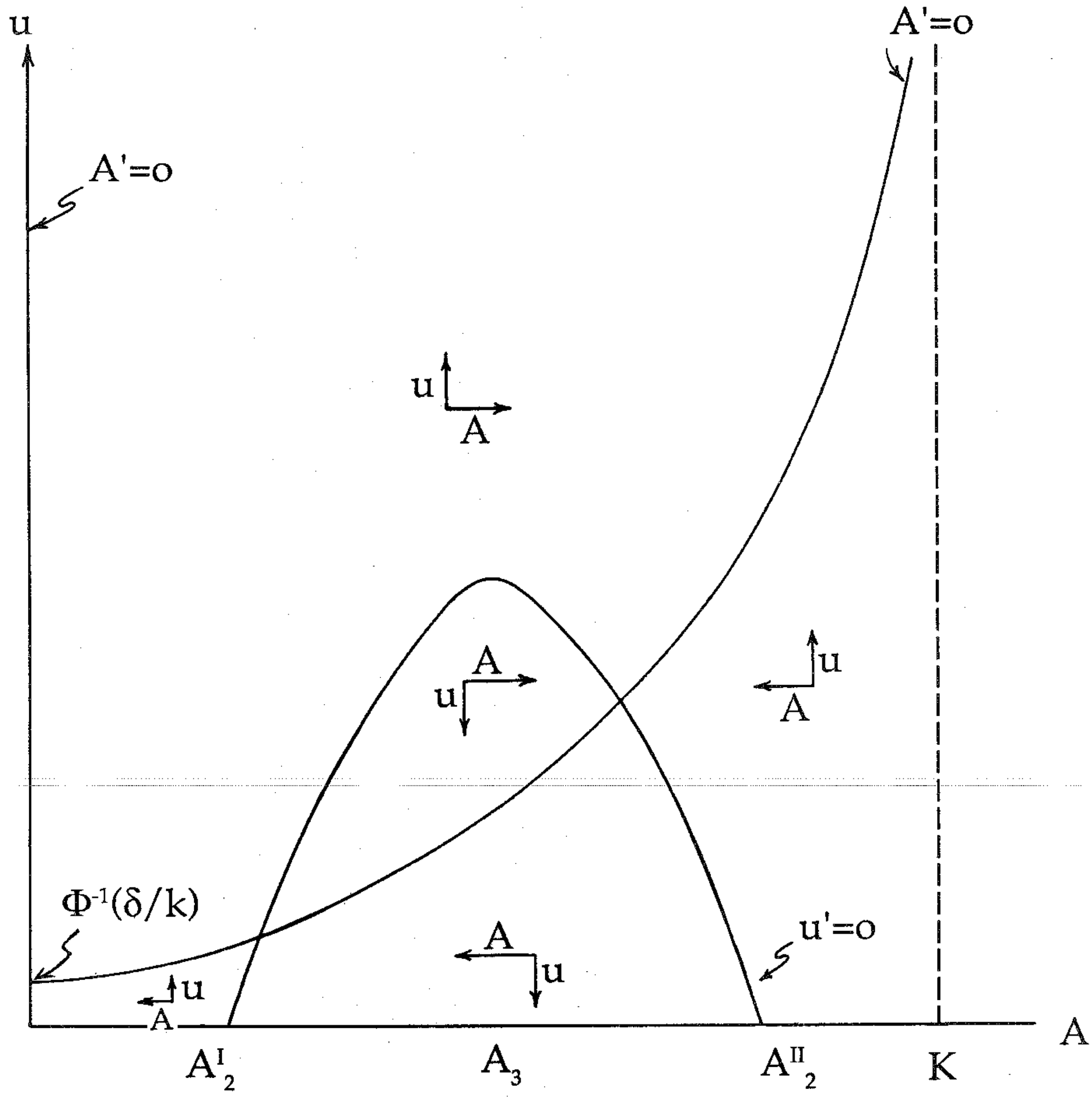

Figure 6 : Isoclines (Ozga-Gould Diffusion Model)

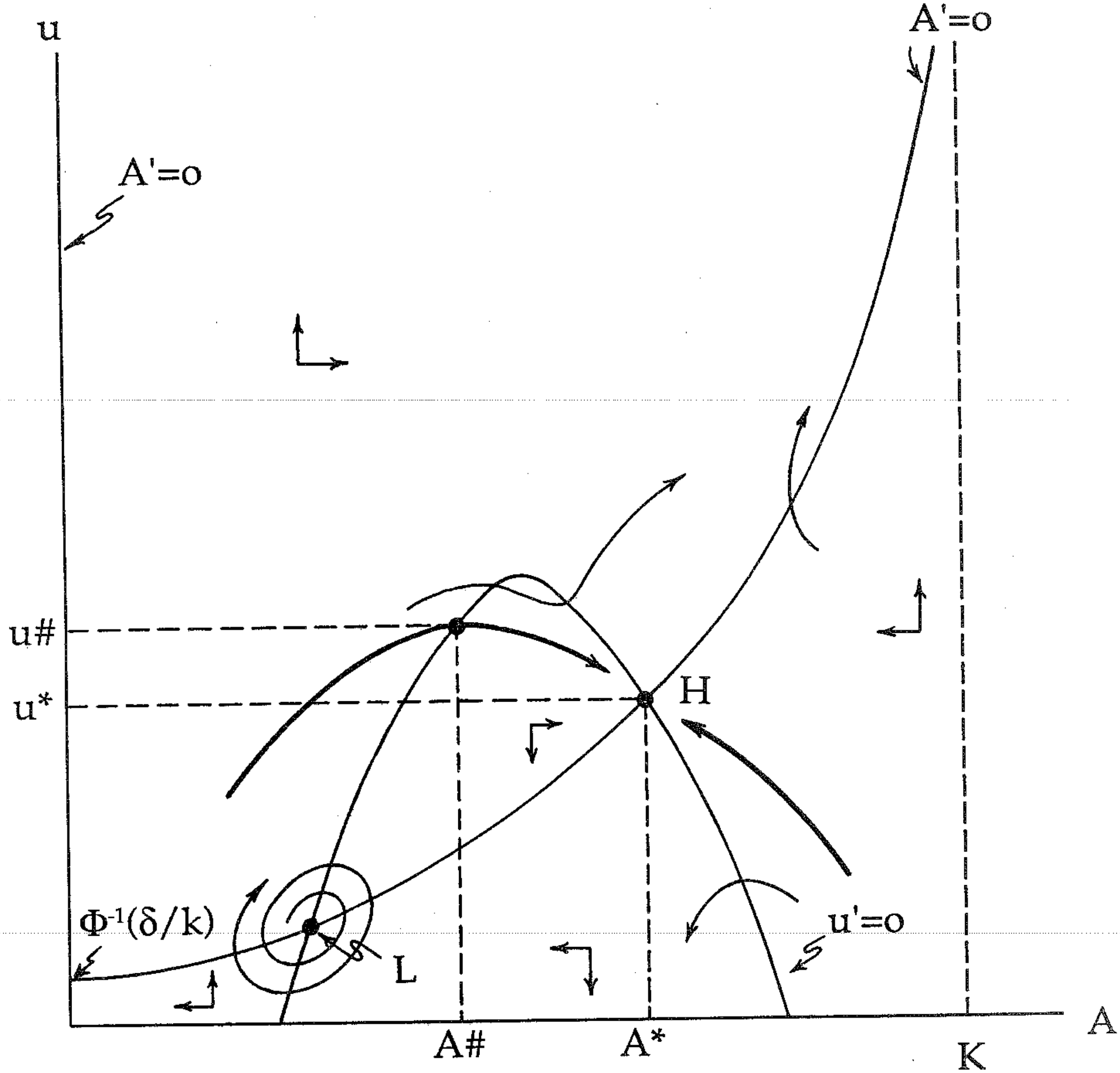

Figure 7 : Phase Diagram (Ozga-Gould Diffusion Model)

SUMMARY AND DISCUSSION

Three basic models were considered in this paper. Both the linear and nonlinear versions of the models were discussed. Nerlove and Arrow's capital theoretic approach treats advertising as an investment of the firm on goodwill which affects current and future demand. In Vidale and Wolfe's sales response model, advertising is viewed as a means of capturing some portion of the market potential up to a saturation point. Gould's diffusion approach to advertising explicitly admits the interaction between the uncaptured and the captured portions of the market either through inanimate media advertising (Stigler-Gould) or through word-of-mouth (Ozga-Gould).

