TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

D I S S E R T A T I O N

Optimal Control of a Two-State Diffusion Model:

Susceptibles and Adopters in Marketing and Drug Consumption

ausgeführt zum Zwecke der Erlangung des akademischen Grades einer Doktorin der Sozial- und Wirtschaftswissenschaften unter der Leitung von

Em.O.Univ.Prof. Dr.phil. Gustav Feichtinger
E105 Institut für Wirtschaftsmathematik

eingereicht an der Technischen Universität
Fakultät für Informatik

von

Mag.rer.soc.oec. Andrea Seidl
Matrikelnr. 0025151
Wienerstr. 230
3400 Klosterneuburg

Wien, am 17. Februar 2009

Kurzfassung

Bei Diffusionsmodellen geht es darum wie sich eine bestimmte Innovation, Idee oder ein Produkt sich unter den Mitgliedern einer Gesellschaft über die Zeit verbreitet (vgl. Rogers, 2003). Ein zentraler Punkt hierbei ist die Interaktion zwischen potentiellen und tatsächlichen Nutzern einer Sache, denn oft entscheiden Erfahrungen und Eindrücke dieser beider Gruppen über den Erfolg eines Produktes.

In dieser Dissertation wird ein Diffusionsmodell betrachtet, das aus zwei Zustandsvariablen, der Anzahl potentieller bzw. tatsächlicher Nutzer eines Produktes besteht. Mit Hilfe der optimalen Steuerungstheorie wird basierend auf der Dynamik dieser Zustände ein Maximierungsproblem im Anwendungsbereich Marketing behandelt. In diesem hat ein Entscheidungsträger die Möglichkeit das Ergebnis mittels Preisreduktionen, die das Produkt attraktiver machen, sich allerdings negativ auf den Profit auswirken, zu beeinflussen. Ohne Verwendung konkreter empirischer Daten wird gezeigt dass die optimale Lösung sowohl vom Startpunkt wie auch von den Parametern abhängt und komplexeres Verhalten wie DNSS-Kurven und Grenzzyklen auftreten kann.

Dieses Marketing Problem wird in weiterer Folge durch eine zweite Stufe erweitert, in der sich der einstige Monopolist plötzlich mit perfektem Wettbewerb konfrontiert sieht. Es wird gezeigt, dass wenn Konkurrenz nur negative Auswirkungen für den Entscheidungsträger bringt, würde er versuchen diesen Status so lange wie möglich zu erhalten. Allerdings wird auch gezeigt, dass der Umschaltzeitpunkt unter bestimmten Vorraussetzungen optimal bestimmt werden kann, insbesondere wenn besonders aggressiv gehandelt werden kann um möglichen Mitbewerben den Markteintritt zu verwehren.

Weiters wird ein Minimierungsproblem im Anwendungsbereich Drogenpolitik behandelt, im dem ein Entscheidungsträger mittels entsprechender Maßnahmen die Möglichkeit hat das durch Drogenkonsum entstehende Leid zu reduzieren. Solche Maßnahmen haben allerdings die Auswirkung, dass Drogenkonsum für potentielle Nutzer attraktiver wird, und sich so das durch

Drogen entstehende Leid wieder vergrößert. Wieder wird gezeigt, dass die optimale Strategie vom verwendeten Startpunkt und den Parametern abhängt und Grenzzyklen und schwache DNSS Kurven auftreten können.

Abstract

A diffusion process considers how and why a certain good, idea or innovation spreads over time among the members of a society (cp. Rogers, 2003). Crucial in this process is the interaction between potential and actual users of a good, because often the experiences and impressions of the members of these groups decide whether a product becomes successful or not.

In thesis a diffusion model is introduced, which consists of two states describing the number of actual and potential users of a good respectively. Applying tools from optimal control theory, a maximization problem in the field of marketing is considered. It is discussed how a decision maker can influence the development of the two states by giving price reductions, which make the product more attractive to potential customers, however, lead to smaller profits per user. Without usage of concrete empirical data, it will be shown, that the optimal solution depends on the initial values of the state variables as well as on the used parameters and that rather complex behavior such as limit cycles and DNSS curves can occur.

The marketing problem then is extended by including a second stage, in which the former monopolist suddenly has to face perfect competition. It is shown that if competition only has negative consequences for a decision maker, he would try to remain monopolist as long as possible. Under certain circumstances, one can determine the switching point between the stages optimally, i.e. when the decision maker can act particularly aggressive to deter market entry of possible competitors.

Further, a minimization problem in the field of drug policy is discussed, where a decision maker can reduce harm caused by drug usage through certain measures. However, such measures make the drug more attractive for potential consumers and more harm arises through a larger number of users. Again it will be shown that the optimal solution depends on the initial values of the two states and the used parameters and that limit cycles and (weak) DNSS curves can be found.

Preface

The aim of this thesis is to give some insights on how to optimally control a model describing the influence of the number of potential and actual adopters on the spreading of a certain good over time. While there are certainly more applications for the proposed dynamical system, this work focuses on marketing and drug policy. Without the use of concrete data, it is investigated which outcomes are possible under certain conditions in order to see what might be explained with the help of such a model.

This work mainly consists of two parts, which can be basically read independently of each other: In the first part a maximization problem in the field of marketing is presented, the second part deals with a minimization problem, considering the impact of harm reduction measures. The results of this thesis are found with the help of optimal control theory; a short overview of the used terminology can be found in the appendix of this work. The necessary numerical calculations are done with the help of the MATLAB toolbox OCMat, which was developed to adequately deal with optimal control problems; some remarks on the ideas, on which these calculations are based, are briefly described in this thesis together with the according results. The main capabilities of the toolbox are presented in the appendix.

The first chapter of this thesis will provide a basic introduction including a description of the underlying dynamics of the model and give two short, introductory examples of how a result of such an optimal control may look like. In the second chapter a marketing model will be presented and analyzed, where the decision maker can influence his profits by giving price reductions; in the third chapter this model will be extended to consist of two stages. In the first stage the provider of the good is a monopolist, and has to face perfect competition in the second.

As the spreading of a product is not the only application that can be described with the help of the used dynamical system, a minimization problem concerning the spreading of a good will be analyzed in the fourth chapter, where the control instruments are harm reducing measures and a particular focus will be put on the role of the used objective function.

In the appendix the reader can find more detailed information about how the results of the work were derived, providing a short introduction on optimal control theory. It also contains a short description of how one can use the OCMat toolbox to find the numerically calculated results and how this toolbox actually works.

Contents

Chapter 1

Introduction

1.1 A Two-State Diffusion Model

In a diffusion process an idea, practice or object is communicated "through certain channels over time among the members of a social system" (cp. Rogers, 2003). Diffusion can be considered as a special type of communication distinguished by spreading a new[1] idea, good, etc. which can have an impact on a social level, sometimes smaller, sometimes larger in significance.

The fact that information exchange among people can be crucial for the success or failure of a certain good, idea or technology and has to be taken into account when trying to influence its spreading, has lead to a large number of diffusion models in many areas such as marketing, technology adoption, drug and crime policy, fashion and cultural development.

Since people are often influenced in their decision on whether to purchase a product or not by other people, who already use it, there is a big interest in marketing to gain more insights into diffusion processes in order to be able to better understand the development of sales and the ways to influence them in a reasonable way (cp. Mahajan et al., 1990).

In order to influence the spreading of a product and to maximize the profit, a decision maker has various instruments (e.g. advertisement, price), yet the question arises how to optimally use them. Due to the dynamic character of such problems some answers can be provided with the application of optimal control theory. Good introductions into this field are provided, e.g., by Grass et al. (2008); Feichtinger & Hartl (1986); Léonard & Long (1992).

[1]New does not mean in this context that this idea or good has had to be recently developed or introduced, it rather concerns the consideration of the adoption decision (cp. Rogers, 2003).

As sometimes key aspects concerning the problem might change abruptley, such as the underlying parameters, one aim of this thesis is to analyze how this affects the optimal solution. Therefore, the model will be extended to consist of two stages, one where it is assumed that the decision maker has a monopoly and in a second one faces competition.

Another field considered in this thesis in which diffusion plays an interesting role is drug policy. The harm caused by the usage of illicit drugs is not only present to people directly affected, but impacts society itself on a social and economic level. There is again a diffusion process; a drug might or might not spread among the members of a certain social system and one of the main reasons why people start taking drugs is, because they want to imitate some existing users, e.g., due to some social pressure.

It must be a policy maker's objective to minimize the harm caused by drug users. Usually this is done by trying to keep the number of drug users as small as possible. Another plausible, but quite controversial approach toward this problem, which will be considered in this thesis, is to reduce the harm caused by drug usage itself, e.g., by measures such as needle exchange or drug substitution programs (cp. MacCoun et al., 1996; Lenton & Single, 1998). Yet harm reduction might make people believe, to a certain extend at least, that it is safer to take drugs and therefore they might be more easily or more frequently willing to do so (cf. MacCoun, 1998). This, for example, has been shown to be true for cigarettes.

The underlying dynamic, autonomous system used in this thesis has already been considered in a marketing/fashion context in Seidl (2005) and drug diffusion context in Caulkins (2004), but without any control variable to optimally influence the outcome. There are two states: the potential consumers, the so-called susceptibles, and the consumers of a certain good. In marketing models a distinction between these groups of people is certainly not new, however in many models (e.g., Gould, 1970; Muller, 1983) population (or market) size is assumed to be fixed at a certain level. Therefore they do not include the susceptibles as separate state variable and consider the potential clientel as those who have not adopted the product yet. A model where this was not the case was introduced by Feichtinger (1992), however, the slightly different dynamical system was not optimally controlled. The approach to include susceptibles in drug policy models is relatively new, Wallner (2008) showed that it makes sense to do so using data from Australia and the U.S. concerning cocaine to validate a closely related model.

The social interaction between potential and actual adopters of a good is crucial for the development of their sizes, because, although there might be a number of susceptibles, which decides independently of others to use a certain good, people are often encouraged to imitate others due to some information exchange. Based on an earlier edition of Rogers (2003) such a distinction was made in a mathematical diffusion model in Bass (1969), one of the most important articles in marketing literature, a model that has found many applications and extensions (cp. Bass, 2004).

