Working Paper

AN OPTIMAL CONTROL MODEL FOR
THE DIFFUSION OF INNOVATION

Peter Markowich

February 1980
WP-80-108

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

AN OPTIMAL CONTROL MODEL FOR THE DIFFUSION OF INNOVATION

Peter Markowich

February 1980
WP-80-108

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

PREFACE

The diffusion of technology has been modeled using different modeling techniques, and these models should fulfill different needs. For example, there are spatial models, models describing the technological side or the consumer side, and so on. On the other hand there are many models which just describe the process but few really give advice for necessary regulations or outside influence. Therefore, I have set up a model called DIFFOPT, which should describe the production aspect and the societal aspect of the diffusion process, and should also give explicit advice for investment and price policy of innovative technologies. This paper deals with the development of DIFFOPT which has the structure of an optimal control model, and with the mathematical description.
Moreover the descriptive model which is basic for the optimal control model is tested computationally.

CONTENTS

SOME SUBSTITUTION MODELS

We will first look at Peterka's substitution model, which was described in Peterka (1977). This model describes, in its most general setting, the development of market shares--in terms of production--of n technologies competing for the same market, and it is in the case of two competing technologies consistent with the well-known Fisher and Pry logistic substitution model.

Peterka's basic model equation is:

$$\alpha_i \dot{P}_i(t) = P_i(t)(p(t) - c_i) + q_i(t) \qquad i = 1(1)n \quad .$$

where:

$P_i(t)$	production of i-th technology at time t,
α_i	investment needed to increase production by one unit,
$p(t)$	reference price at time t,
c_i	production costs of one unit of i-th commodity including amortisation of material goods used in production and eventually taxes set on the product,
q_i	external capital flow.

The first assumption Peterka makes is in order to derive an equation with market shares of the n technologies as state variables. He states that the external capital flow is zero from a certain time level up to the end of the time interval, over which we want to observe the development of n technologies. What does this mean for the real world? As Peterka says, each technology has to grow on its own (from a certain time level on). Of course this is not very useful for our purposes because when technologies are in the start-up phase having only a small market share, the price per unit would be very high. Therefore, the commodity would be totally unattractive. The main problem concerning the self growth of new technologies is the efficiency problem (see H.D. Haustein and H. Maier). A new technology can only grow on its own if the ratio of marginal and mean efficiency is greater than 1 and this cannot be fulfilled in the start-up phase. So this assumption restricts the validity of Peterka's model to the case that each technology has already gained a certain market share, let us say a certain percentage, and so the model cannot give a good description of the first phase of the diffusion process.

Another thing that we may easily observe when looking at the model equation, is that the production of the i-th commodity grows exponentially whenever the difference between commodity price and production costs is positive. Assuming high production costs, and (because there is no external investment) consequently higher prices, which makes the commodity totally unattractive on the market, the production will still grow exponentially, which is also unrealistic. This was improved by Spinrad (1979) by introducing the system's growth rate $\rho = \dot{P}/P$ where P is the total production of all n technologies. Peterka also assumes implicitly that the market in which the n technologies are involved is clear, this means that all units produced at time t are sold at the same time. This is not true for all technologies, especially in market economies. There are many technologies where there is a large inventory so that there is a (sometimes large) time delay between production and sale.

However, assuming the situation described in the model's equation, Peterka uses some mathematical techniques to eliminate the price and to introduce instead of production, market shares into the system. This has the advantage, as observed by Fisher and Pry, that market shares of technologies behave more regularly than production quantities. Finally, Peterka is able to give an algorithm to estimate the unknown parameters using historical data of market shares and to estimate error probabilities for forecasting. This is the greatest advantage of this model. The algorithm for parameter estimation is, although derived by difficult mathematical methods, very capable and simple.

But what of forecasting? Of course the system cannot recognize future events like births of new technology, political change, and new energy resources. It can, however give some reliable information about how the market is shared in the immediate future between the given technologies presuming the structure of the model. It also teaches us another important feature of the diffusion process of new technologies: the importance of time cycles. Looking at Peterka's computed curves for energy substitution we see that the diffusion of new primary energies takes its time. This means that we cannot stop oil usage and start up with solar energy at once. This is, more or less, valid for all types of innovation. Every innovation takes its preparatory time, of course, basic innovations take a longer and more intensive preparation than improvement innovations. (For the explanation of these terms see Maier and Haustein 1979). Basic innovations are more investment intensive, and more R & D intensive than improvement innovations. Therefore the start-up phase is more difficult, and consequently for the IIASA innovation task group, more interesting. It turns out therefore, to be necessary to model the start up phase of innovations.

Peterka's model equation also shows that it is necessary to have improvement innovations following basic innovations in order for them to remain competitive. This is because improvement innovations have a great influence on the coefficients c_i. Successful improvement innovations decrease the production costs of new technology and so the price p(t) can be decreased, and the production will still increase. The attractiveness of the technology to the consumer is also increased.

On the other hand, the model was applied to primary energy and so the consumer side is not very important. But the innovation process is an interaction between supply and demand, and the demand side seems to be more probabilistic than deterministic because the consumer's preference may be influenced by several factors such as personal taste, personal innovativeness, success or failure of advertisements, or political circumstances. This does not mean that the process is not controllable and observable but it's nature is very aptly described by the mathematical term "random process" which means that one cannot say deterministically what will be, but that one can compute probabilities for certain circumstances which may arise.