Since goodwill is not observable and proxies such as consumer disposition are difficult and expensive to obtain, it is convenient to view sales as a manifestation and, hence, a practical proxy of goodwill.[17] Accordingly, the Nerlove and Arrow goodwill model becomes a special case of the Vidale-Wolfe sales response model in the sense that the former assumes an infinitely large market saturation. Since sales of any product is finite, the approach of Vidale-Wolfe is more appealing from an empirical standpoint.

In the models which take a diffusion approach to advertising, the state variable is the number of individuals who are aware of the characteristics of the promoted product. The goal of promotion is to increase the number of individuals who are aware of the product. Typically, awareness data measure only the proportion of individuals who are aware out of some population; not how intense that awareness is. To quantify the "overall" awareness of the population requires measures on both the extent and the intensity. Thus, the difficulty of incorporating both aspects into a single measurement poses a serious empirical challenge in implementing this model. However, one approach is to view sales as a manifestation and, hence, a practical proxy of the overall measure of awareness. Accordingly, the Stigler-Gould media type

17 This is not to mean that sales data are easy to come by. For example, it is very difficult to obtain sales figures encompassing both at-home and away-from-home consumption situations. A detailed discussion on the issues and difficulties involved in collecting sales data can be found in Forker, Liu, and Hurst.

diffusion model becomes identical to the Vidale-Wolfe sales response model, and the Ozga-Gould social contact type diffusion model represents a variant of Vidale-Wolfe's model.

Given the above discussion, the models reviewed in this paper can be summarized as follows. The objective functional is:

(1) $$\max_{\{u(t)\}} \int_0^\infty e^{-rt} [\Pi(A(t),Z(t)) - u(t)] \, dt.$$

While in the case of media advertising, the state equation is

(17) $$A'(t) = \Phi(u(t)) (K - A(t)) - \delta A(t),$$

and in the case of word-of-mouth type promotion, the state equation is

(24) $$A'(t) = \Phi(u(t)) A(t) (K - A(t)) - \delta A(t).$$

Equations (1) and (17) embrace the nonlinear version of the Vidale-Wolfe sales response model as well as the nonlinear version of the Stigler-Gould diffusion model. Also, the model reduces to the nonlinear version of the Nerlove-Arrow goodwill one if K is treated as infinitely large. Equations (1) and (24) encompass the nonlinear version of the Ozga-Gould diffusion model. The optimal policy derived from (1) and (17) is to advertise most heavily in the initial periods and continually decrease the level of advertising efforts as the state variable approaches its long-term equilibrium from below. In contrast, the optimal policy derived from (1) and (24) begins with a low level of expenditures, building up to a maximum which is greater than the long-term equilibrium expenditure level, and then cuts on efforts as the state approaches its long-term equilibrium from below.

In the special linear case where $\Phi(u(t)) = \phi\, u(t)$, equations (1) and (17) represent the linear version of the Vidale-Wolfe sales response model as well as the linear version of the Stigler-Gould diffusion model. Also, the model reduces to the linear version of the Nerlove-Arrow model as K approaches infinity. With the same linear assumption on Φ, equations (1) and (24) constitute the linear version of the Ozga-Gould diffusion model. Imposing the control constraints that u(t) must lie between zero and u^{max} [i.e. equation (3)], the above linear models prescribe a bang-bang-singular advertising policy; the steady state is to be approached as rapidly as possible and maintained thereafter. Without the above control constraints, the models result in an impulse advertising policy implying a jump to the steady state at the initial period of the control horizon.

In concluding the discussion, it is important to mention some limitations of the models. First, it is clear that equations (17) and (24) represent two polar cases. Equation (17) represents a process in which the influence on the state is completely due to media advertising while equation (24) represents a process in which the influence is entirely due to social contact. A more realistic model would be one which combines both effects, although the algebraic manipulation for solving the problem becomes more complicated (see Kotowitz and Mathewson for an initial attempt). Second, with regard to the diffusion process of (24), it may be unrealistic to assume that all the informed agents of the population are linked with all the uninformed agents in an all-channel social structure. This naive approach ignores the effects of social segmentation in the process of diffusion (see Bernhardt and Mackenzie for an extension). Finally, based on empirical evidence, it is dubious that advertising expenditures generate maximum response immediately and then decay exponentially over time (see Mann and Bultez and Naert for various extensions and the resulting policy implications).

The brief discussion on the model limitations has focused only on the specification of the state equation which distinguishes various models considered in this paper. Other extensions can be found in Sethi (1977[a]) and Little. Also, to adapt the firm-type models to generic commodity promotions, extensions have to be made to account for such issues as supply response (entry), government reaction in support policy, and allocation of funds between primary and secondary markets. As the model becomes more realistic and hence complex, however, it is less likely that one will be able to solve the problem analytically or numerically in a precise manner. In such a case, obtaining an approximate solution may be the only alternative. Whether or not an approximate solution to a more precise problem is preferred is a judgement left to future empirical efforts.

REFERENCES

Arrow, K. J. "Optimal Capital Policy, the Cost of Capital, and Myopic Decision Rules." Annals of the Institute of Statistical Mathematics 16(1964): 21-30.

Bernhardt I., and K. D. Mackenzie. "Some Problems in Using Diffusion Models for New Products." Management Science 19(1972): 187-200.

Bultez, A. V., and P. A. Naert. "Does Lag Structure Really Matter in Optimizing Advertising Expenditure?" Management Science 25(1979): 454-65.

Clark, C. W. *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*. New York: John Wiley & Sons, 1976.

Forker, O. D., D. J. Liu, and S. J. Hurst. *Dairy Sales Data and Other Data Needed to Measure Effectiveness of Dairy Advertising*. A.E. Res. 87-25, Department of Agricultural Economics, Cornell University, October 1987.

Glaister, S. "Advertising Policy and Returns to Scale in Markets where Information is Passed between Individuals." Economica 41(1974): 139-56.

Gould, J. P. "Diffusion Processes and Optimal Advertising Policy," in E. S. Phelps, ed., *Microeconomic Foundation of Employment and Inflation Theory*. New York: Norton, 1970, pp. 338-68.

Kamien, M. I., and N. L. Schwartz. *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management.* New York: North Holland, 1981.

Kinnucan, H. W. "Evaluating Advertising Effectiveness Using Time Series Data," in *Proceedings from Research on Effectiveness of Agricultural Commodity Promotion Seminar*. Arlington, Virginia, April 9-10, 1985.

Kinnucan, H. W. *Demographic Versus Media Advertising Effects on Milk Demand: The Case of The New York City Market*. Staff Paper No. 82-5, Department of Agricultural Economics, Cornell University, March 1982.

Kotowitz, Y., and F. Mathewson. "Informative Advertising and Welfare." American Economic Review 69(1979): 284-94.