In this thesis one control variable is considered, by which a decision maker has some influence on the number of users (and therefore also on the number of susceptibles) and their effect on his or her objective. In the marketing model, where it is the decision maker's objective to maximize the profit, he can influence the spreading of the product by advertisement and giving price reductions. In the harm reduction model, in which the goal is to minimize costs arising from drug usage, he can decide about the percentage reduction of the arising harm. In order to be able to provide a plausible economical interpretation, it is necessary to include control constraints meaning that the control always has to be non-negative and must not exceed one.

Since the use of the control does not only have positive effects on the outcome (costs can arise that might outweigh the utility of the application of the control), the objectives of this work are to look at the optimal application of the control, its effect on the diffusion process, at the shapes of these optimal solutions and on possible choices that a decision maker might or might not have. The question then arise on whether does an optimal solution still have the same shape if different premises are considered in terms of looking at what happens if some parameters (e.g., the strength of the influence of the users on susceptibles) or functions (e.g., the objective function) change.

Applying Pontryagin's Maximum Principle, the optimal system, its steady states and their stability properties will be derived analytically as far as possible, but mainly the results of this models will be calculated numerically with some existing and newly developed programs in the MATLAB environment (OCMat, see App. B, and MATCONT, see Dhooge et al., 2003). The parameters used are not based on concrete data, but are chosen in a way that some general economic interpretations can be made in order to gain some insights about optimal application of the control under different premises.

It will be shown that the shape of the optimal solution depends on the initial number of susceptibles and users and that it is necessary for a decision maker to adapt the use of the control over time depending on the current number of potential and already existing users, but also that he might sometimes have more than one option when considering the optimal strategy. For example he might have the option to decide to do little, save costs and have little consumers of a good or decide to spent much and have many users. For some functions a cycle can occur for certain parameters: in some phases it is then optimal to spend much in others to spend little on advertising or harm reduction (because there are sometimes many users, sometimes only a few). We will see that because of the magnitude parameters, it might sometimes be best to apply a strategy that is possibly a bit controversial, e.g. accepting many drug users or not promoting a product, because it would lead to no profit anyway.

It will be an issue of this thesis to show and compare different optimal solutions due to different parameters and functions, which might in some cases only differ in the magnitude of the control and the number of susceptibles and users, but in other cases have completely different shapes. It will be shown that the change of some parameters has a larger impact to the optimal solution than the change of others. The economic interpretation in the fields of marketing and drug policy / harm reduction of such aspects of the models will always play a central role.

1.2 Model Description

With the help of a diffusion model, one can try to follow the adoption of a certain good among members of a social system and explore why it spreads in a certain way or why it fails to do so.

The model described in this thesis consists of two states: the *susceptibles* (potential consumers) S and the adopters (consumers/customers/users) A of a certain good. We will see in this thesis that these dynamics makes sense both in the fields of marketing as well as in drug research. The incentive behind considering the consumers of a good is pretty obvious: it is them who create the profits of a product by purchasing it, but it is also them who cause the harm through drug usage.

The inclusion of susceptibles into the consideration allows to single out a more homogeneous group of people from a heterogeneous total population. The common attribute among this kind of people in this framework is that it is them, who are attracted to the purchase and/or usage of the good. Reasons why some are more inclined to the usage of a certain product can be found in age, income, education, religion, social commitment, peer pressure, location and many more, often reflecting that information exchange happens most effective among people who share the same background and/or interests (cmp. Rogers, 2003, Chap. 8). However, such a homophily among people might also prevent the spreading of a certain good, particularly if people only start using a product if others do so too. Therefore, one can gain valuable insights by including such people and their development into consideration.

It is also possible in this model to consider a distinction between different kinds of adopters (without introducing a third state). Defined by Bass

(1969), a central work in the field of diffusion models in marketing literature, one can differentiate between *innovators* and *imitators* among those who decide to adopt a product. Innovators are individuals who make their decision to adopt/purchase/use a certain good independently of other members of the considered social system; imitators, however, are influenced in their adoption decision (and in its timing) by already existing users of this product. It certainly might be a hard task to identify members of both groups, however there are certain characteristics correlated to innovativeness such as education, income, etc. (see Lilien & Kotler, 1983; Rogers, 2003), which increase or decrease the likelihood of being an innovator. The influence that imitators are exposed to, which makes the product more attractive to them, can get across people by direct communication - people share their experience about the product via word-of-mouth and/or might even introduce others to the usage of the product by letting them try it - or by observability, i.e. people can see the benefits of using the product, may it be the capabilities or only the status associated with the good. Other product characteristics might also play a big role when it comes to its diffusion, such as compatibility (with society or with other goods - particularly when considering network goods, i.e., products of which the utility increases with the number of people using it), availability, complexity or other benefits (see Lilien & Kotler, 1983).

As described in Rogers (2003) there are incentives of different types to spread a good: Not only an adopter might have reasons to start using the product, because of the expected advantages of usage, the users also can have incentives to persuade susceptibles, e.g., it might increase their own utility of the good, particularly when network effects play a role or they might even receive a premium (such particular incentives are, however, not included in the models here). Advantages from product adoption might not only arise for the adopters themselves but also for a (social) system / company they use this good for and the advantages might not occur immediately but only after some time.

The state dynamics used in this thesis has already been considered by Seidl (2005) in a marketing relevant context, trying to explain the development of fashion goods and in Caulkins (2004) considering a very basic model regarding the spreading of a drug (cocaine in the US). However, these models were not optimally controlled, which changes in this thesis, which aims at exploring how and why the optimal application of different control instruments influences development of the two states in a certain way.

The state equations used in this thesis are as follows:

$$\dot{S} = k - \delta S - f(A)Sg(v) \tag{1.1}$$

$$\dot{A} = f(A)Sg(v) - \mu A \tag{1.2}$$

The parameter k describes the inflow to the susceptibles per time unit. The number is constant and describes the number of people of a country/city/ethnic group etc. who start to fulfill certain criteria, due to which they might be attracted to use the described good (e.g. reaching a certain age, becoming exposed to a certain social surrounding, etc.). The percentage of susceptibles quitting the system without ever using the product is given by parameter δ, reasons for this might be reaching a certain age, dying or maybe also simply leaving the surrounding.

The function $f(A)$ is crucial regarding the flow between the two states, since it reflects the interaction between users and susceptibles. Also called *initiation function*, it gives the percentage of susceptibles which start to use a certain good per time unit. As previously described, people are often introduced by other people to the usage of a good. In this work, it is assumed that $f_A > 0$ meaning that on the one hand it reflects that the overall perception of using the good is positive and potential consumers are attracted to the product because of existing users. On the other hand if the number of users is big, the consumers exercise a larger influence or pressure on the susceptibles to imitate their behavior as if there would only be a few of them. The shape of this function is always assumed to be $f(A) = a + bA^{\alpha}$ with $\alpha > 1$ in this work. This convexity is supposed to reflect that a person is more likely to start becoming consumer of a certain good if there are many people already using it and that it even becomes very hard to escape a certain pressure of adopt a certain product if everyone else uses it. Note, however, that for some goods initiation might be more appropriately described by a concave function, for example if attractivity of a product is not so big when everyone else has it (e.g. certain fashions in clothing) or if the need to also own such a good is not so large anymore if a closely related person already got it.

This function does (at least in most of the cases described in this work) not only include the percentage of susceptibles willing to imitate the behavior of users (bA^{α}) - previously described as imitators -, but also the percentage of people who start using the product for individual reasons (given by parameter a) - the innovators.

The application of the control instrument influences the number of users (and therefore also the number of susceptibles) in a direct and an indirect way: The function $g(v)$ describes in both applications the influence of the control on the initiation, or more precisely it reflects the percentage change of the flow from susceptibles to users due to the application of the control. It is assumed that $g(0) = 1$, meaning doing nothing neither leads to an increase nor decrease of initiation. It is also valid for both marketing and harm reduction that $g_v > 0$ since it is assumed that any increase of marketing activities makes a product more attractive on the one hand, and on the other hand less severe consequences of drug use would also lead to a higher incentive to become a drug user.

But there is also an indirect effect leading to an increase of users due to the application of the control. As the flow from susceptibles to users increases, the number of consumers grows as described by using the control instrument. Since this means that the influence of existing consumers on susceptibles becomes larger too, initiation and the number of users will further increase.

Finally, parameter μ describes the quitting rate of the consumers, e.g. due to the reach of a certain age, death or a loss of the appeal of the good for other reasons.

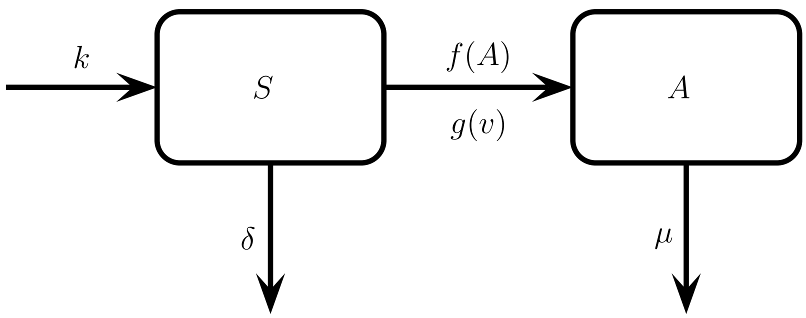

Figure 1.1: Flow diagram

This work will analyze a range of different problems where it makes sense to use this particular dynamical system to describe the development of potential consumers and adopters of the according good. The aim of taking this particular dynamics for the application in very different fields, i.e. in marketing, where the decision maker seeks to maximize his profits and in harm reduction where the decision maker wants to minimize the costs arising by drug usage, is to see how the inclusion of this state describing the development of susceptibles can affect the optimal strategy and in a way to show the potential of such an approach.