We have now reached a very good starting point for starting the discussion of the second of Fleck's substitutions models, which is described in Peterka and Fleck (1978). It describes the diffusion process from quite a different point of view, namely, it reflects the consumer's side of the process. The mathematical tool used is a discrete Markoff process, which is a sequence of random variables X_n which have the special feature in that X_n only depends on the realisation of X_{n-1}, but not on the previous X_i. Following the theory of the Markoff process, Fleck assumes that the probability of a consumer owning the commodity produced by old technology turns to the new one at the time point t, is linearly dependent of the market share, the new technology has gained the time t-1. Calling that probability $P_o(t,t-1)$ this means

$$P_o(t,t-1) = \rho f(t-1) + \gamma$$

where f is the market share of new technology now in terms of sales. Further, he assumed that having reached the state "new technology" the consumer cannot get back to "old technology" which is quite a realistic assumption, especially if the new technology is investment intensive, because if the consumer had to put a lot of money in, it will be very hard to withdraw that money without having great losses. It is not so realistic to assume that the transition probability from 'old' to 'new' only depends on the market share reached one time unit before, but it allows us to use the very convenient mathematical tool, the "Markoff Process".

But what do the constants ρ and γ mean? γ gives the transition probability having a zero market share at the last time point, so it is a measure for the individual's "innovativeness", and ρ describes the imitation behavior of the consumer. The larger γ is the more 'innovative' the consumer, and the larger ρ is, the more he imitates others. γ and ρ are functions of the kind of innovation with which we are dealing. For example, the adoption of a basic innovation is in general a more expensive and important matter than the adoption of an improvement innovation, so the imitation coefficient ρ should be smaller in the case of basic innovations than in the case of improvement innovations. By aggregating all possible consumers Fleck gains a difference equation for the market share of new technology with the parameters ρ and γ from which he can easily derive a differential equation with the same features. The parameters ρ and γ can be estimated from historical data by least-square-estimation. The greatest flaw in this model is the aggregation, because all consumers do not show equal behavior, and that is exactly what we loose in the performance of aggregation. Fleck also applied his model to several cases, and it showed that the resulting curves have a similar behavior to those gained by Fisher and Pry's empirical model. On the other hand, it does not give much advice to decision makers because it is only descriptive. One interesting feature is that the model is much more sensitive to the consumer's 'innovativeness' than to the imitation factor. This might mean that it makes much more sense (for technological change) to educate people to be well informed, modern thinking, and capable of taking well considered risks, rather than educating them to make imitations.

But Fleck's model has another disadvantage too; looking at the derived differential equation for new technology's market share one notices that its derivative is always greater than or equal to zero and that means that new technology's market share is always a non-decreasing function of time unless $\rho = \gamma = 0$. Applying this to the real world would mean that no innovation could fail in the long run; it would always penetrate the market because more and more individuals would adopt it. Considering all these facts, it is obvious that the importance of Fleck's model is not to give advice for innovation policy or to forecast future chances of special innovations in penetrating the market, but to help clarify the decision-making process of potential adopters.

The third kind of models which will be discussed here are those describing the spatial diffusion of innovation (see for example, Lawrence Brown 1968). The basic assumption of that theory is, as its architect T.Hägerstrand 1952 stated, that diffusion and adoption of an innovation is a learning process of its potential adopters which is initiated by the persuasion of former adopters. The main barriers for the diffusion are distances between information carriers and potential adopters, potential adopters and places where they can buy the new commodity and non-adopters. The effect of mass-media, like TV, radio and newspapers, can be taken into account, too. The mathematical tools used is a random or biased net which is a set of nodes connected by one way edges which expand or contract over time. This means that more nodes are incorporated into the graph. After having divided the population which is located in the nodes of the net, into groups characterized by the adopting or non-adopting behavior, the information flow in the net can be described. This is done by computing probabilities for the events, that certain nodes are contacted in certain time levels. The diffusion (learning process) is then completely described on different aggregation levels, for example, single persons, households, farms or a whole country. The special facilities of a country or landscape can be incorporated into the model by net-functions.

Another spatial diffusion model uses the idea that diffusion of innovation behaves in a similar way to the spread of disease. This has been mentioned by many social scientists, for example see E. Rogers (1962). In that context epidemiological models can be used for the description of the diffusion of innovation, such as the epidemiological model described by Brown (1968) which divides the population located in a certain position into non-adopters, active adopters, and passive adopters with respect to a time level t and ends up in a system of integro-differential equations using the proportions of these population parts as state variables.

The common feature of the first and second kind of spatial diffusion processes is that 'learning' and spread of disease take place through interpersonal networks and that the process of transferring information has the same attributes as that of transferring diseases.

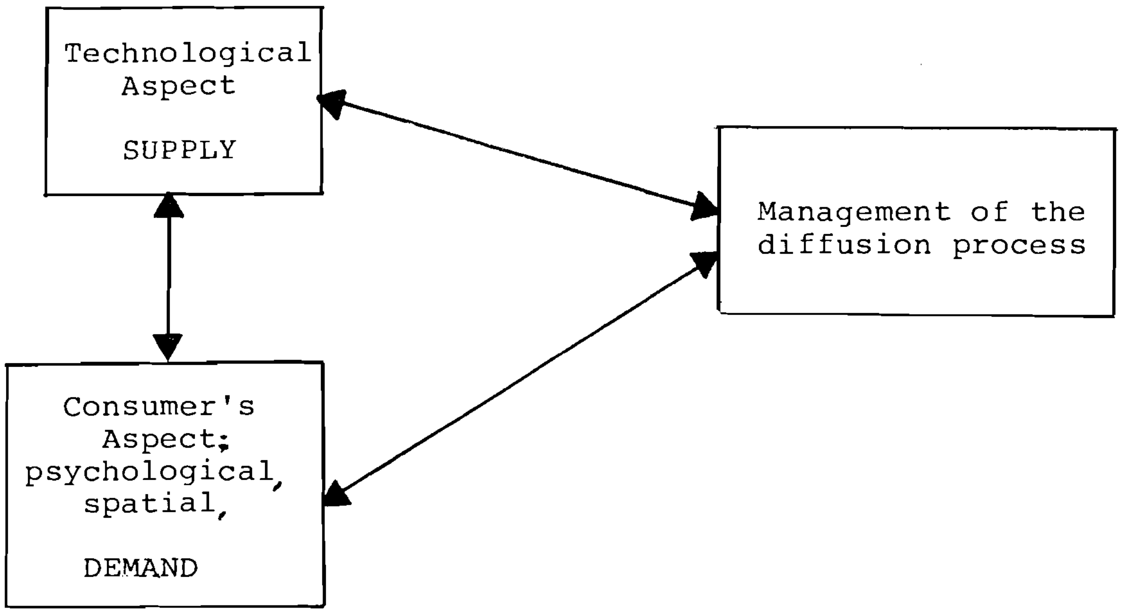
But what do these models of the diffusion process show? They show that this process is very complex in the sense that it consists of two different aspects:

-- the technological aspect reflecting the supply of the commodity produced by the new technology; and

-- the consumer's aspect reflecting the demand of the new commodity.

The interaction of these two aspects determines the economic success or failure of the innovation when it makes no difference if we talk about basic, improvement, product or process innovation.

One can show the diffusion process with the following diagram:



Of course there are supply as well as demand driven innovations as Maier and Haustein (1979) pointed out. So the diffusion process can also be supply or demand driven, but both aspects have to contribute in order to result in a successful innovation. Mostly process innovations are supply driven, and product innovations are demand driven.

Another issue indicated by the models, especially Peterka's substitution model, is that there seems to be two different phases of the diffusion process:

-- the initial phase, where the new technology has to try to gain a good starting point in the market. For that, great investment and support seems to be necessary so that the technology can gain a few percent of the market share.

-- the continuation phase, where the technology has to try to penetrate the market and to substitute other technologies.

The success or failure of the innovation seems to be mostly determined by the initial phase, that means the suppositions for a good take-off are made there and the development of the innovation (which became a new technology) over the continuation phase is mainly determined in the initial phase.

CONSTRUCTION OF AN OPTIMAL CONTROL SUBSTITUTION MODEL

Let us assume the following situation: there are two technologies T_0 and T_1, producing a commodity that fulfills the same need, competing for the market. Each one is represented by its production P_0 resp. P_1 and by its sales S_0 resp. S_1. The following arguments are quite similar to those in Peterka (1977). Assume that the two technologies want to increase their production from $P_i(t)$ to $P_i(t + \Delta t)$ $(i = 0,1)$. Therefore a certain investment

$$\alpha_i(t)[P_i(t + \Delta t) - P_i(t)] \quad , \qquad i = 0,1 \tag{1}$$

is necessary. $\alpha_i(t)$ is the capital needed to increase the production of the i-th technology at time t by one unit and is called specific investment. To cover this investment a certain amount of capital is necessary. One possibility is the capital accumulated by the producer in $[t,t + \Delta t]$

$$\int_t^{t+\Delta t} [S_i(\tau)(p_i(\tau) - b_i(\tau)) - c_i(\tau)P_i(\tau)]d\tau \tag{2}$$

where

- $p_i(t)$: price of i-th commodity at time t per unit
- $b_i(t)$: distribution costs of i-th technology per unit at time t
- $c_i(t)$: specific production costs of i-th technology per unit including amortization of material goods used in production at time t.

Another way to cover the necessary investment is external capital. Let $Q_i(t,\Delta t)$ be the capital which was given to the producer of of i-th technology in the time period $[t,t + \Delta t]$. But we take the external capital flow $q_i(t)$, which is defined by

$$Q_i(t,\Delta t) = \int_t^{t+\Delta t} q_i(\tau)d\tau \quad . \tag{3}$$

Another possibility of covering the investment is that the producer kept some capital in the past to invest in the future. We call this capital $\Delta_i(t,\Delta t)$. In this case $\Delta_i(t,\Delta t)$ is positive, but if the producer keeps some capital in the time period $[t,t + \Delta t]$ in order to invest it later, $\Delta_i(t,\Delta t)$ is negative. Therefore we can make a balance and get:

$$\alpha_i(t)[P_i(t + \Delta t) - P_i(t)] = \int_t^{t+\Delta t} [S_i(\tau)(p_i(\tau) - b(\tau)) - c_i(\tau)P_i(\tau)]d\tau + \Delta_i(t,\Delta t) \tag{4}$$

or

$$\int_t^{t+\Delta t} [\alpha_i(t)\dot{P}_i(\tau) - (S_i(\tau)(p_i(\tau) - b_i(\tau)) - c_i(\tau)\cdot P_i(\tau) - q_i(\tau))]d\tau = \Delta_i(t,\Delta t) \quad . \tag{5}$$

We can regard $\Delta_i(t,\Delta t)$ as a random variable and assume that it has a mean value of zero. Consequently equation (5) is a stochastic equation. Now we look at a certain realization of that stochastic equation:

$$\int_t^{t+\Delta t} [\alpha_i(t)\dot{P}_i(\tau) - (S_i(\tau)(p_i(\tau) - b_i(\tau)) - c_i(\tau)P_i(\tau) - q_i(\tau))]d\tau = 0 \quad . \tag{6}$$