Little, J. D. C. "Aggregate Advertising Models: The State of the Art." Operations Research 27(1979): 629-67.

Liu, D. J., and O. D. Forker. "Generic Fluid Milk Advertising, Demand Expansion, and Supply Response: The Case of New York City." American Journal of Agricultural Economics 70(1988): to appear in May issue.

Mann, D. H. "Optimal Theoretical Advertising Stock Models: A Generalization Incorporating the Effects of Delayed Response from Promotional Expenditure." Management Science 21(1975): 823-32.

Nerlove, M., and K. J. Arrow. "Optimal Advertising Policy under Dynamic Conditions." Economica 29(1962): 129-42.

Nerlove, M., and F. V. Waugh. "Advertising Without Supply control: Some Implications of a Study of the Advertising of Oranges." *J. Farm Econ.* 43(1961):813-37.

Ozga, S. "Imperfect Markets through Lack of Knowledge." Quarterly Journal of Economics 74(1960): 29-52.

Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *The Mathematical Theory of Optimal Processes.* New York: John Wiley, 1962.

Samuelson, P. A. "A Catenary Turnpike Theorem Involving Consumption and the Golden Rule." American Economic Review 55 (1965): 486-96.

Sethi, S. P. "Optimal Control of the Vidale-Wolfe Advertising Model." Operations Research 21(1973): 998-1013.

Sethi, S. P. "Dynamic Optimal Control Models in Advertising: A Survey." SIAM Review 19(1977[a]): 685-725.

Sethi, S. P. "Optimal Advertising for the Nerlove-Arrow Model under a Budget Constraint." Operational Research Quarterly 28(1977[b]): 683-93.

Stigler, G. "The Economics of Information." Journal of Political Economy 69(1961): 213-25.

Vidale, M. L., and H. B. Wolfe. "An Operations Research Study of Sales Response to Advertising." Operations Research 5(1957): 370-81.

APPENDIX A: NERLOVE-ARROW FINITE PLANNING HORIZON MODEL

Due to management considerations, it is sometimes desirable to specify the problem with a finite planning horizon T. In this case, the objective functional is:

(A1) $\max_{\{u(t)\}} \int_0^T e^{-rt} [\Pi(A(t),Z) - u(t)]\, dt.$

The state equation is (2) and the control constraint is (3). For conveniences, they are reproduced here as:

(A2) $A' = \phi\, u(t) - \delta\, A(t)$, with $A(0)=A_0$,

(A3) $0 \leq u(t) \leq u^{max}$.

The current-value Hamiltonian and the current-value Lagrangian are still (4) and (5) and they are reproduced here as:

(A4) $H = \Pi(A(t)) - u(t) + \psi(t)\, [\phi\, u(t) - \delta\, A(t)]$,

(A5) $L = H + \lambda_1(t)\, u(t) + \lambda_2(t)\, [u^{max} - u(t)]$.

Accordingly, the optimal control u^* must satisfy (7) and the adjoint variable ψ must satisfy (8). These two conditions are reproduced here as:

(A6) $$u^*(t) = \begin{cases} 0 \\ u^s \in [0, u^{max}] \\ u^{max} \end{cases} \quad \text{if} \quad \begin{cases} \sigma(t) < 0 \\ \sigma(t) \equiv -1 + \psi(t)\, \phi = 0 \\ \sigma(t) > 0. \end{cases}$$

(A7) $\psi' = (r + \delta)\, \psi - \Pi_A$.

The only difference is the transversality condition which requires the shadow price of the state be driven to zero at the terminal time. This condition is:

(A8) $\psi(T) = 0$.

Since all the necessary conditions for optimality are the same except (A8), the optimal policy should resemble the one under infinite time horizon. That is, the optimal policy characterization is a bang-bang control followed by a singular path. However, modification of the final phase of the control is needed in order to satisfy (A8).

From (A6), the switching function $\sigma(t) \equiv -1/\phi + \psi(t) = 0$ along the singular path. Also, in view of (A8), $\sigma(T) = -1/\phi$ which is negative. Thus, one must leave the singular path before T. While off the singular path, (A6) says that one must use bang-bang control of $u^*(T) = 0$ because $\sigma(T)$ is negative. Since the switching function $\sigma(t)$ is a continuous function of time and it is zero on the singular path while negative at T, proceeding backward in time, one must find a $t^{\#} < T$ such that $\sigma(t^{\#}) = 0$ and $\sigma(t) < 0$ for $t > t^{\#}$. This means that the singular path should be followed up to $t^{\#}$ only and then the control should be switched to $u^{*} = 0$.

Thus, as shown in Figure A1, the optimal policy is a bang-bang-singular control, followed by a bang-bang control with $u^{*}(t) = 0$ for $t > t^{\#}$. The last phase of the control is to drive $\psi(t^{\#}) = 1/\phi$ to $\psi(T) = 0$. In other words, with zero gross goodwill investment, $T\text{-}t^{\#}$ is the time required for the equilibrium goodwill shadow price to depreciate to zero exactly at time T.$ The policy of getting on the singular path as early as possible and off the path as late as possible is referred to as the turnpike property (Samuelson).

In the above, it is implicitly assumed that T is sufficiently large (in the sense that $t^{\#} > t^{I}$ in the case where $A_0 < A^{s}$, and $t^{\#} > t^{II}$ in the case where $A_0 > A^{s}$). When T is sufficiently small relative to the time required for A to approach A^{s} from A_0, however, it will not be possible to attain the singular path and meanwhile satisfy the transversality condition (A8). For example, consider Figure A2 where $A_0 < A^{s}$ and $t^{I} > t^{\#}$. On the one hand, one wishes to get on the singular path as rapidly as possible. On the other hand, however, one knows that once on the singular path there will not be enough time to get off the path and meet the endgame requirement. As a compromise, the best policy in this case is $A_0SA(T)$ which minimizes the deviation from the singular path while satisfying the endgame condition.

$ Numerical methods for the determination of $t^{\#}$ can be found in Sethi (1977[b]).