However, the parameters used in this work are mostly not based on real data. While this means that the actual outcomes will hardly fit for a certain concrete situation, but it still allows some general observations and economic interpretation about the reasons why a system behaves the way it does and why a decision maker should apply a certain strategy.

1.3 Two Motivating Examples

This section now will provide two motivating examples how a decision maker can optimally influence his outcome, taken from the two fields considered in this thesis, namely harm reduction and marketing.

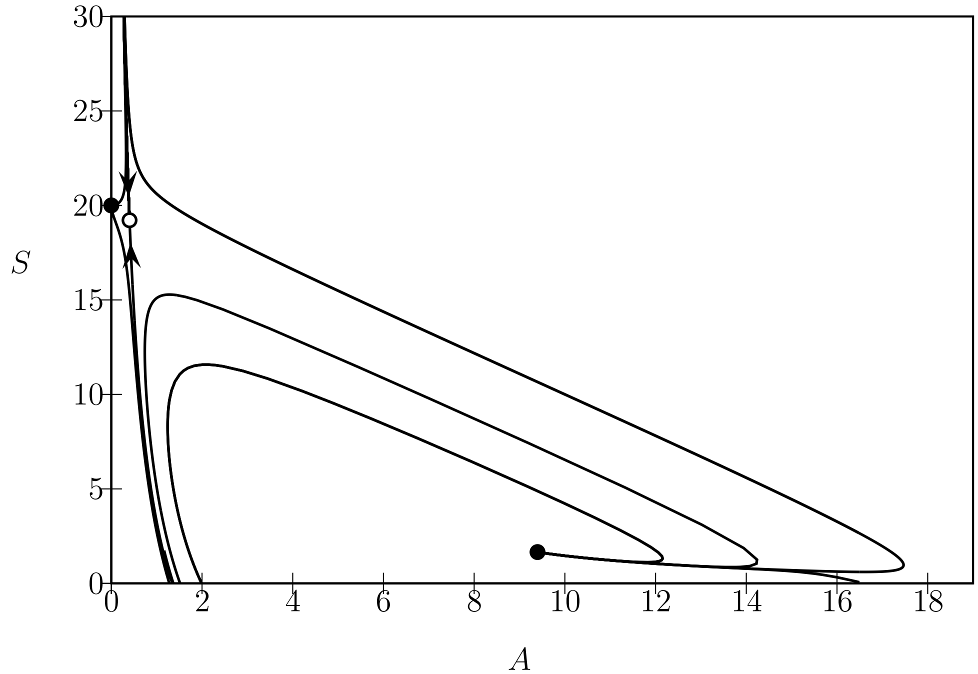

Figure 1.2: Phase portrait for the harm reduction case, showing the number of susceptibles S and users A for different optimal solutions. Depending on the starting point, there are two possibilities of where an optimal solution can end: in a steady state (illustrated by a $\bullet$) with many or in a steady state with no users, seperated by a threshold emerging from a non-optimal steady state $\circ$.

1.3.1 Harm Reduction in Illicit Drug Consumption

Some politicians / decision makers are confronted with the problem whether it makes sense to consider harm reduction measures such as needle exchange programs or legalization in order to avoid harm to drug users and, consequentially, society arising through certain drugs. While supporters of harm reduction emphasize the advantages of a lower harm caused by drug usage, such as, e.g., lower health risks, people opposing to such efforts say that such programs might lead to higher incentives for some people to start taking drugs or increase the amount of drugs taken.

In Chap. 4 we will see that the optimal application of harm reduction instruments depends on the existing drug situation, particularly how many users and susceptibles are there in the beginning (i.e. the initial value of the states), and on the attractivity of the drug to potential consumers, the influence of existing users, etc. (i.e. the parameters).

Let us now consider an example how and when a decision maker should or should not try to influence the development of the drug situation with harm reduction measures.

In this drug policy model the objective is to minimize the total arising harm caused by users as well as the arising costs caused by the control instrument, i.e. any harm reduction programs. There are two ways in which the users negatively influence the outcome: On the one hand there is the direct harm, caused by their usage of the drug and on the other hand there is some kind of indirect harm: In the model here it is assumed that the drug users exert some kind of influence on non-users (e.g. peer pressure, a promise of image change) to follow their example to take drugs. Of course, if there are more users, this influence becomes larger and larger. There are certainly also users, who try drugs independently of others, but for some drugs their number might be neglectible (see, e.g., Wallner, 2008).

Fig. 1.2 shows a phase portrait of the case where there are no innovators. Then there are two possible final outcomes, depending on the starting point. If the number of users and susceptibles is initially small, they do not have the means to cause much harm neither directly nor indirectly. Then the number of users cannot grow and will even start to decline due to the small inflow. However, if their number exceeds a certain threshold, which is more closely shown in Fig. 1.3, drug use will escalate and lead to a final outcome with many drug users. This threshold is a weak DNSS curve (see Appendix A).

It can also be seen (particularly in Fig. 1.3), that also the number of susceptibles plays a certain role whether a drug finds many users or not. If the initial number of potential consumers is large, then, as they reach more people, less users are necessary in order to convince the same amount of people compared with the case when there are only a few or no susceptibles. As such, if starting close to the weak DNSS curve the initial number of susceptibles can make a difference regarding whether the problem escalates or not.

The optimal control strategy takes regard of the actual number of users and susceptibles. In the case where the influence of users is not strong enough to make the drug situation escalate, it does not make sense to do any harm reduction. On the one hand, the harm of the users is not large enough to be of real concern for the decision maker; the costs of harm reduction would exceed the utility of it, respectively. On the other hand as already described any

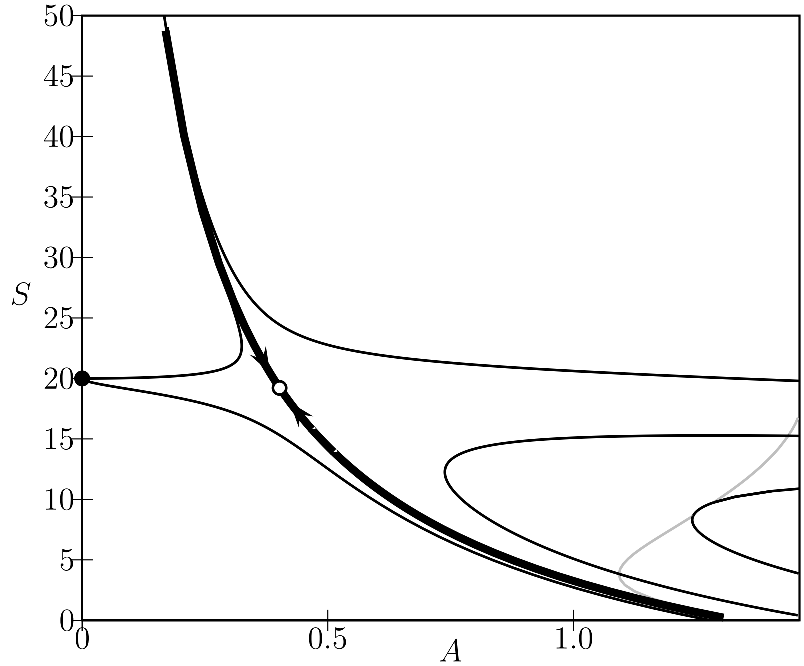

Figure 1.3: Zooming (in different scale) of the threshold of the harm reduction model of Fig. 1.2.

effort in harm reduction would not only decrease the harm, but also motivate some additional susceptibles to become users. Especially if the number of users and susceptibles is close to the weak DNSS curve it would be fatal to do anything.

However, if the initial states are such that their number exceeds this threshold depicted in Fig. 1.3, a decision maker also has to adapt his strategy according to how many users and susceptibles there are at each instant of time: If the number of users is small, a decision maker should do nothing for the same reasons as described before, but if the number of users becomes so big that the harm caused by them cannot be neglected anymore, the decision maker should start with harm reducing measures: Even though it leads to an additional increase of drug consumers, it has such an important impact on the harm directly caused by the users, that this additional inflow simply has to be tolerated. If the number of users becomes larger it is required to spend more and more for harm reduction until the maximum possible amount is reached. As the inflow to susceptibles is constant, i.e. when the population grows at a constant rate,the flow to users will be limited in growth after some time and may even decrease if there are not enough susceptibles to support initiation anymore, until their steady state value is reached. As such the steady state is not always directly approached.

It is not optimally possible in this case to use the control instrument in a way to give a decsion maker the choice between having finally to deal with many or no users at all. Note, however, that a decision maker could, if he knows that such a weak DNSS curve exists, try to evaluate whether there are any possibilities (e.g. by using more traditional drug policies, such as prevention, treatment or law enforcement) that are not captured by the model, by which he could manipulate the initial number of users and/or susceptibles so that he can reach the more convenient outcome.

This is not the only possible outcome for the harm reduction model. It will be shown in Chapter 4, that for different parameter values the number of steady states and their stability properties can change, i.e. there might be only one steady state, where optimal solutions can end in a limit cycle can occur, or there might be a case with one steady state and a limit cycle separated by a weak DNSS (threshold) curve.

1.3.2 An Example from Marketing - Taking Price Reduction Measures

When trying to sell a product it often proves to be of concern how a price should be adapted in order to make a product more interesting for potential consumers. A company can give price reductions in order to promote the sales, but since a smaller price means less profit, the question arises when should a decision maker reduce the price how much in order to maximize his profits. In Chapter 2 we shall see that the optimal solution depends on the parameters and the initial values of the states. Similar to the harm reduction case, it can occur that for some starting points nothing can be done; a steady state with no users will always be reached. If, however, the number of existing users is large enough, then the influence of these consumers is so big, that (also with the provision of additional incentives to purchase the product through price reductions) the number of users can grow until there are not enough susceptibles to allow a further increase. In the price reduction case a decision maker might even have the choice between letting the product fail or, with high efforts, making it successful. However, unlike in the harm reduction model, control application would be large if the initial number of users is small in order to built up a consumer base, who then can influence further people in their purchase decision.