This is now valid for all $t,\Delta t$ greater than 0. Therefore we conclude (relabeling t and τ)

$$\alpha_i(t)\dot{P}_i(t) = S_i(t)(p_i(t) - b_i(t)) - c_i(t)\ P_i(t) + q_i(t)$$

$$i = 0,1 \quad . \tag{7}$$

These two equations cover the technological side of the diffusion process. We are now going to look at the demand side of this process. We introduce the market share of T_i in terms of sales

$$f_i(t) = \frac{S_i(t)}{S_0(t) + S_1(t)} \; ; \qquad i = 0,1 \tag{8}$$

$$f(t) := f_1(t)$$

$$1-f(t) = f_0(t) \quad ,$$

and the total amount of sales (market size)

$$S(t) = S_0(t) + S_1(t) \quad .$$

As in Fleck (1978) we use a discrete Markoff chain in order to describe the customer's behavior and assume that the probability of an 'over age' customer owning the commodity produced by T_0 adopts the commodity produced by T_1 during the time period $[t - 1,t]$ has the following form:

$$Pr_{01}(t,t-1) = (\beta(t-1).g(\frac{p_1(t-1)}{p_0(t-1)}) + \delta(t-1).h(\frac{P_0(t-1)}{P_1(t-1)}))\cdot$$

$$\cdot(\rho(t-1)f(t-1) + \gamma(t-1)) \qquad (10)$$

where

$$g,h:[0,\infty] \rightarrow [0,1] \text{ and } g,h\downarrow 0 \quad \text{for } t\rightarrow\infty.$$

So this probability is the smaller the greater the ratios $\frac{p_1}{p_0}$ and $\frac{P_0}{P_1}$ are.

The weighted sum $\beta(t)g + \delta(t)h$ means that the consumer's will to buy the commodity produced by the new technology is a weighted average of the ratios of prices and productions of the two technologies. It expresses the fact that a commodity can be expensive if it has gained a big market share in terms of production because the consumer cannot really choose (the production of T_0 is small compared with that of T_1). So the coefficient $\delta(t)$ is a measure of the necessity of the kind of goods produced by T_0 and T_1. $\delta(t)$ is the greater the more necessary these goods are.

The coefficient $\beta(t)$ can be interpreted as a measure of the quality of the commodity produced by T_1, because it shows how much the customer cares about prices. Therefore δ will be small if T_1 produces a commodity with high standard quality and will be greater if the standard of quality is small. The coefficients ρ and γ have about the same meaning as it is explained in the previous chapter.

Using the same mathematical methods as Fleck did we can derive an ordinary differential equation for the market share (in terms of sales) of T_1:

$$\dot{f} = (1-f)(\beta(t)g(\frac{p_1(t)}{p_0(t)}) + \delta(t)h(\frac{P_0(t)}{P_1(t)}))(\rho(t)f + \gamma(t)) \quad . \qquad (11)$$

We can combine the differential equations (7) and (11) using

$$S_i(t) = f_i(t) \cdot S(t) \quad ; \quad i = 0,1 \qquad (12)$$

and gain:

$$\dot{P}_0 = -\frac{c_0(t)}{\alpha_0(t)} P_0 + (1 - f)\cdot\frac{S(t)}{\alpha_0(t)}\cdot(p_0(t) - b_0(t)) + $$

$$+ \frac{1}{\alpha_0(t)} q_0(t)$$

(13) $$\dot{P}_1 = -\frac{c_1(t)}{\alpha_1(t)} P_1 + f\cdot\frac{S(t)}{\alpha_1(t)}(p_1(t) - b_1(t)) + \qquad t_0 \leqslant t \leqslant T$$

$$+ \frac{1}{\alpha_1(t)} q_1(t)$$

$$\dot{f} = (1 - f)(\beta(t) g(\frac{p_1(t)}{p_0(t)}) + \delta(t) h(\frac{P_0(t)}{P_1(t)}))\ \cdot$$

$$\cdot(\rho(t) f + \gamma(t))$$

if $\alpha_i(t) \neq 0$ in $[t_0, T]$

with given initial values

$$P_0(t_0) = P_{00},\quad P_1(t_0) = P_{10},\quad f(t_0) = f^0 \qquad (14)$$

This is a nonlinear (because of the equation for $\dot{f}$!) first order system of three differential equations with given initial values.

Let us now think a little about these three state equations. The first equation describes the development of the old, the second the development of the new technology. So these two equations describe the technological aspect of the diffusion process, while the third describes the consumer's aspect or the societal aspect and the diffusion process as a whole is an interaction of these two aspects.

Knowing all the system's parameters and parameter functions we could solve this system of differential equations numerically, but we take a different way. We want to force T_1 to gain an optimal profit by the means of optimal price and investment policy. Assuming a given price and investment policy of T_0, a certain parameter (function) constellation concerning $\alpha_i(t)$, $c_i(t)$, $b_i(t)$, $\rho(t)$, $\beta(t)$, $\gamma(t)$, $\delta(t)$, P_{00}, P_{01}, f^0 and a given distribution of the market size $S(t)$ we can influence the system (13) by varying $p_1(t)$ (price of T_1 per produced unit) and $q_1(t)$ (external investment flow) within certain bounds. Let us first consider the case of the price $p_1(t)$. Of course there must be a minimum price (greater 0) because otherwise the technology would work completely without income, and on the other hand there could be governmental or other restrictions concerning the highest possible price; and if not we take a big enough number as a bound.

Looking at the investment which can be given to T_1 it is obvious that there must be an upper bound because no technology can have an unlimited amount of capital available and on the other hand there should be a minimum of investment promised to be given to T_1.

In order to express these thoughts mathematically, we write:

$$p_{min} \leq p_1(t) \leq p_{max} \quad , \quad t \in [t_0,T] \tag{15}$$

$$q_{min} \leq q_1(t) \leq q_{max} \quad , \quad t \in [t_0,T] \tag{16}$$

The second assumption means that

$$q_{min}(T - t_0) \leq Q_1(t_0,T - t_0) \leq q_{max}(T - t_0) \tag{17}$$

where $Q_1(t_0,T - t_0)$ is the investment that is given to T_1 during the time period $[t_0,T]$.