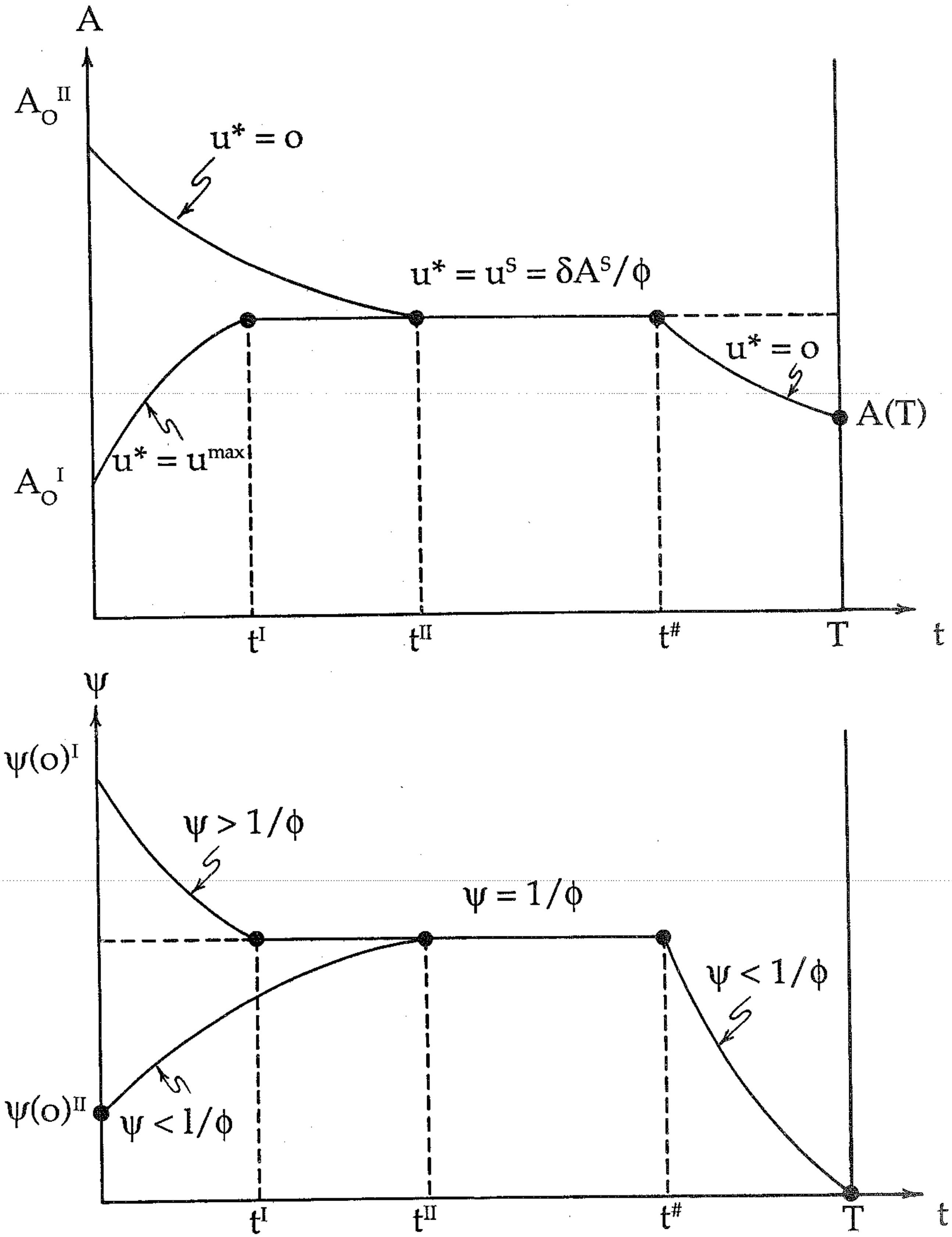

Figure A1: Bang Bang, Singular, Bang Bang Control

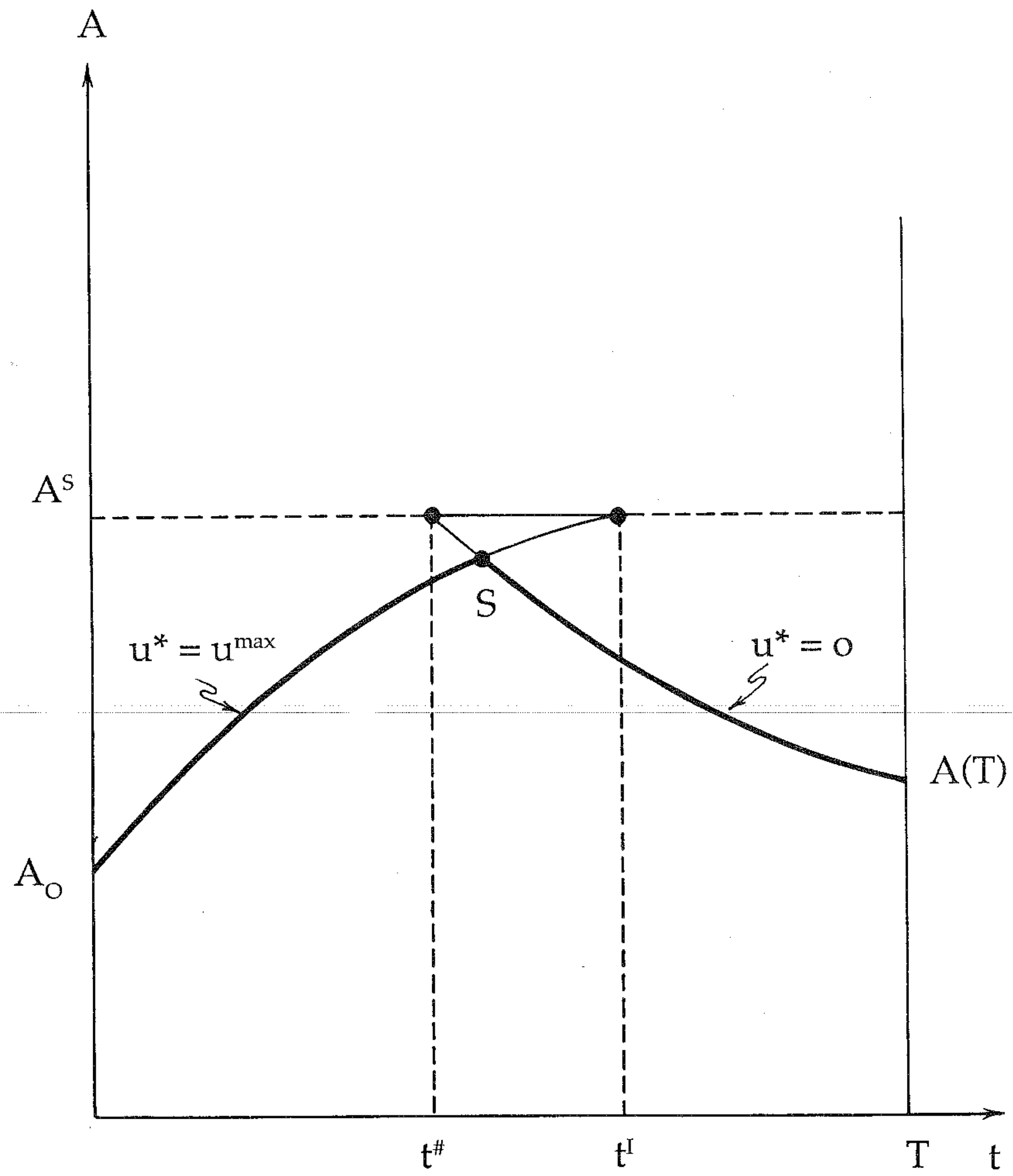

Figure A2: The Case of Small T

APPENDIX B: INFEASIBLE SINGULAR CONTROL

Depending on the control constraint, the singular path may not be attainable. Arrow has introduced the term "blocked interval" to refer to any time interval during which the constraints of a linear control problem prevent the path A(t) from following a nonstationary singular path $A^s(Z(t))$. A blocked interval can also occur in the case where the singular path is stationary. The initial bang-bang adjustment phase discussed is an example of a blocked interval as the control is blocked by its constraints (zero and u^{max}).

An interesting case arises where u^{max} is strictly less than the singular control $u^s = \delta A^s/\phi$. Consider the scenario where $A_0 < A^s$. In accordance with the most rapid approach discussed, the maximum control $u^*(t) = u^{max}$ should be applied until A^s is attained. For a given u^{max}, (2) implies $A(t) = \phi u^{max}/\delta$ as $t \to \infty$, indicating that the singular A^s will never be attained if u^{max} is strictly less than $\delta A^s/\phi$ (i.e. u^s). However, since the optimal path must lie as close as possible to the singular path, applying the maximum policy u^{max} is still the best strategy. In this case goodwill stock will accumulate (or decumulate if $A_0 > \phi u^{max}/\delta$) over time and approach $\phi u^{max}/\delta$ as $t \to \infty$ (Figure B1).