Another example of a possible outcome is that of a limit cycle (see Fig. 1.4). Then the number of susceptibles and users is not fixed in a steady state

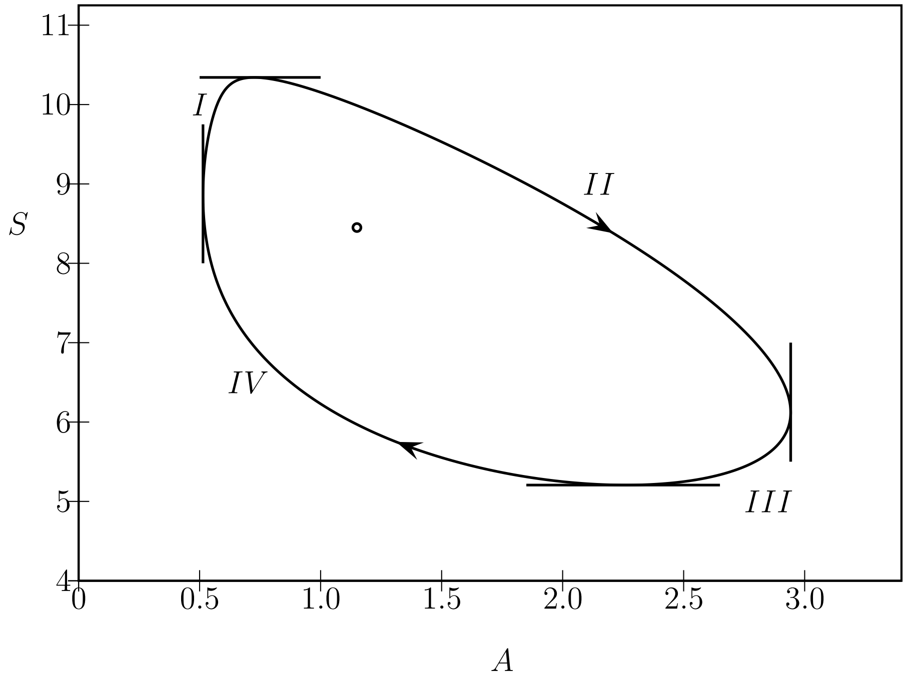

Figure 1.4: Limit cycle in the marketing/price reduction case. ○ again depicts a steady state which is not a candidate for an optimal solution to end in.

at the end of the solution path, but will periodically increase and decrease. Moving now to the field of marketing, a reason for such a behavior of the system can be found , e.g., in a low consumer loyalty or a high quitting rate of the users, respectively. Unlike before the fraction of innovators among people who start to use a certain product is not zero anymore, however, it is possible to find limit cycles without people whose adoption decision is not based on others as we shall see later. The optimal strategy then has to be adapted according to which phase (compare Fig. 1.5) of the cycle one currently is: If there are only a few users, but many susceptibles (phase I), the number of consumers can grow, simply because the peer audience of this product is so big that it is quite easy to find people willing to purchase such a product. It makes sense then to support initiation and increase the attractivity of the product to potential consumers by certain marketing activities, such as giving price discounts. As the flow from susceptibles to users increases, so does their influence of the users and it becomes less and less necessary to do anything to promote the product (II). After a while the number of potential consumers decreases (phase II) as their growth is limited by a constant inflow, and an increasing outflow, due to the the growing strength of the influence of the consumers, i.e. the attractivity of buying a certain product increases. But

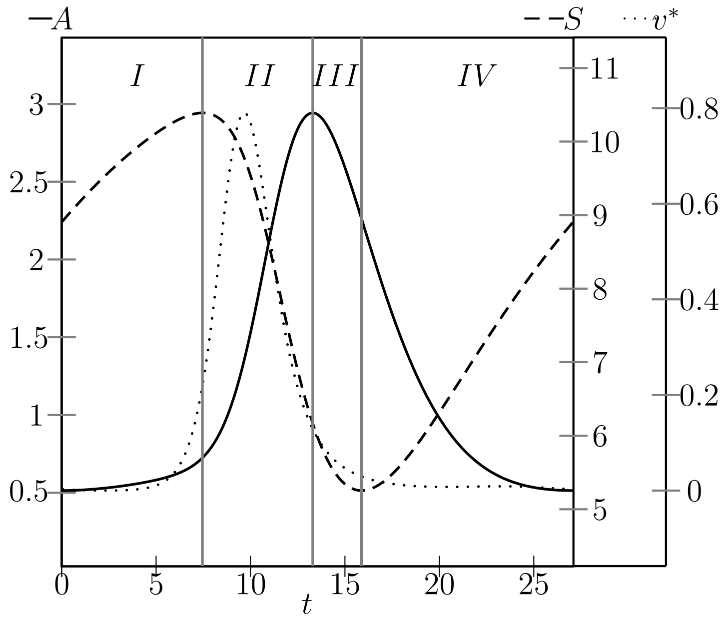

Figure 1.5: Timepath of the limit cycle depicted in Fig. 1.4

this decrease and later lack of people who could potentially become attracted to use the product, leads to a slow down and later on a stop of the growth of users (III), because on the one hand there are too few people who can be recruited, on the other hand the decision maker has to pay tribute to the high quitting rate, due to which the outflow from users starts to exceed the initiation flow due to the constant inflow limited initiation. As the number of users and their influence with it decreases, the number of potential users can "recover" (IV) until the market for this good becomes again large enough, so that it is easy to find new people willing to purchase the product. Note that one can, similarly to Feichtinger (1992), name the four different parts of the limit cycle *prosperity*, *saturation*, *declining* and *recovery phase*.

These are not the only possible outcomes regarding system behavior and optimal control strategy. Depending on the problem and on the used parameters one can find very valuable insights about how a decision maker should apply his available control instruments for an optimal outcome, may it be the maximization of profits in marketing or the minimization of costs in the harm reduction case.

Part I

Marketing

Diffusion processes play an important role in marketing, since any decision maker certainly wants to have deeper insights into the product adoption process in order to be able to adequately influence it. There is a large number of diffusion models in marketing literature, the most important probably being Bass (1969); some other models are Gould (1970); Muller (1983) and Hartl & Kort (2005). For a short overview of marketing diffusion models using methods of optimal control theory see Feichtinger & Hartl (1986); see also Mahajan et al. (1990).

The interaction between users and non-users is often crucial for the success of a certain good for different reasons. The utility of network goods, for instance, depends on the number of people using it, and it might be very difficult to influence anyone buying such a product if no one else has it (e.g. a telephone). Cultural acceptance can also play a role when it comes to the spreading of a certain good (see, e.g., Rogers, 2003), particularly if people try to avoid the risks that come along with trying out a new product. The positive example of other people using such a product is then absolutely necessary then for the success of a good.

When people use a certain product, they not only gain experience, which they might share with others, but usage might also have some symbolic character and express one's status, personality, taste,etc. (cp. Hoyer & MacInnis, 2004). Often even attitudes and emotions are associated to the usage of a product such as joy, love, hope, excitement, etc. Non-users might be attracted by what usage and as such being a user of such a good represents. Another example where the existing users of a product have a large impact on potential adopters is fashion. If many people own a certain product, it might create a certain urge among non-users to get such a product too. But there are not only people who start to use a product because of the example other people serve as (imitators), but also people who start the product independently of others (innovators). It will be seen that the size of the fraction of innovators and imitators can contribute to the success of a product.

There are many marketing instruments by which a decision maker can try to influence the number of sales of his product (see, e.g., Kotler & Armstrong, 2008), such as advertisement, various pricing strategies and product quality. In the following model we take a closer look at how promotional price reductions can influence the development of the number of adopters of a certain good. The purpose of promotional price changes is to create some excitement and urgency toward the potential customers. The price of a product has the property that it is the only element that leads to revenues for a company, therefore any reduction of the price reduces the profit. Even though, according to Lilien & Kotler (1983); Ofir & Winer (2002), (potential) customers might not always react to changes in the price in a straightforward way[2], it is assumed that a smaller price leads to an increase of purchases. For this, however, the consumers' knowledge about the price is also relevant (cp. Ofir & Winer, 2002), i.e. they must have some judgment about the amount of money a product costs and how much it is actually worth. A potential customer might also react differently to a temporary price reduction as to permanent change of the basic price, particularly if he is aware of the nature of the change and it is a part of the purchase decision whether the product is still affordable when no temporary price reduction is given. Price reductions do not only have a direct, short-term effect, i.e. an increased number of newly entering customers, but also an indirect, long-term effect, since existing users influence other people mostly positively in their product adoption decision.

While in marketing it often might not be too difficult finding out who the actual customers are (e.g. by customer loyalty programs), determining who the potential consumers are for a certain product, can be quite hard sometimes, particularly if the number of actual customers is quite low and reference data are rare(cmp. Kotler, 1999). Yet, it can be very crucial for the success of a certain product, because addressing the wrong audience means missing the really relevant group of people and their awareness of the product despite a possibly high effort, and in the worst case influence the real potential customers negatively in their purchase decision. (Although certainly the saying "There is no such thing as bad press" is often valid, e.g., reducing the price for a good basically attracting to a clientel with very exclusive needs might not be the best thing to do). The efficiency of different marketing measures also depends crucially on the knowledge, who the people deciding about a purchase are and by whom they are influenced (e.g., employees might not always have direct input about which equipment they have to work with). Therefore it makes sense to identify and include the dynamics of a group, who a company can address in order to find new adopters of a product.

According to Hartl & Kort (2005) one can distinguish between two different kinds of marketing models, one focusing on how to gain new customers, the other one trying to explain how to keep them. The models in this work focus on the optimal way of finding new customers, particularly examining the role that the potential customers of the product play.