Now we have to specify mathematically what we want to reach, that is, highest possible profit:

total profit = profit of sales - costs of production -
- given investment (18)

or

$$G = \int_{t_0}^{T} (f(t)(p_1(t) - b_1(t))S(t) - c_1(t)P_1(t) - q_1(t))dt \tag{19}$$

Our aim is now to maximise G within the given restrictions (15) (16), following the system of differential equations (13)(14). Abstractly speaking we got a problem of the following kind:

$$\dot{x} = F(x,u,t) \quad \text{represented by (13)}$$
$$x(t_0) = x_0 \quad \text{represented by (14)}$$
$$u \in U \quad \text{represented by (15)(16)} \qquad (20)$$
$$G(u) = \int_{t_0}^{T} \tilde{L}(x,y,t)\,dt \quad \text{represented by (19)}$$
$$G(u) \to \max \quad \text{optimality condition}$$

where

$x = (P_0, P_1, f)^T$: state variables

$u = (p_1, q_1)^T$: control variables

U is the set of all piecewise continuous functions $u = (u_1, u_2)^T$ defined on $[t_0, T]$ fulfilling

$p_{min} \leq u_1 \leq p_{max}$

$q_{min} \leq u_2 \leq q_{max}$,

U is called the set of admissible controls.

$x_0 = (P_{00}, P_{10}, f^0)$

G(u) is the total profit ; performance index

The assumption "piecewise continuous" is convenient for the following mathematical analysis (see next chapter).

The problem (20) is called an optimal control problem with given restrictions on the controls. Mathematical tools for these kinds of problems, will be explained in the next chapter.

PONTRYAGIN'S MAXIMUM PRINCIPLE AND ITS APPLICATION TO THE PROFIT OPTIMAL DIFFUSION PROCESS

Consider the following general optimal control problem:

a) $\dot{x} = F(x,u,t)$; $t_0 \leq t \leq T$, with $x \in R^n, u \in R^m$

b) $x(t_0) = x_0$ (21)

c) $u \in U$

d) $G(u) = \int_{t_0}^{T} L(x,u,t)\,dt \to \min_{u \in U}$

That means that we look for an admissible control u and for a state vector x(t) for $t \in [t_0,T]$ fulfilling the differential equation (21a) and the initial condition (21b), which maximize the performance index G(u). Because of mathematical convenience we request G(u) to be minimized from now on (that means $L = -\tilde{L}$). In order to solve this optimization problem we use one of the various formulations of Pontryagin's Minimum principle which is very well described in M. Athans and P. Falb, (1966) starting from 5.11.

First we have to assume that the set of all admissible controls U is constructed in such a way that every control in U has to be piecewise continuous on $[t_0,T]$ that means that the set of discontinuities is numerable and that at each of these points u has a finite limit from the right and the left, and further that every component of the vector u has to be bounded as a function on $[t_0,T]$, that means that $u(t)\in\Omega$ for all t in $[t_0,T]$ where Ω is a bounded subset of R^m.

Furthermore we assume that $F(x,u,t)$, $\frac{\partial F}{\partial x}(x,u,t)$ and $\frac{\partial F}{\partial t}(x,u,t)$ ($\frac{\partial F}{\partial x}$ denotes the Jacobian of F with respect to the vector x if F is a vector-valued function and it denotes the gradient vector of F with respect to x if it is a scalar function) are together with $L(x,u,t)$, $\frac{\partial L}{\partial x}(x,u,t)$ and $\frac{\partial L}{\partial t}(x,u,t)$ continuous on $R^n \times \bar{\Omega} \times [t_0,T]$ ($\bar{\Omega}$ denotes the closure of Ω).

The next step is that we form the function:

$$H(x,y,u,t) = L(x,u,t) + \langle y,F(x,y,t)\rangle \qquad (22)$$

which is called the Hamiltonian of the system (21) ($\langle , \rangle$ indicates the scalar product in R^n). y is called the costate vector of the system. Now we can formulate the Minimum principle in a form which is useful for our purposes.

Let $u^*(t)$ be an admissible control and let $x^*(t)$ be the trajectory of the system (21a) corresponding to $u^*(t)$ and fulfilling the initial condition (21b). In order that $u^*(t)$ be optimal for the functional (21d) it is necessary that there exists a function $y^*(t)$ such that

$$\dot{y}^*(t) = -\frac{\partial H}{\partial x}(x^*(t),y^*(t),u^*(t),t) \qquad (23)$$

$$y^*(T) = 0 \qquad (24)$$

and the function $H(x^*(t),y^*(t),u,t)$ has an absolute minimum as a function of u over Ω at $u = u^*(t)$ for $t \in [t_0,T]$; that is

$$\min_{u\varepsilon\Omega} H(x^*(t),y^*(t),u,t) = H(x^*(t),y^*(t),u^*(t),t) \qquad (25)$$

Please observe that this condition is necessary for optimality but not sufficient!

In the following part of this chapter we are going to apply this principle to our special problem (13),(14),(15),(16) and (19).

We notice that the first thing we have to do is to find the optimal control $u^* = (p_1^*, q_1^*)$. So we form the Hamiltonian H:

$$H(P_0,P_1,f;p_1,q_1;t) = - f(p_1 - b_1(t))S(t) + c_1(t)P_1 + q_1 +$$

$$+ y_1(\frac{1}{\alpha_0(t)}q_0(t) - \frac{c_0(t)}{\alpha_0(t)} P_0 + (1 - f)\frac{S(t)}{\alpha_0(t)}(p_0(t) - b_0(t)) +$$

$$+ y_2(\frac{1}{\alpha_1(t)}q_1(t) - \frac{c_1(t)}{\alpha_1(t)} P_1 + f\frac{S(t)}{\alpha_1(t)}(p_1 - b_1(t)) +$$

$$+ y_3(1 - f)(\beta(t)g(\frac{p_1}{P_0(t)}) + \delta(t)h(\frac{P_0}{P_1}))(\rho(t)f + \gamma(t)) \quad . \qquad (26)$$

From the condition (25) of the Minimum principle it follows that

$$H(x^*(t),y^*(t),u^*(t),t) \leq H(x^*(t),y^*(t),u(t),t) \qquad (27)$$

for all $u \varepsilon U$.