Now, consider the opposite scenario where $A_0 > A^s$. Since u^s is not admissible, a reasonable guess of the optimal policy would be to allow A_0 to depreciate to A^s by applying the minimum control (i.e. zero) and then switching to the maximum policy (i.e. u^{max}) at t^{II} with the goal of minimizing the deviation from the singular path (Figure B2). This policy, in general, will not be optimal. To prepare for the fact that the singular path cannot be maintained once it is reached, it may be profitable to switch to the maximum policy at $t^{\#} < t^{II}$, even though the goodwill stock is still above the singular level at that time. As shown in Figure B2, the total area of deviation from A^s is smaller for the path (sketched in heavy line) with a switching time at $t^{\#}$ than for the path with a switching time at t^{II}. The optimal switching time ($t^{\#}$) should be determined in such a way that the objective functional in (1) is maximized; a problem solvable by calculus and/or numerical methods.

In the preceding discussion, the exogenous variable Z(t) is assumed to be constant and, hence, the equilibrium goodwill stock level $A^s(Z)$ is stationary. For a time dependent Z, $A^s(t) = A^s(Z(t))$ will be a time function (Figure B3). In this situation, the characterization of the optimal policy remains the same in the sense that the singular path should be approached as rapidly as possible and maintained whenever possible. The cyclical pattern of the singular path in Figure B3, however, opens up the possibility of blocked intervals even when u^{max} is large enough to maintain the singular path for most of the time period. As before, in this case the bang-bang policy of applying either the minimum or the maximum control must become effective and the optimal time interval must be determined for this control as it will in general not coincide with the interval of infeasibility.

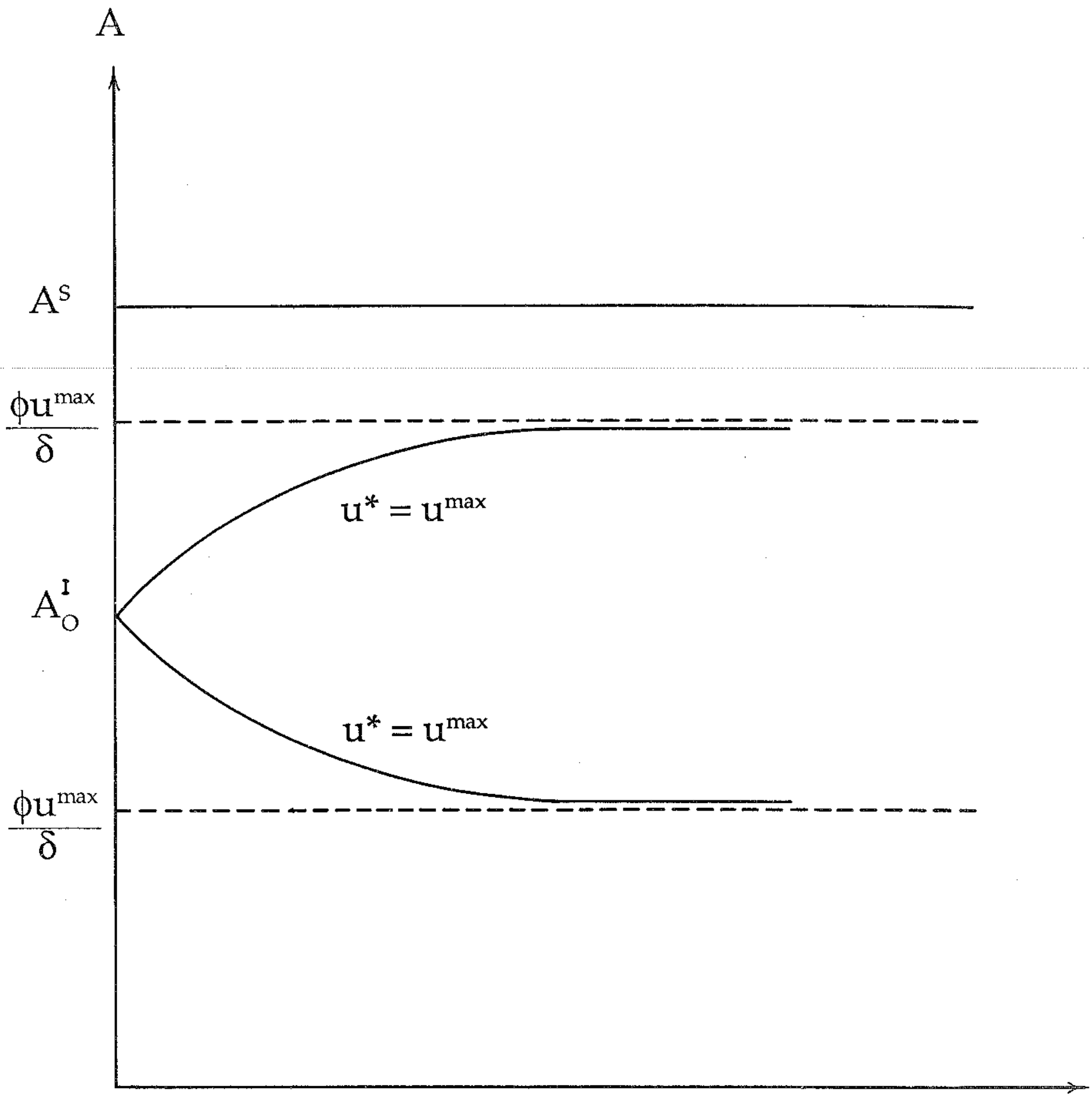

Figure B1: Infeasible Singular Control ($A^O < A^S$)