There are plenty of models considering the potential customers, many of them, however, assume that the total population size to is fixed at a certain level, see, e.g., Bass (1969) and expressed the number of susceptibles

[2]A price reduction might not always contribute in a positive way to potential consumers' adoption decision, particularly since the price is according to Ofir & Winer (2002) also some kind of communication device, signaling potential customers for example high quality or trendiness.

as being the proportion of the population who is not already a customer. A different approach was followed by Mahajan & Peterson (1978), who did not assume the number of the potential adopters to be fixed, but as a function of relevant factors. However, they did not include a seperate state equation for the susceptibles. Muller (1983), who assumed the total population size to be constant, considered two different kinds of people who might be potential buyers of the good. He distinguished between people among the susceptibles who are and people who are not aware of the product. He included state dynamics for both of these groups. However, under the assumption that total population size was fixed, he expressed the number of potential customers aware of the product as a function of the other two states. Feichtinger (1992) introduced a model consisting of potential and actual customers, which was dealt with in more detail in Feichtinger et al. (1995). While it was shown that limit cycles might occur, the model was not optimally controlled and some assumptions made in the dynamics were different, such as that there was no outflow from the potential customers out of the system, but a backflow from users to susceptibles.

The model used in this work, however, includes one state variable describing the development of susceptibles and no explicit assumptions are made about the size of the total population. No distinction is made between people who are aware of a certain product and those who are not. Some reasons for the spreading of a good, i.e. being influenced by the example of existing users and the application of the control instrument, however, certainly contribute both to an increased awareness of the product among the susceptibles. The model analysed in the next chapter, considers a trade-off between making a product more attractive by giving price reductions and lower profits, and will take a particular focus on the adoption process.

Chapter 2

One-Stage Marketing Models

The next chapter describes and analyzes a one-stage marketing model, focusing on the impact that a price reduction can have on the number of users and potential consumers of a product as well as their interaction. The results will be calculated analytically and numerically. They will give some idea about why a the number of (potential) customers of a product evolves the way it does and how one should optimally apply the control instrument, i.e. price reductions. The model used here will be extended into a multistage version in Chapt. 3.

2.1 A Price Reduction Model

In the following model a decision maker, who is the only provider of the good, wants to maximize his profits. Knowing that more people are likely to be attracted towards the product if on the one hand the price is low and if on the other hand there are many people, who serve as some kind of example, a decision maker can reduce the price to increase the number of people newly adopting the product. This, however, has some negative impact on the profits since it is the price which creates the revenues for a company.

It is assumed that each user of the product contributes each time unit to the profits of the company, either by a frequent purchase of the non-durable good, by having to pay for the usage of the product (e.g., telephone or internet fees) or by buying a complementary goods also provided by the company (e.g., add-ons of a certain software). The objective function is then given by

$$\max_{0 \le v \le v_{\max}} \int_0^\infty e^{-rt} (\pi A(1-v)) \mathrm{d}t, \tag{2.1}$$

Parameter π denotes the average profit per consumer per time that a company makes when the price is not reduced in a promotional effort. It consists of the difference between total sales and total costs per user. As such, π displays the profits if the product is sold at the standard price. Assuming that consumers are aware that a price reduction is only a promotional effort to gain more users, π, the profit per user, serves as some kind of reference regarding the attractiveness of a temporary price reduction. Therefore, initiation depends on how big the percentage reduction of the average profit per user caused by a price reductions is.

The discount rate is given by r, A denotes the number of users, and the control instrument v is the percentage reduction of the profit caused by price reductions[1]. For obvious reasons, it is assumed that smaller price reductions do not have a big impact on the profits, while large price reduction take away much of the profit. It makes sense to introduce constraints for this control, i.e.

$$v \geq 0 \qquad \text{and} \qquad v \leq v_{\max} \leq 1, \tag{2.2}$$

signifying that a decision maker can choose somewhere between giving no discount, leading no cuts of the profit and giving a price reduction at a maximum possible level $v_{\max}$, which is exogenously determined, and reflects e.g. possible corporal specifications of not going below a certain price level. It is not allowed in this model to charge a price that would cause the average profit per user to be negative, i.e. lead to losses. As such $v_{\max}$ has to be assumed to be smaller or equal to one.

The state equations are given by

$$\dot{S} = k - \delta S - f(A)Sg(v) \tag{2.3}$$

$$\dot{A} = f(A)Sg(v) - \mu A. \tag{2.4}$$

$$\text{s.t.} \quad S(0) = S_0 \text{ and } A(0) = A_0 \tag{2.5}$$

As described in Chap. 1 S denotes the number of potential consumers / susceptibles, k the inflow rate, which is assumed to be constant, δ the quitting rate of the susceptibles, $f(A)$ the initiation function, $g(v)$ the influence of a price reduction on initiation and μ the quitting rate of the users, reflecting also the consumer loyalty.

Condition (2.5) states that an optimal solution has to start at a certain exogenously given starting point (A_0, S_0). It is not necessary to assume even when the product is newly launched at a certain market, that the initial

[1] Basically, it would also be possible to interpret it as some other marketing measure, that, by its application, reduces the profits, but leads to an increased attractiveness of the product, so that more people are willing to buy it.

number of consumers is equal to zero. This can occur when the initial demand for this product is so big that immediately a consumer base arises. Reasons for such a large initial demand can be that either the product is really needed or that previous marketing measures that are not captured by this model lead to some customer excitement and urgency to immediately buy the product (e.g, the newest movie, computer game, mp3-player, etc.) (cmp. Kotler, 1999).

The number of people starting to use a product can be separated into two different groups: the innovators and the imitators. The initiation function $f(A)$ therefore consists of two parts, with parameter a denoting the percentage of susceptibles, who decide for their very own reasons to start using a product and bA^α describes the percentage of susceptibles who is influenced by the number of existing users (A) in their purchase decision. Parameters b and α serve to weight the strength of the influence of the consumers. The initiation function therefore has the following shape:

$$f(A) = a + bA^\alpha. \tag{2.6}$$

It is assumed that $f_A \geq 0$, meaning that each additional user makes the product more attractive to others by providing a higher degree of information and an increasing awareness of the potential customers regarding this product. In this work α is greater than one, reflecting that the pressure, incentive or urge to purchase a product becomes stronger the greater the number of existing users is. This can be potential consumers might want (when it is particularly trendy to use have a good, e.g. toys, technical gadgets, certain fashions) or might even have (when interaction and compatibility play a big role, e.g. certain soft- or hardware products, or when it is needed by the customer to remain capable for competing) to fit in. The specific parameters for a and b have to be greater than zero, so that the initiation function fulfills the condition $f(A) \geq 0$ for all A, meaning that a flow from A to S is not allowed within the initiation function, which only serves to describe the product adoption process.

Then the question arises how does the application of the control influence initiation? According to Kotler & Armstrong (2008) consumers can react very different to any change of price depending on the nature of the product and the image of the corresponding company. Here it is assumed that higher price reduction makes the product more attractive, i.e. $g_v \geq 0$ and that for no control application $g(0) = 1$. Because of the Legendre-Clebsch condition it is to be assumed that function g is concave ($g_{vv} \leq 0$, see Appendix C).

A function fulfilling these conditions is

$$g(v) = 1 + \beta v^\omega \qquad \text{with} \qquad 0 \leq \omega \leq 1 \tag{2.7}$$

Regarding economic interpretation, this assumption means that increasing the application of the control is most efficient in terms of the percentage increase of the flow from the first to the second state when the control is small. This means that if a company gives a large price reduction, and faces therefore lower profits per user, it is not much important for the customers if the company increases or decreases the price a little bit – the reduced price is still perceived as good. If a company would give only small price reductions, and then even further decrease them, far less people would be attracted to the good, simply because it would not really pay off for most of those who can be reached by the control measures to take advantage of the temporal price reduction.

Note that using this specified function $g(v)$, $f(A)S$ describes the number of people who start to use the product if the product is sold at the standard price.

Application of Pontryagin's Maximum Principle

The Hamilton function then can be formulated as

$$\mathcal{H} = \pi A(1-v) + \lambda_1(k - \delta S - f(A)Sg(v)) + \lambda_2(f(A)Sg(v) - \mu A) \tag{2.8}$$

where λ_1 and λ_2 represent the costates corresponding to the first (S) and the second state variable (A), respectively. They are required, because control application does not only affect the profits, i.e. the objective function directly, but also by influencing the development of the users and susceptibles, i.e. the state equations. The Hamiltonian function therefore has to include all these functions and the costate variables serve in some sense to weight the impact of the state equations. The economic interpretation of the costates is that they represent a shadow price. It measures the impact of an additional susceptible (λ_1) or user (λ_2), respectively on the objective function. As such it represents the maximum amount a decision maker is hypothetically willing to pay for an additional unit of the corresponding state (see, e.g., Grass et al., 2008).

Applying now Pontryagin's Maximum Principle to the Lagrange function[2] which is given by $\mathcal{L} = \mathcal{H} + \nu_1 v + \nu_2(v_{\max} - v)$ we find that

$$\begin{aligned} \mathcal{L}_v &= -\pi A + (\lambda_2 - \lambda_1)f(A)Sg_v + \nu_1 - \nu_2 = 0 \qquad (2.9)\\ \mathcal{L}_{vv} &= (\lambda_2 - \lambda_1)f(A)Sg_{vv} \leq 0 \qquad (2.10) \end{aligned}$$

[2]The first Lagrange multiplier ν_1 corresponds to the control constraint $v \geq 0$ and the second one ν_2 to the constraint that $v \leq v_{\max}$.

Appendix C.1 shows why (2.10) is always fullfilled for $g_{vv} \leq 0$ if no control constraints are active. Note that in this work only the necessary and not the sufficient conditions are fulfilled and therefore the found solutions are basically only extremals, but are assumed because of the economic interpretation to be optimal. The complementary slackness conditions

$$\nu_1 v = 0 \qquad \text{and} \qquad \nu_2(v_{\max} - v) = 0 \tag{2.11}$$

with $\nu_1, \nu_2 \geq 0$ have to be fulfilled. Note that both constraints cannot be violated at the same time, meaning that at least one of the Lagrange multipliers always has to be zero.