In our case (the argument t is omitted in most cases from now on):

$$H(P_0^*,P_1^*,f^*;y_1^*,y_2^*,y_3^*;p_1^*,q_1^*;t) \leq H(P_0^*,P_1^*,f^*,y_1^*,y_2^*,y_3^*;p_1,q_1;t) \qquad (28)$$

for all $p_{min} \leq p_1(t) \leq p_{max}$

and $p_{min} \leq q_1(t) \leq q_{max}$

After some calculations we get:

$$p_1^*\cdot[-f^*\cdot S + y_2^*\cdot\frac{S}{\alpha_1}\cdot f^*] + y_3^*\cdot\beta g(\frac{p_1^*}{p_0})(1 - f^*)(\alpha f^* + \beta) + q_1^* +$$

$$+ q_1^*[\frac{y_2^*}{\alpha_1}] \leq p_1\cdot[-f^* S + y_2^*\cdot\frac{S}{\alpha_1}\cdot f^*] + y_3^*\beta g(\frac{p_1^*}{p_0})(1 - f^*)\cdot(\rho f^* + \gamma)$$

$$+ q_1^*\cdot[\frac{y_2^*}{\alpha_1} + 1] \quad . \tag{29}$$

That means that (p_1^*, q_1^*) minimizes the function

$$E(p_1,q_1) = p_1[- f^*\cdot S + y_2^*\cdot\frac{S}{\alpha_1}\cdot f^*] + y_3^*\cdot\beta(1 - f^*)\cdot(\rho f^* + \gamma)g(\frac{p_1}{p_0}) +$$

$$+ q_1\cdot[\frac{y_2^*}{\alpha_1} + 1] = E_1(p_1) + E_2(q_1) \quad . \tag{30}$$

Because we could separate the variables p_1 and q_1 we can apply the following rule:

$$\min_{\substack{p_{min} \leq p_1 \leq p_{max} \\ q_{min} \leq q_1 \leq q_{max}}} E(p_1,q_1) = \min_{p_{min} \leq p_1 \leq p_{max}} \{p_1[- f^*\cdot S + y_2^*\cdot\frac{S}{\alpha_1}\cdot f^*] +$$

$$+ y_3^*\beta(1 - f^*)\cdot(\rho f^* + \gamma)g(\frac{p_1}{p_0})\} +$$

$$+ \min_{q_{min} \leq q_1 \leq q_{max}} \{q_1[\frac{y_2^*}{\alpha_1} + 1]\} \quad . \tag{31}$$

It is obvious that the function:

$$q_1^*(t) = \begin{cases} q_{min} & \text{if} \quad \frac{y_2^*(t)}{\alpha_1(t)} + 1 > 0 \\ q_1^*(t-0) & \text{if} \quad \frac{y_2^*(t)}{\alpha_1(t)} + 1 = 0 \\ q_{max} & \text{if} \quad \frac{y_2^*(t)}{\alpha_1(t)} + 1 < 0 \end{cases} \tag{32}$$

minimizes $E_2(q_1)$ assuming that $\left(\frac{y_1^*(t)}{\alpha_1(t)} + 1\right)$ vanishes only in distinct points of the interval $[t_0, T]$.

Now we have to minimize that part of $E(p_1, q_1)$ that depends on p_1, which has the following principle structure:

$$E_1(p_1) = \tilde{\alpha} p_1 + \tilde{\beta} g\left(\frac{p_1}{p_0}\right) \quad (33)$$

where

$$\tilde{\alpha} = f^* \cdot S \cdot \left(\frac{y_2^*}{\alpha_1} - 1\right) \text{ and } \tilde{\beta} = y_3^* \beta (1 - f^*)(\rho f^* + \gamma) \quad (34)$$

Let us first assume that $\tilde{\alpha} \leq 0$ and that $\tilde{\beta} > 0$. $E_1(p_1)$ is a monotonically decreasing function of p_1, because g is decreasing, and a function of that kind takes its minimum at its right boundary value, which is p_{max}. In the opposite case ($\tilde{\alpha} \geq 0$ and $\tilde{\beta} < 0$) we get an increasing function of p_1 and it takes its minimum at p_{min} (left boundary).

Now two cases are left: a) $\tilde{\alpha} \leq 0$, $\tilde{\beta} \leq 0$

b) $\tilde{\alpha} \geq 0$ $\tilde{\beta} \geq 0$

In these cases $E_1(p_1)$ can take its minimum on the boundary as well as in the interior of the interval $[p_{min}, p_{max}]$. Assuming that the first order derivative of g exists on $[p_{min}, p_{max}]$ and is continuous, we get using that