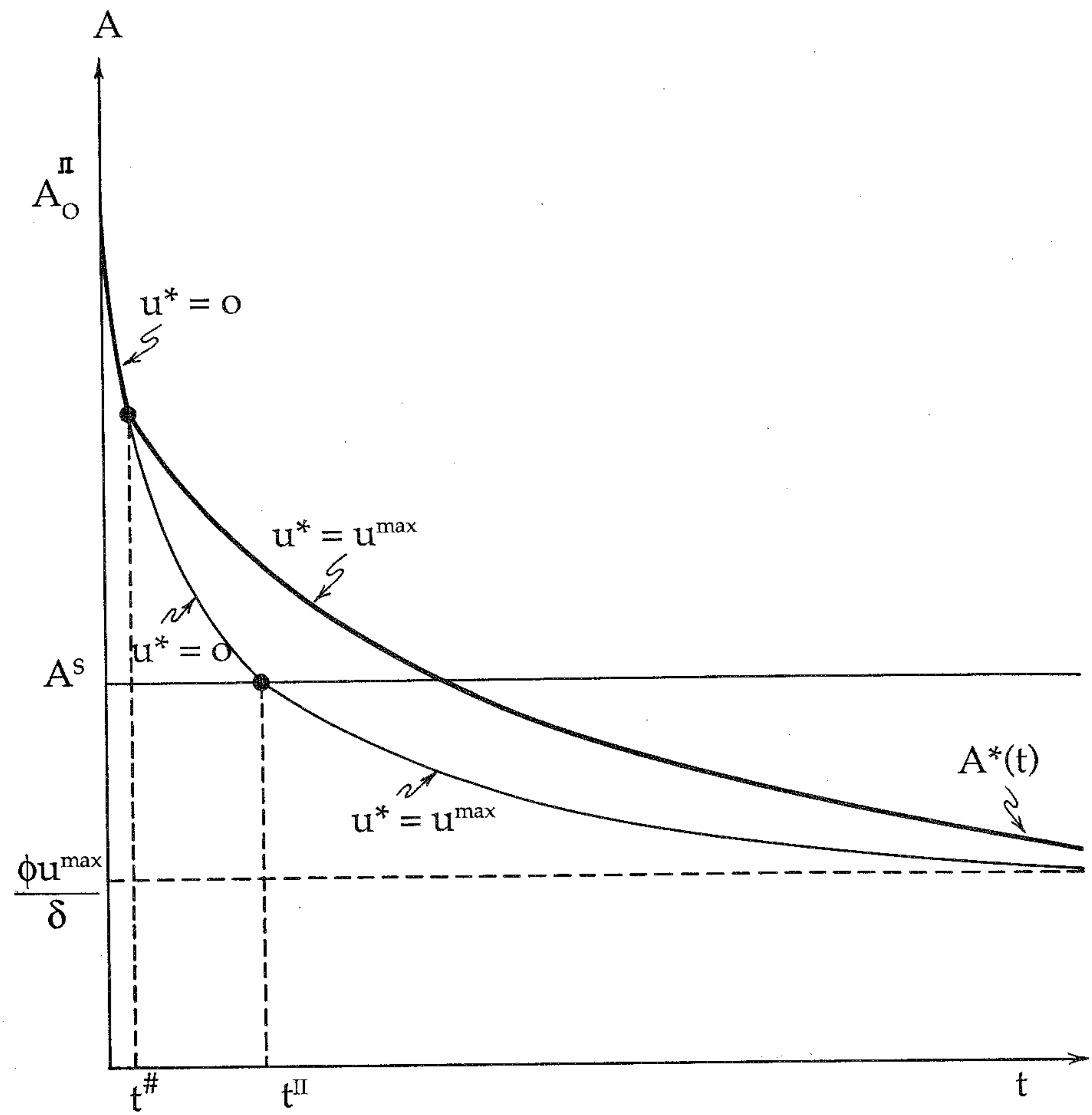

Figure B2: Infeasible Singular Control ($A_o > A^s$)

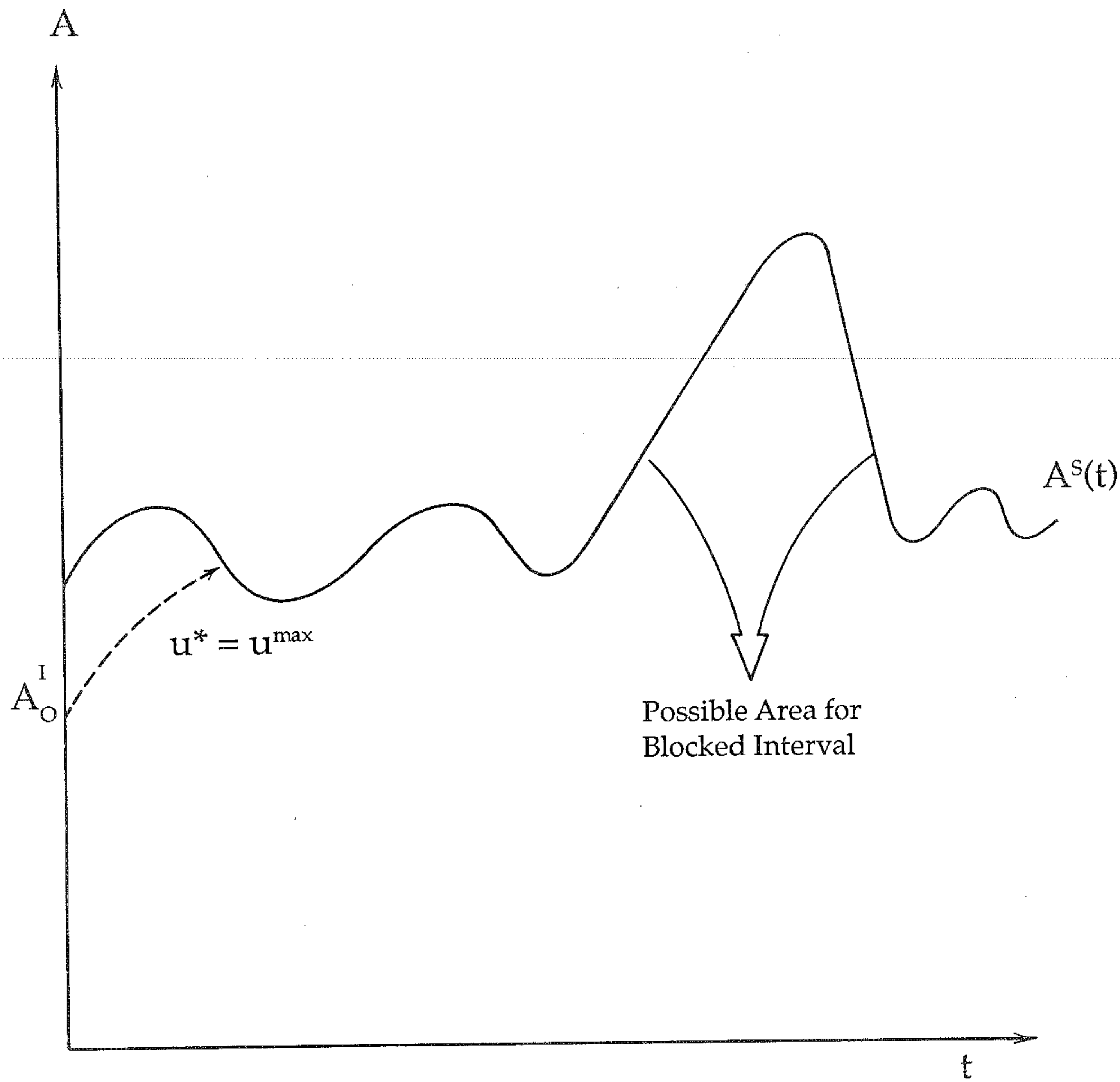

Figure B3: Nonstationary Singular Path

APPENDIX C: VIDALE-WOLFE MODEL WITH A LINEAR PROFIT

Following Sethi (1973), Π(A) is assumed to be a linear function of the state A:

(C1) $\max_{\{u(t)\}} \int_0^{\infty} e^{-rt} [\pi A(t) - u(t)]\, dt,$

s.t.

(C2) $A'(t) = \phi\, u(t)\, (K - A(t)) - \delta\, A(t),$

(C3) $0 \leq u(t) \leq u^{max}.$

The current-value Hamiltonian from (C1) and (C2) is

(C4) $H = \pi A(t) - u(t) + \psi(t) [\phi\, u(t)\, (K - A(t)) - \delta\, A(t)],$

and the current-value Lagrangian is

(C5) $L = H + \lambda_1(t)\, u + \lambda_2(t)\, [u^{max} - u(t)].$

The Kuhn-Tucker conditions $L_u = 0$, $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, $\lambda_1 u = 0$, and $\lambda_2(u^{max}-u) = 0$ imply the following:

(C6) $$u^*(t) = \begin{cases} 0 & \sigma(t) < 0 \\ u^s \in [0, u^{max}] \text{ if } & \sigma(t) \equiv -1 + \psi(t)\, \phi\, (K - A(t)) = 0 \\ u^{max} & \sigma(t) > 0. \end{cases}$$

The singular condition in (C6) indicates that the last dollar spent should equal the shadow price of sales times the marginal increase in sales due to that spending. Also necessary is the adjoint equation:

(C7) $\psi' = (r + \delta + \phi\, u)\, \psi - \pi,$

Equation (C7) indicates that the change in the imputed value of the state and its marginal contribution to the cash flow must equal the marginal opportunity costs of having the state at that level. Here the opportunity costs include not only the time costs and the depreciation costs but also costs of the diminished marginal advertising sales effectiveness which arises as the saturation point is approached. Obviously, there is a tradeoff; advertising increases sales, but the added sales depress the effectiveness of future advertising.