From (2.7) we find that $g_v = \omega\beta v^{\omega-1}$; by inserting this into (2.9), the optimal control can be expressed as

$$v^* = (\frac{(\lambda_2 - \lambda_1)\omega\beta f(A)S}{\pi A - \nu_1 + \nu_2})^{\frac{1}{1-\omega}}. \tag{2.12}$$

It can be seen that if an additional user would be particularly valuable (reflected by a large costate λ_2), it would be optimal to give larger price reductions, while if more susceptibles are desired (reflected by a larger costate λ_1) a decision maker should do less in order to avoid the additional flow caused by control application. If initiation, i.e. $f(A)S$, is high or control application has higher impact (represented by the parameters β and ω), more efforts should be put into control measures, simply because they are more efficient then. If the profits made by all customers are large, a decision maker should consider to do less in order not to cut his profits. Apparently, the relation between initiation and profits plays a large role when determining the optimal value of the control, which is not surprising considering that while price reductions are good for initiation, they are bad for the profits.

The costate equations are derived as required by Pontryagin's Maximum principle as

$$\dot{\lambda}_1 = (r+\delta)\lambda_1 + (\lambda_1 - \lambda_2)f(A)g(v) \tag{2.13}$$
$$\dot{\lambda}_2 = (r+\mu)\lambda_2 - \pi(1-v) + (\lambda_1 - \lambda_2)f_A S g(v) \tag{2.14}$$

Remark on the Numerical Calculations *The first step to do numerical calculations of an optimal control problem is to state the first order necessary optimality conditions obtained by the application of Pontryagin's Maximum Priciple. This includes the statement of the canonical system, the optimal control value by the Hamiltonian maximizing condition, etc. With given information about the state equations, objective function and control constraints, the OCMat-Toolbox is able to handle this by first deriving these* *conditions analytically using Matlab's Symbolic Math Toolbox and then generating the necessary files needed for further computations automatically. See Appendix B for more details.*

Equilibria

Since the system is too complex to find all of the steady states analytically, it is necessary to compute them numerically as will be done in the subsequent sections. However, some properties of some of the steady states can be found analytically.

By setting the state equations (2.3) and (2.4) to zero, we find that in a steady state

$$\hat{A} = \frac{k - \delta\hat{S}}{\mu}. \tag{2.15}$$

This means the number of users will be large in a steady state if the inflow to the system is large and the quitting rate of susceptibles is low, both indicating that it is advantageous to have many potential first time consumers available. If the quitting rate of users is high, sometimes reflecting a not particularly attractive product, the number of consumers will be lower compared to a case where this rate was small.

From (2.13) we find that

$$\hat{\lambda}_1 = \hat{\lambda}_2 \frac{f(\hat{A})g(\hat{v})}{r + \delta + f(\hat{A})g(\hat{v})} \tag{2.16}$$

confirming that in the steady state the second costate, i.e. λ_2, has to be larger than λ_1, since due to the assumed positivity of the parameters the numerator in (2.16) is smaller than the denominator. Since the interpretation of the costates is that of a shadow price, describing the highest price a rational decision maker would pay for an additional unit of the state variable, this basically says, that an (additional) user is more valuable than an (additional) susceptible. While it certainly would be good to have additional potential consumers among which one can recruit new users, it is better to have more users since they directly lead to revenues for the company by paying for the good, but also exert some influence on susceptibles by causing imitational behavior of susceptibles.

The optimal control applied in the equilibrium depends on the number of users and susceptibles in the steady state: If there are many users it is better to do nothing and exploit the high profit as well as the fact that existing users widely influence potential ones. If the number of susceptibles is big,

Table 2.1: Steady state values for $a = 0$

$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
$\frac{k}{\delta}$	0	0	$\frac{\pi}{r+\mu}$	0	0	0

one would spent more on marketing efforts compared to a case with a small number of susceptibles, since more can be gained then except in the following case:

Taking now a closer look at the initiation function $f(A) = a + bA^\alpha$, if parameter a is zero, then at least the following equilibrium can be found, displayed in Tab. 2.1. The eigenvalues of this state state $\xi_1 = -\mu$, $\xi_2 = -\delta$, $\xi_3 = r + \mu$, $\xi_4 = r + \delta$ confirm that this steady state is indeed a saddle point and the Legendre-Clebsch condition is trivially satisfied with $L_{vv} = 0$ at this steady state.

In later sections we will see that this steady state is not the only one for $a = 0$, i.e. there will be another one with a large number of users. This analytically found steady state represents the case where the number of existing users is so small that hardly anyone ever owns this product - leading to no imitation - and no one is really interested for reasons independent of other people of doing so. Something like this can happen if (a) the product is useless if no one else owns it (e.g. network goods, such as telephones, instant messaging software, etc.), or people are particularly afraid (b) that due to the lack of users the company will soon cancel all support services, (c) that the product is faulty or not trustworthy or (d) that it simply has not found acceptance in the dominating culture.

As there is no flow from susceptibles to users, the number of susceptibles will remain at a relatively high level. The only other positive steady state value is that of the second costate. While nothing could be gained by an additional susceptible, an additional user would very well be welcomed, however, due to the lack of existence of users who serve as examples to potential users, this is not possible and any control effort would be useless leading to an optimal control equal to zero.

Remark on the Numerical Calculations *Matlab provides functions for equation solving, which are used by OCMat to numerically calculate the steady states. OCMat also automatically computes the Jacobian matrix, where its eigenvalues determine the stability properties of the steady state. All the information of the steady state, needed for subsequent computations, are returned in an Matlab object.*

Table 2.2: Parameter values for the single steady state case

r	π	k	δ	μ	a	b	β	ω	α
0.04	1	1	0.05	0.12	0.02	0.025	1	0.75	1.75

Table 2.3: Admissible steady state values for the parameters described in Tab. 2.2

$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
0.99	7.92	6.42	7.02	$1^{-5} * 0.88$	0	0

2.2 One Long Run Steady State

For many parameter constellations, such as seen, e.g., in Tab. 2.2, only one admissible steady state, serving as candidate where an optimal solution can end, can be found[3]. The equilbrium values of the state, costate and control variable can be found in Tab. 2.3, revealing that the number of users is unlike the number of potential adopters rather large in this steady state, and that due to this not much control efforts are required, because (a) the influence of the users of the product on the adoption decision of potential customers is rather strong and working in favor of the decision maker and (b) the number of people who could be recruited to start using the product is so small that any control effort would not very effective.

Fig. 2.1 shows a phase portrait revealing the shapes of five different optimal trajectories, which differ in their starting point. Considering the trajectories in clockwise direction, in the first one it is assumed that there are initially no customers and no susceptibles, in the second case the number of initial susceptibles is high, however the number of users is still low, in the third and fourth case the number of users is assumed to be high, the number of susceptibles to be high and low, respectively. The gray line shows the points of the stable manifold where a control constraint becomes inactive: If the number of users is smaller than the corresponding point with the same

[3]Other non-admissible steady states can be found, where the Lagrange multiplier is negative or the control outside the admissible control region. However, since they are for the most part not relevant for economic interpretation, they are not described more closely in this work.

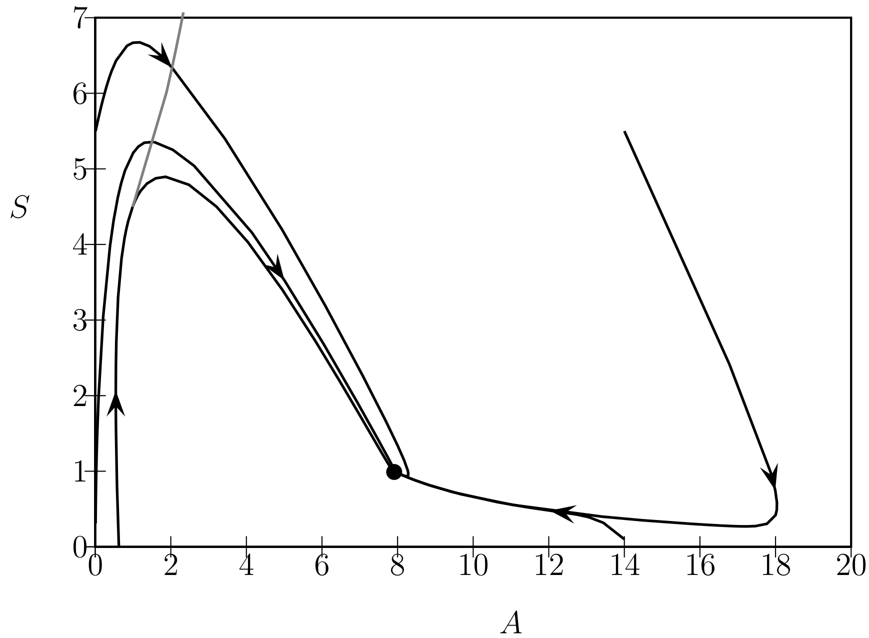

Figure 2.1: Phase portrait of the single steady state case of the marketing model; the black lines show optimal solutions starting at different initial points (A_0, S_0) ending in the equilbrium. On the left side of the gray curve it is optimal to apply the maximum possible control. On the right side the constraint is not active anymore.

number of susceptibles one would have to give price reductions as large as possible. (A higher number of users would mean one could determine the control optimally within the admissible control region.) The fifth trajectory shows the last optimal solution, where no constraint is active; there are initially no susceptibles, but some consumers of the product. Starting only with a little bit less users, one would would first give the maximum amount of control efforts, and then reduce them and exploit the high number of users. At this trajectory the maximum admissible price reduction is only required at one single point, i.e. where it touches the curve, where one switches from active to inactive control constraint.

In the first and second case one would give initially price reductions as large as possible in order to ensure the growth of the paying customers. While in both case the flow from susceptibles to users would in the beginning still be so small that despite the increased efforts to make those people buy such a product the number of potential users can still grow. Particularly in the second case one would soon reach such a high number of users that the number of susceptibles would decrease after a while due to the user's stronger influence and the incentives provided by the company in terms of price reductions. This means the outflow from susceptibles both out of the system and to the other state soon would exceed the constant inflow of people to this state. The number of customers then would grow, as would their impact on the potential users; since their inflow is restricted by a constant inflow, their number would soon be so low that no further increase of customers is possible and a steady state is reached. As the number of users grows it is not required anymore to give large price reductions; the influence of the users works very much in the favor of the affected company. Note that the steady state is not always approached directly: If really many people start adopting a product, the number of users overshoots its steady state value, but after some times not enough susceptibles are left to permit further growth. The constant inflow to this state is too slow to even avoid that the number of actual users decreases for some time.