$\bar{p}_1^*$ is defined by the zero of the equation

$$g'\left(\frac{\bar{p}_1^*}{p_0}\right) = -\frac{p_0 \tilde{\alpha}}{\tilde{\beta}}, \quad \tilde{\beta} \neq 0 \quad (35)$$

and

$$m: = \begin{cases} \min(E_1(p_{min}), E_1(P_{max}), E_1(\bar{p}_1^*)), & \text{if } \bar{p}_1^* \in [p_{min}, p_{max}] \\ \min(E_1(p_{min}), E_1(p_{max})), & \text{if } \bar{p}_1^* \notin [p_{min}, p_{max}] \end{cases}$$

the following optimal p_1^*:

$$p_1^*(t) = \begin{cases} p_{min} & \text{, if } \tilde{\alpha} \geq 0 \text{ and } \tilde{\beta} < 0 \text{ or } \tilde{\alpha} > 0 \text{ and } \tilde{\beta} \leq 0 \text{ or} \\ & \text{otherwise if } m = E_1(p_{min}) \\ p_{max} & \text{, if } \tilde{\alpha} \leq 0 \text{ and } \tilde{\beta} > 0 \text{ or } \tilde{\alpha} < \hat{0} \text{ and } \tilde{\beta} \geq 0 \text{ or} \\ & \text{otherwise if } m = E_1(p_{max}) \\ \bar{p}_1^* & \text{, if } \tilde{\alpha} > 0 \text{ and } \tilde{\beta} > 0 \text{ or if } \tilde{\alpha} < 0 \text{ and } \tilde{\beta} < 0 \\ & \text{and if } m = E_1(\bar{p}_1^*) \\ p_1^*(t-0) & \text{, if } \tilde{\alpha} \doteq 0 \text{ and } \tilde{\beta} = 0 \end{cases} \tag{37}$$

again assuming that $E_1(p_1^*(t))$ vanishes only in distinct points of the interval $[t_0,T]$. The optimal control (p_1^*,q_1^*) is a switching function.

Lastly we have to calculate the so-called adjoint equation (23) with boundary condition (24). We arrive at this by using the gradient of the Hamiltonian:

$$\dot{y}_1^* = \frac{c_0}{\alpha_0} y_1^* - y_3^* (1 - f^*)(\rho f^* + \gamma)\cdot\delta h'(\frac{P_0^*}{P_1^*})\frac{1}{P_1^*}$$

$$\dot{y}_2^* = -c_1 + \frac{c_1}{\alpha_1}\cdot y_2^* + y_3^* \cdot (1 - f^*)\cdot(\rho f^* + \gamma)\delta h'(\frac{P_0^*}{P_1^*})\cdot\frac{P_0^*}{(P_1^*)^2}$$

$$\dot{y}_3^* = (p_1^* - b_1)s + y_1^* \frac{s}{\alpha_0}(p_0 - b_0) - \frac{s}{\alpha_1}\cdot(p_1^* - b_1)\cdot y_2^* - $$
$$- y_3^*(\beta g(\frac{p_1^*}{p_0}) + \delta h(\frac{P_0^*}{P_1^*}))(\rho - \gamma - 2\rho f^*) \quad . \tag{38}$$

$$y_1^*(T) = y_2^*(T) = y_3^*(T) = 0 \tag{39}$$

Next we substitute the optimal control (p_1^*,q_1^*) defined by (32), (33) - (37) for (p_1,q_1) in the equation (23) and put this optimal control in the equation (38). Then with the initial conditions (24) and the boundary conditions (39) we got a six-dimensional two-point-boundary value problem on the interval $[t_0,T]$ which is nonlinear. The only way to obtain a solution of that problem, if one exists, is numerically, for example, by using combined multiple shooting and extrapolation methods (Diekhoff, et al. (1977)) or collocation methods (see R. Weiss (1974)) with attention to the possible discontinuities.

REFERENCES

Athans, M., and P. Falb. 1966. Optimal Control : An Introduction to the Theory and its Applications. New York: McGraw-Hill

Brown, L. 1968. Diffusion Dynamics. Review and Revision of the Quantitative Theory of the Spatial Diffusion of Innovation. Lund Studies in Geography, Ser. B, Number 29.

Diekhoff, H.-J., P. Lory, H.J. Oberle, H.-J. Pesch, P. Rentrop, R. Seydel. 1977. Comparing Routines for the Numerical Solution of Initial Value Problems of Ordinary Differential Equations in Multiple Shooting. Technische Universität München, Institut für Mathematik. Bericht 7520.

Hägerstrand, T. 1952. The Propagation of Innovation Waves. Lund Studies in Geography, Ser. B, Number 4.

Haustein, H.-D., and H. Maier. 1979. Basic, Improvement and Pseudo-Innovations and Their Impact on Efficiency. WP-79-96. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Peterka, V. 1977. Macrodynamics of Technological Change: Market Penetration by New Technologies. RR-77-22. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Peterka, V., F. Fleck, et al. 1978. The Dynamics of Energy Systems and the Logistic Substitution Model. AR-78-1A,B,C. Laxenburg, Austria: International Institute for Applied Systems Analysis. Report prepared for Stiftung Volkswagenwerk.

Spinrad, B.I. 1979. A Generalized Model for Market Substitution. WP-79-53. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Weiss, R. 1974. The Application of implicit Runge-Kutta and Colloocation Methods to Boundary Value Problems. Math. Comp. 28.

APPENDIX A

The basic model (13),(14) which describes the diffusion process assuming a certain scenario determined by a parameter constellation was implemented in FORTRAN and tested on the IIASA PDP-11 computer. This model is posed as a three-dimensional nonlinear dynamical system with the state variables production of old and new technologies and market share of new technology in terms of sales. Numerically an initial value problem had to be solved. This was done by using the IMSL Routine DREBS which employs an extrapolation method.