According to (C6), advertising will be at either its lower or its upper bound except when the switching function $\sigma(t)$ is zero. In this latter case, $\psi(t)\, (K - A(t)) = 1/\phi$. Since ϕ is a constant, one has

(C8) $0 = d[\psi\, (K - A)]/dt = \psi'\, (K - A) - A'\, \psi.$

Substituting into (C8) A' from (C2), ψ' from (C7), and ψ (evaluated at the singular path) from (C6), one obtains

(C9) $\phi \pi (K - A^s)^2 - r (K - A^s) - \delta K = 0$.

Equation (C9) is quadratic in $(K - A^s)$ and it has two roots:
$(K - A^s) = [r \pm (r^2 + 4 \phi \pi \delta K)^{1/2}] / (2 \phi \pi)$.

Since $K \geq A$ by definition, only the positive root can be relevant. Hence, the singular state level is

(C10) $A^s = K - [r + (r^2 + 4 \phi \pi \delta K)^{1/2}] / (2 \phi \pi)$.

Further, $A^s \geq 0$ is also needed for sense. Setting $A^s = 0$ in (C10), one finds that $A^s \geq 0$ requires $\pi \phi K \geq r + \delta$. Expressing the requirement in terms of the sales response coefficient ρ in Vidale-Wolfe equation (15), one has:

(C11) $\pi \rho \geq r + \delta$.

Stated in words, condition (C11) says that advertising is worthwhile expending only when the market condition (π) and the sales response coefficient (ρ) are large enough, relative to the opportunity costs of advertising arising from interest rate (r) and sales decay (δ).

Substituting A^s from (C10) into (C2) and taking into account that A' is zero along the singular path, the singular advertising policy is

(C12) $u^s = (\delta A^s) / [\phi (K - A^s)]$,

providing that the singular control is feasible:[$]

(C13) $u^s \leq u^{max}$.

In sum, if advertising is worthwhile [i.e. (C11)] and the steady state control is feasible [i.e. (C13)], the optimal singular path is characterized by (C10) and (C12) and it should be approached as rapidly as possible by selecting either the maximum or the minimum control [i.e. (C6)].

$ The explicit condition for (C13) can be found by substituting (C10) into (C12) and then the resulting expression into (C13). If the singular control is infeasible (i.e. $u^s > u^{max}$), then the policy should follow those discussed in Figure B1 and Figure B2.

APPENDIX D: THE DERIVATION OF PHASE DIAGRAM (OZGA-GOULD MODEL)

From (28a), u′ will be zero for points (u,A) satisfying:

(D1a) $1/\Phi_u = [\Pi_A A (K - A)] / [r + \delta A / (K - A)]$

$\equiv h(A)$.

Since Φ^{-1} is convex in u, the left-hand side of (D1a) is an increasing function of u. By inspection, the right-hand side depends only on A and it is zero at both A=0 and A=K. Thus, h(A) is increasing for small A and decreasing for large A. Figure D1 illustrates the identification of pairs (u,A) satisfying (D1a).

As illustrated, when u is small, there will be two values of A associated with each such u. For example, at $u=u_1$, (D1a) is satisfied for $A=A_1^{I}$ and $A=A_1^{II}$. Similarly, the values A_2^{I} and A_2^{II} are associated with $u=u_2$. Note that as u gets larger, the associated values of A get closer together. For $u=u_3$, there is only one unique $A=A_3$, and (D1a) has no solution for those u greater than u_3. Finally, since $\Phi_u > 0$ (and hence $\Phi_u(0)^{-1} > 0$) and h(0) = h(K) = 0, it follows that at u=0, the u-isocline intersects the A axis at two points between zero and K (i.e. A_z^{I} and A_z^{II}).

Given the discussion above, the u-isocline is shown in Figure D2. The directional arrows there are determined as follows. Differentiating u′ with respect to u in (28a), one obtains $\partial u'/\partial u|_{u'=0} = \Pi_A \Phi_u A (K - A) > 0$. Thus, for points above the u′=0 locus, u′ > 0 and for points below u′=0 locus, u′ < 0.

Consider the A-isocline. Setting A′=0 in (28b), one finds that there are two A-isoclines; the line A=0 and:

(D1b) $A = (\Phi K - \delta) / \Phi$.

From (D1b), it holds that A=0 implying $\Phi(u) = \delta/K$ and, hence, $u = \Phi^{-1}(\delta/K) > 0$. Also, u grows without bound as A approaches K. Using (D1b), one finds $dA/du = \delta \Phi_u / \Phi^2 > 0$ and $d^2A/du^2 = [\delta \Phi_{uu} \Phi - 2 \delta \Phi_u^2] / \Phi^3 < 0$. Thus, the A-isocline described in (D1b) is upwardly sloped and convex in u-A plane. Further, from (28b), $\partial A'/\partial u = \Phi_u A (K - A) > 0$ indicates that above this isocline A is increasing and A decreases below the isocline (Figure 6).

There are two stationary points (denoted by L and H) in the phase diagram (Figure 7). It is necessary to examine the behavior of the system at each. From the directional arrows in the diagram, it appears that the high equilibrium H (the one with larger coordinates) is a saddle point. All the paths near the separatrices (sketched in heavy lines) seem to diverge. Also, it appears that the trajectory near the low equilibrium L is cycling. However, it is not clear whether the cycle converges to L or not (though a diverging cycle is showing in the diagram). To verify that H is a saddle and to investigate whether the trajectory near L converge or not, linearize (28) around an equilibrium point (u^*, A^*) and express the result in the format of (13). One finds:

$\beta_{11} \equiv \partial u'/\partial u = \Pi_A(A^*)\ \Phi_u(u^*)\ A^*\ (K - A^*) \qquad > 0$

$\beta_{12} \equiv \partial u'/\partial A \qquad ?\ 0$

$\beta_{21} \equiv \partial A'/\partial u = \Phi_u(u^*)\ A^*\ (K - A^*) \qquad > 0$

$\beta_{22} \equiv \partial A'/\partial A = \delta - \Phi(u^*)\ K \qquad < 0.$

In arriving at the sign of β_{22}, one uses the previous results that the u-isocline has a positive slope and it intersects the u-axis at the point where $\Phi(u) = \delta/K$. Given the signs of α_{ij}, one knows that saddle point condition (14) holds only if α_{12} is not negative. However, it can be shown that the sign of α_{12} depends on the size of A^* relative to K and, hence, is ambiguous. The condition for saddle point in (14) is equivalent to:

(D2) $\quad -\alpha_{12} / \alpha_{11} < -\alpha_{22} / \alpha_{21}.$

Now, from (13), the linearized u'=0 locus has slope $du/dA|_{u'=0} = -\alpha_{12} / \alpha_{11}$ (whose sign is ambiguous) and the linearized A'=0 locus has slope $du/dA|_{A'=0} = -\alpha_{22} / \alpha_{21}$ (whose sign is positive). In view of these and (D2), it is clear that the saddle point is the equilibrium where the linearized A-isocline is steeper than the linearized u-isocline. Examining Figure 7, the high equilibrium H is characterized by the A'=0 locus being steeper than the u'=0 locus and, hence, is a saddle point. On the other hand, the low equilibrium L is totally unstable since A'=0 locus is less steep than the u'=0 locus.