If the number of initial susceptibles is high (unlike the number of users) then it makes sense to give price reductions, simply because, when reaching a large audience, they are most effective. However, if the number of users is really big too, such measures are not required, at least not in a similar magnitude. While one would certainly exploit the large market potential, the large number of existing customers has enough influence that no further measures are necessary, and the number of paying customers will still grow, at least for a while. However, starting at such a point, the number of susceptibles and also then the number of users would fall after a while due to the constant inflow to S. The same is valid if the initial number of susceptibles is low, but the number of users high.

Remark on the Numerical Calculations *The optimal solution paths (trajectories) solving the system of ODEs (ordinary differential equations) can be gained by inserting (2.12) into (2.3)-(2.4) and (2.13)-(2.14). These can be found by either solving boundary value problem (BVP) or an initial value problem (IVP). Provided that the optimal solution is controlled into a steady state* $(\hat{A},\hat{S})$, *the necessary boundary conditions for given initial states* (A_0,S_0) *become* $A(0) = A_0$, $S(0) = S_0$ *and ends up in a steady state of the canonical system with* $A(\infty) = \hat{A}$, $S(\infty) = \hat{S}$. *For the numeric calculation the latter condition has to be replaced by so called asymptotic boundary conditions (i.e. the trajectory has to end at the linearization of the stable manifold of the steady state; see Grass et al., 2008). OCMat provides functions for the automatic formulation and solution of such problems(*`initocmat` *and* `occont` *- see Appendix), using by default Matlab's* `bvp4c`*-solver.*

When one of the control constraints gets violated, OCMat adds further

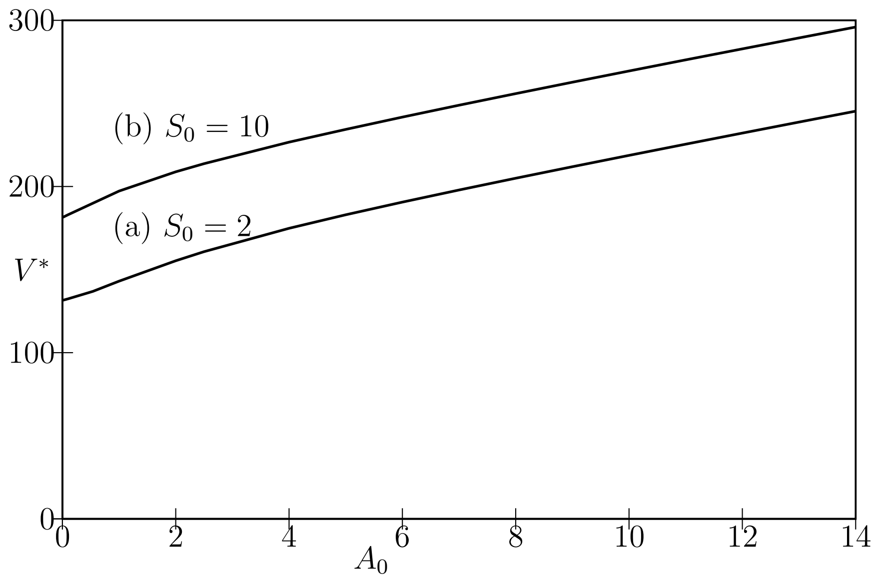

Figure 2.2: Value of the objective function for different initial values of users and susceptibles (a) $S_0 = 2$ and (b) $S_0 = 10$. .

conditions to the boundary value problem in order to be able to proceed with the calculations. These conditions are as follows: the optimal solution has to consist of two parts, one where the constraint is active and one where it is not, and that these two parts have to end/start at the same point (states and costates have to be continuous), i.e. the point where the control constraint becomes (in)active. And a further condition determining the time point of the switching between the two regimes, given by the continuity of the Hamiltonian.

A (potential) optimal solution can also be found by solving an initial value problem, knowing that the trajectory has to end in the specified steady state, or equivalently formulated lying in the stable manifold of the steady state. Then starting from this equilibrium (more exact near the steady state on the linearized stable manifold) one can calculate backwards in time using e.g. OCMat's `odesolve`*-function, which uses by default Matlab's* `ode45`*-solver.*

Fig. 2.2 reveals the value which the objective function takes for different initial values of the users while the number of susceptbiles is assumed to be fixed at (a) $S_0 = 2$ and (b) $S_0 = 10$. It can be seen that the higher the number of initial customers is, the higher will be the value of the objective function: When there are many consumers of a good, they do not only contribute directly to the profits of the company by paying for the good, their high number also leads to a higher initiation and therefore, an easier customer acquisition since then profit diminishing price reductions are not necessary.

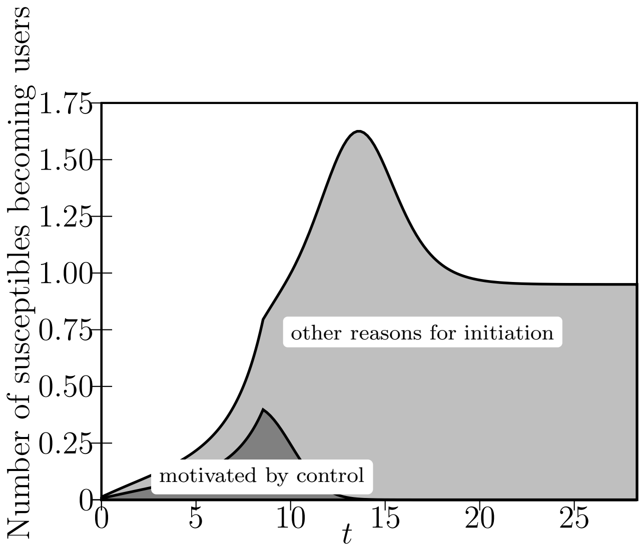

Figure 2.3: Number of susceptibles starting to use a certain product on different point of times on the trajectory starting with hardly any users and susceptibles motivated or not by incentives provided by a lower promotional price

If the initial number of susceptibles is high, Fig. 2.2 shows that this is better for a decision maker compared to a scenario where their initial size was low. The reason of course is that if the market (potential) is larger it is possible to convince more people to become paying customers of a product over time.

Remark on the Numerical Calculations *For showing how the value of the objective function depends on the initial values of the states, one has to find the optimal solution starting at the specified points. Then one can calculate the value of the objective function by using the value of the Hamlitonian and dividing it by the discount rate (see Appendix A).*

Taking a Closer Look at the Initiation

When trying to optimally influence one's profits, it is certainly of interest to see which people are attracted to the product at which instant of time. The qestion arises how many people are affected by the application of the control instrument or whether product appeals more to innovators or to imitators.

Looking now on the trajectory depicted in Fig. 2.1 where the initial of susceptibles and users both are very, very small, Fig. 2.3 reveals the impact of giving price reduction on the initiation: When there are very little customers they cannot contribute much to the potential customers purchase decision

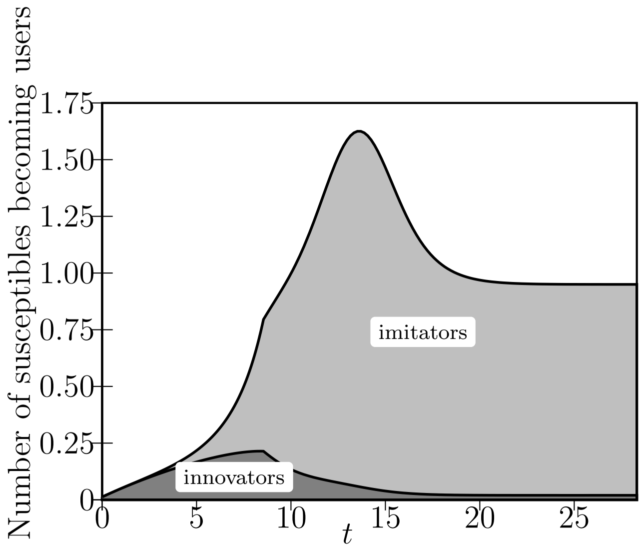

Figure 2.4: Number of susceptibles starting to use a certain product on different point of times on the trajectory starting with hardly any users and susceptibles differentiating between the fraction of innovators and imitators

and many people decide to adopt the product because of the incentives provided by the company. As the number of potential and actual users grows, the control application becomes more and more efficient, leading to a higher initiation. With an increasing customer base, the impact of users serving as example becomes larger and larger, and at a certain point (reflected by the peak of the lower curve), it is possible to reduce the control efforts and exploit the higher profits and the larger size and impact of existing users. It can be seen that the flow from the first to the second state grows, but that after a while, when the pool of susceptibles is exhausted, the flow starts to decline again. At later parts of the optimal solution it is not the price reductions anymore, that are crucial for the spreading of a good, but mainly the propagation of the product's advantages by people who already use it.

Fig. 2.4 tells us more about the role of innovators and imitators on this particular optimal solution. It can be seen that when the number of users and susceptibles is small in the beginning it is only the innovators who start adopting a product, since there is always a certain percentage of potential customers willing to try out the product. The number of innovators grows as the number of suscepitbles does and is positively influenced by the application of marketing measures. It reaches its peak, therefore, at the point where one would stop giving the maximum amount of price reduction, since at this point one still would have the benefits of a lower price but also a high number of potential customers willing to purchase the product independent of others. However as less control efforts are optimal - due to the high number of users and their influence and a decreasing number of susceptibles - the number of innovators falls. In order that there are people who start to use a product because of existing users, serving as example to them, there has to be someone owning and using the product. If the number of users grow, so does their influence on others and if the market potential increases with the susceptible group, there are more people prone to be influenced by others. Price reductions do also have positive effects on the imitators, however, the decrease of the size of this group after some time, relates more to the lowering number of susceptibles than to the smaller price discounts.