Tests were performed with many different scenarios and all results showed some common features. The marketshare of the new technology showed roughly an s-shaped form (therefore the model is qualitatively consistent with the Fisher-Prey model). Furthermore the state variables converged to certain saturation limits which depend heavily on the parameters. These limits were 'reached' at about the same time point by the production quantities as well as the market share. In a success case (which means that the market share of new tech exceeds a reasonable percentage) the production of new tech is an increasing function of the time and the production of old tech is either decreasing or it increases for a short period of time and decreases after that.
Let us now discuss the two program runs which are shown in Figure 1. Exponentially decreasing functions h and g were used for both runs. The following scenario was assumed for the first run (full lines in Figure 1) :

production costs of old technology per unit	= 3800.
production costs of new technology per unit	= 3600.
specific investment for old technology	= 850.
specific investment for new technology	= 950.
distribution costs of old technology per unit	= 500.
distribution costs of new technology per unit	= 490.
unit price of old technology over time	= 8000.
unit price of new technology over time	= 9000.
investment flow of old technology over time	= 1000000.
investment flow of new technology over time	= 120000000.
marketsize over time	= 4000.
measure of necessity of goods produced by the old and new technologies	= 0.9
measure of the quality of goods produced by the new technology	= 0.7
measure of the innovativeness of the average consumer	= 0.08
measure of the attractivity of the new technology	= 0.55

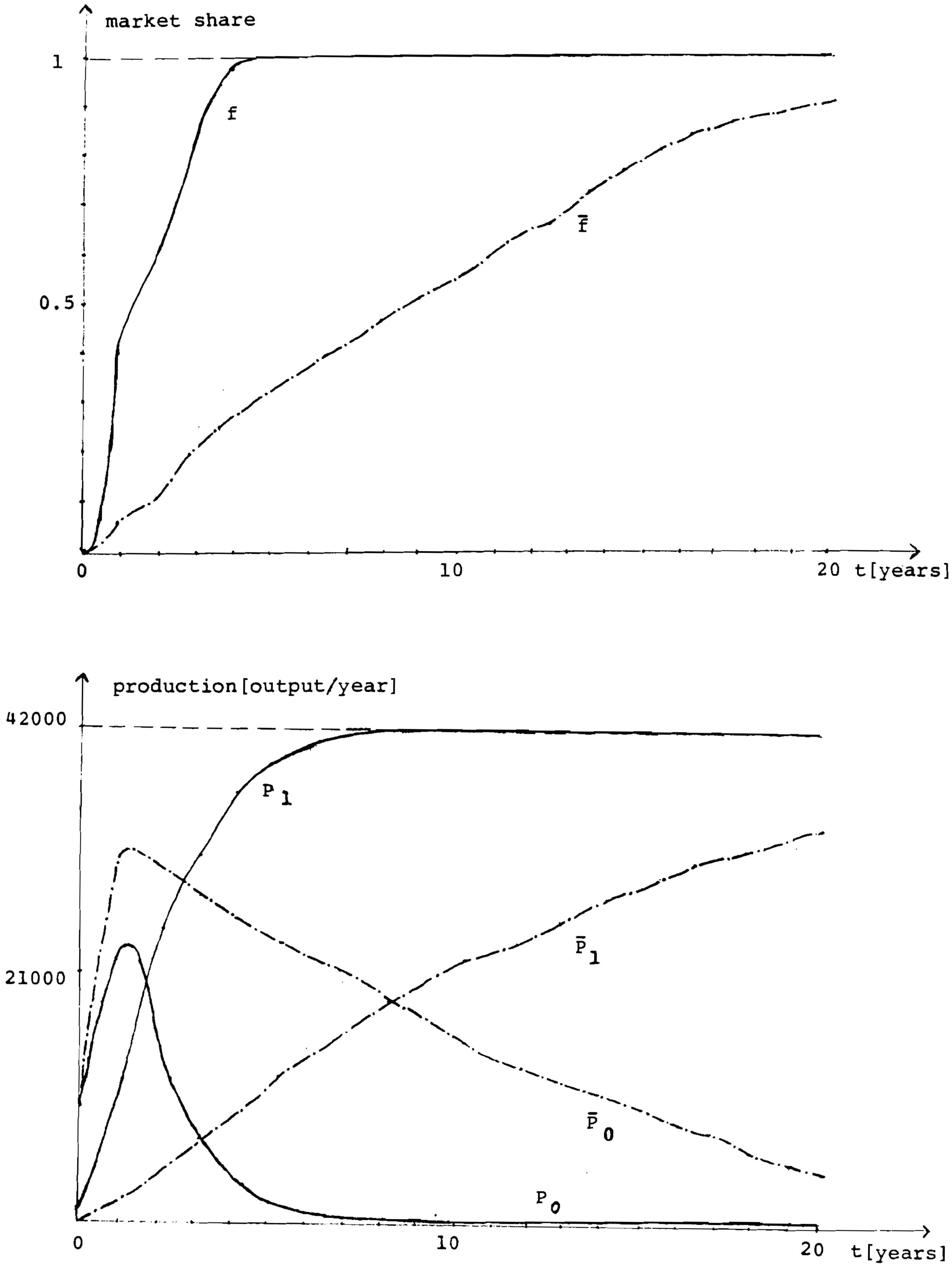


Figure 1. Computed curves

The calculated profit is : -3426474.0

This indicates that a bad investment and price policy was chosen. Looking at the scenario we can easily find out the policy-mistakes. The market size in terms of sales is less than a tenth of the upper bound of the production quantity of the new technology. Too many units are produced, the stock is growing rapidly and therefore the loss gets very large over 20 years. The major driving force for the production is the investment which is given to the new technology at a constant rate. So a better investment policy would be to invest heavily in the first years and to cut down investment rates afterwards. This would imply a large decrease of the production of new tech in the first time but would still keep its absolut value reasonably low.

For the second run which is indicated by lines composed of dots and slashes in Figure 1 the last four parameters of the scenario were changed to 0.6, 0.3, 0.08, 0.1 (in this sequence). The quantities influencing the equations for the production quantities (the technological part of the scenario) remained unchanged. The calculated profit is : -7872268. The situation got worse because we decreased the market attractivity of the new technology without changing the investment and price policy. But even assuming this bad scenario a long run success seems possible if a better policy is chosen.