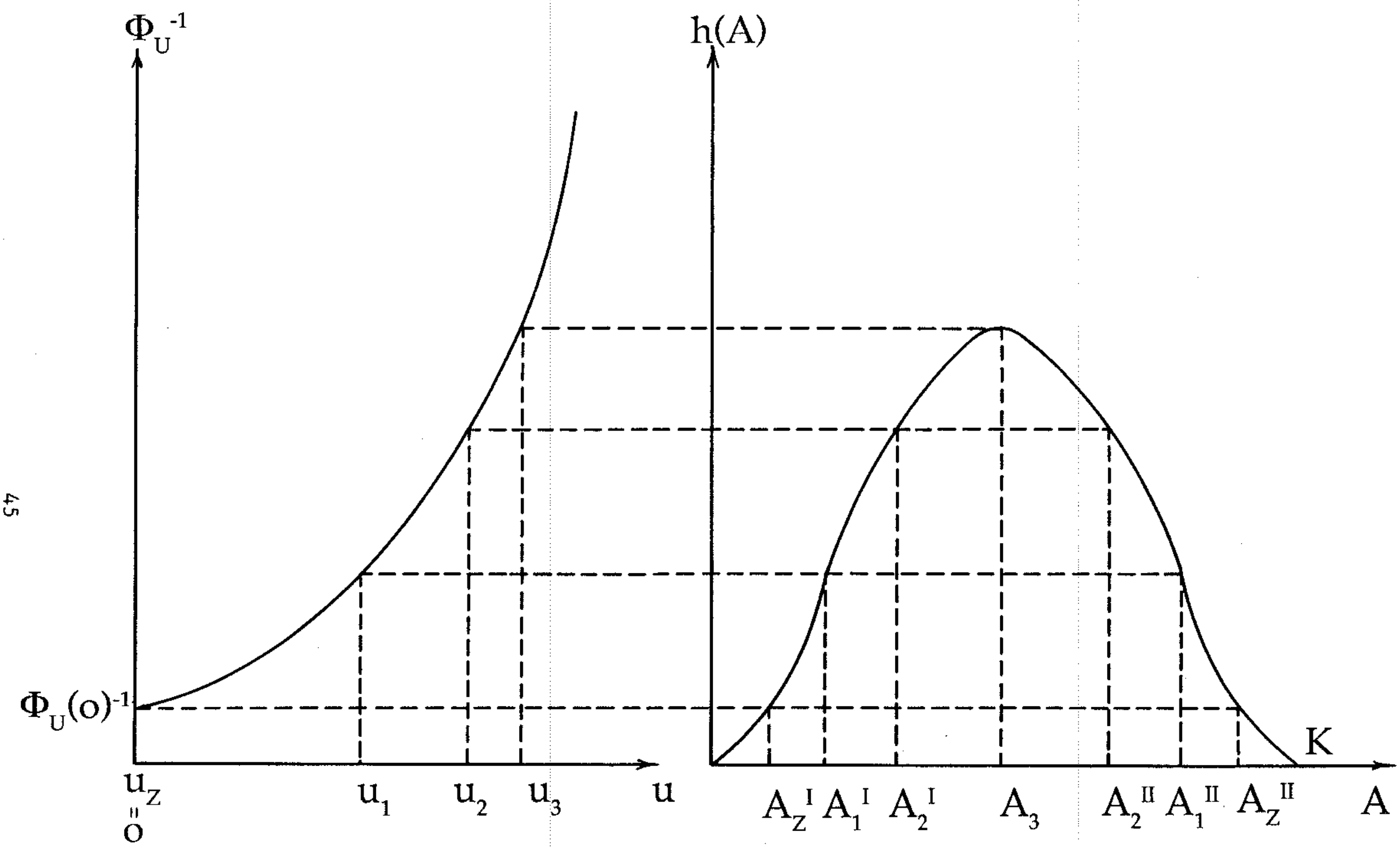

Figure D1: Identification of (U, A) pairs along the u-Isocline

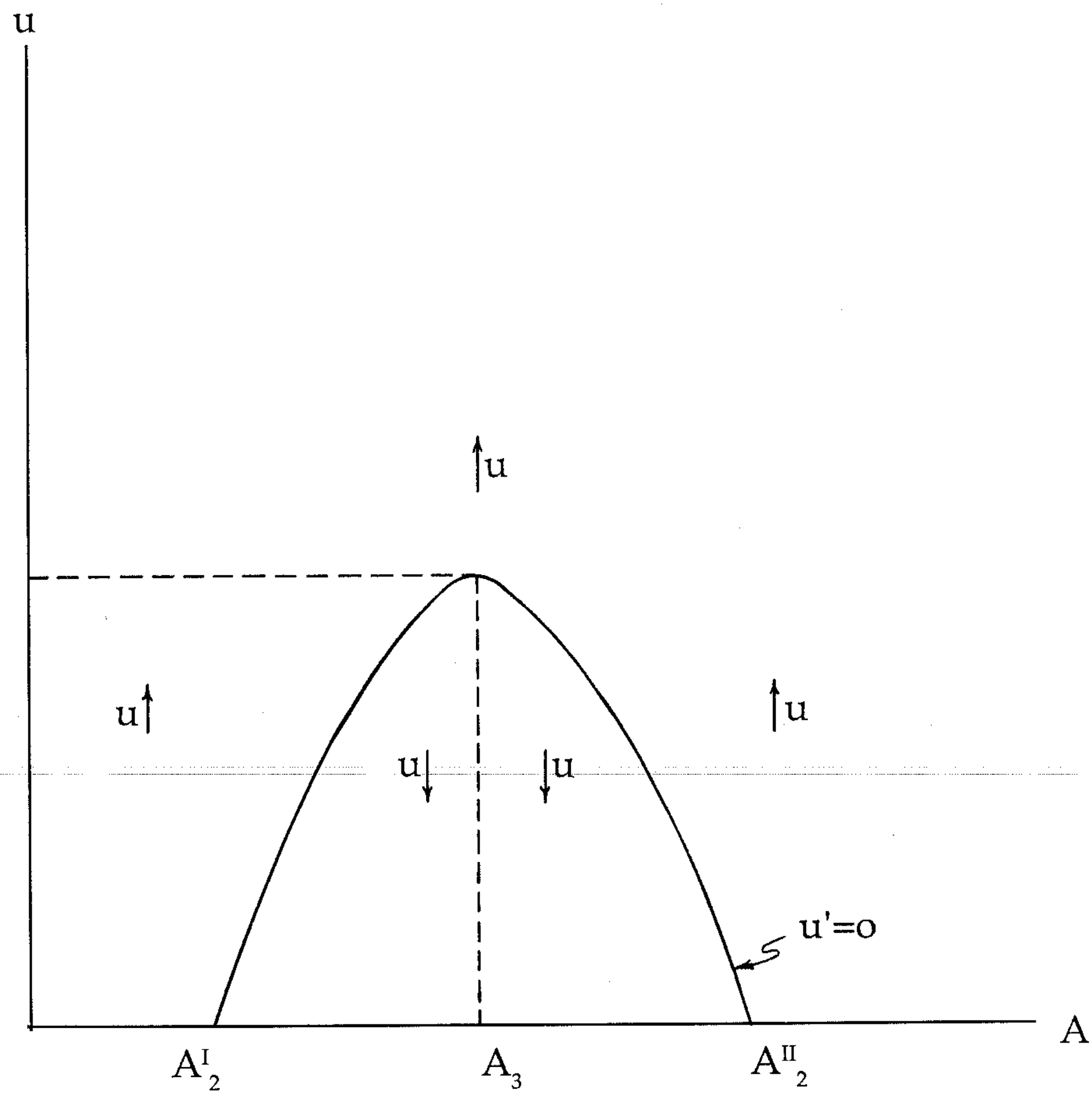

Figure D2: u-Isocline (Ozga-Gould Diffusion Model)