Remark on the Numerical Calculations *The last two figures showed the number of people who decide to start using a product at time* t*, which formally would be given by* $f(A)Sg(v)$*, or, using the previously specified functions, is* $(a + bA^\alpha)S(1 + \beta v^\omega)$*. Then the number of innovators, who decide to start a product independently of others is given by* $aS(1 + \beta v^\omega)$ *and the number of imitators by* $bA^\alpha S(1 + \beta v^\omega)$*. Similarly the number of people who start adopting a product because of the price reduction is given by* $(a + bA^\alpha)S\beta v^\omega$*. Then one can insert the calculated trajectory into these expressions to calculate the corresponding figures.*

The Influence of Parameters on the Steady State

As there is only one steady state, it is particularly interesting to see how it depends on the given parameters, because it gives some insight about why the optimal solutions looks the way it does. The following gives some idea about the impact of some of the parameters:

- a: A large proportion of innovators means that steady state number of users becomes larger, because of the larger outflow from the susceptibles in the initiation function. If the proportion of innovators is zero or very small, a second steady state can emerge with no or hardly any users, being candidate for an optimal solution. Then the impact of the existing customers might be so small that a large spreading of the good is not possible.

- b: Fig. 2.5 shows the impact of parameter b. When the influence of the users is very small, there will not be many other users in the steady state (particularly if the number of innovators is small too), and it is very well possible that the long-term number of potential customer exceeds

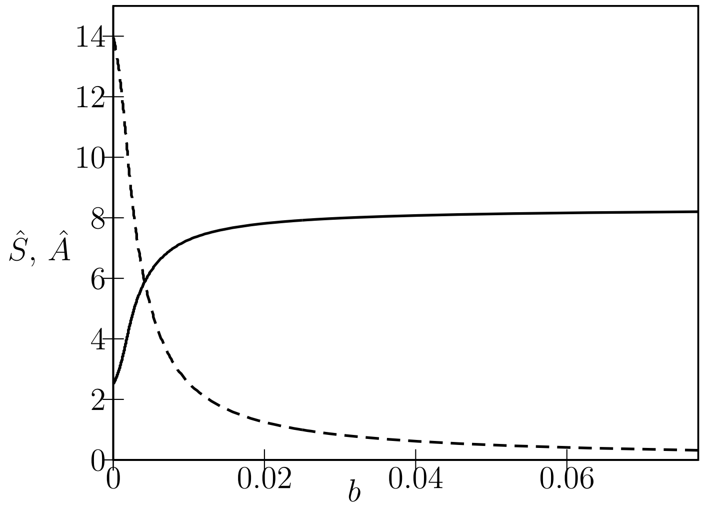

Figure 2.5: Influence of parameter b on the steady state, where the solid line represents the number of users and the dashed line the number of susceptibles

the number of actual customers. When parameters do not work in favor of the decision maker, i.e. they are responsible for why the product is not as successful as it could be, he should give larger price discounts in the steady state to compensate the lower flow. It would then definitely make sense for the company then to investigate why this product is not attractive to potential customers and if possible eliminate the problem. How this should look like, however is not captured by the model.

μ: When the outflow rate from users, reflecting the consumer loyalty, is high then one would try to compensate the higher outflow by giving larger price discounts, yet, the number of users would still be smaller in such an equilbirium as if the consumer loyalty was big. Since then the influence of the users is low, the number of potential customers would be rather big. In such a case it would also make sense for the decision maker to try finding ways to attach the customers more to the brand of the product.

2.3 DNSS case

Recalling now Fig. 2.4, it could be seen that if the initial number of customers is very small, it is only the innovators who start adopting the product in the beginning, since the influence of the very few existing users is too small to have any real impact. But then, what happens if there are no innovators and

Table 2.4: Parameter values for the DNSS case

r	π	k	δ	μ	a	b	β	ω	α
0.04	1	1	0.05	0.12	0.00	0.025	1	0.75	1.75

Table 2.5: Admissible steady state values for the parameters shown in Tab. 2.4

$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
20	0	0	6.25	0	0	0
1.01	7.91	6.44	7.07	$0.98 * 1^{-5}$	0	0

susceptibles are only attracted to a certain product when other people own it too? For instance, certain products are only useful if there is a certain number of other people using it (e.g., network goods, such as a telephone, or peer-to-peer sofware) or if people are very risk averse regarding the introduction of some new innovation. In terms of the model, this would mean that parameter a is very small or zero.

Unlike before it is now possible to find more than one steady state (see Tab. 2.5 and Fig. 2.6): One, similarly to the one shown in Sect. 2.1, consists of comparably many customers, but only a few susceptibles, and there is another one with many susceptibles, but no or only a very few consumers, due to the lack of innovators. These steady states are separated by a DNSS curve (also known as Skiba curve, see Appendix A), on which a decision maker can optimally decide between a solution with finally many users (and high control efforts) or no customers. Fig. 2.6 shows the phase portrait, where one can see these equilibria and also the DNSS curve, which is depicted even more closely in Fig. 2.7.

Intuitively, one might think that it always better to choose a solution leading to the steady state with many paying customers, this, however, is not the case. One reason is that it is not always possible to find a solution for every initial value leading to the "better" steady state, i.e. when the number of exisiting users is too small they are not able to exercise enough impact to convince any susceptible to use the product. On the other hand if the number of users is large enough in the beginning it is not possible to reach the steady state with no users, since the positive influence on other people's

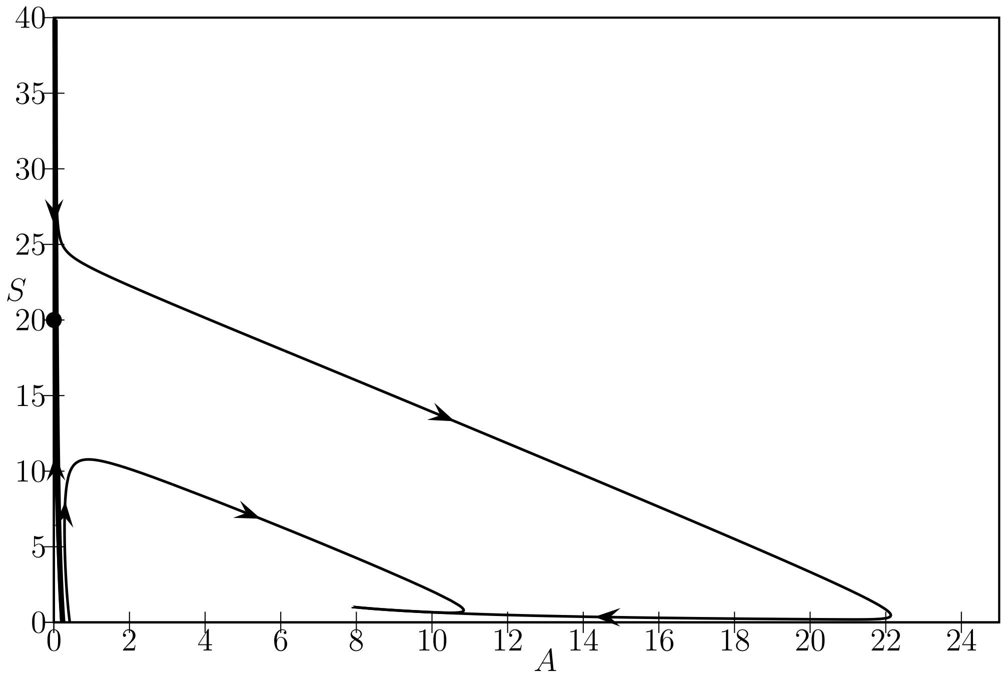

Figure 2.6: Phase portrait of the DNSS case depicting the two steady states and the DNSS curve as well as several trajectories leading to the steady states.

adoption decision of the exisisting users is so strong that their number cannot fall beyond a certain level.

This is reflected in Fig. 2.8, depicting the value of the objective function depending on the initial states; this figure also reveals that for some initial values of S and A a decision maker really has the choice whether he wants to finally have many or no customers. The reason for this can be found in the shape of the stable manifold of the second steady state. As already mentioned, it is not possible to find a candidate for an optimal solution for every initial point ending in the steady state with the many users, particularly if the number of existing users is small. The number of initial users for which this steady state can be reached is only very, very little lower than their number on the DNSS curve that can be seen in Fig. 2.7. Starting on a point on or very close to the DNSS curve, it can be shown that the time needed to reach the higher steady state, as well as the parts of the optimal solution where the number of users is, despite high control efforts, small, are rather large. The closer one is to the minimum number of users needed to be able to reach the steady state with many customers, the longer is the time needed

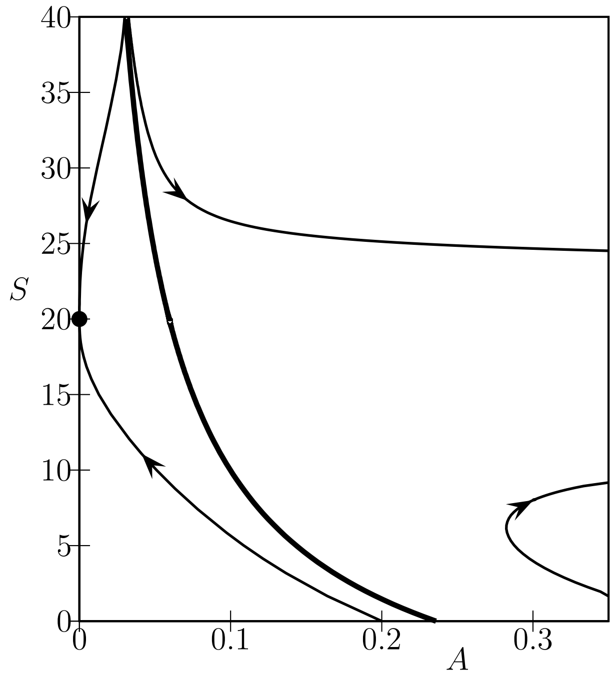

Figure 2.7: Zooming (in different scale) of the threshold shown in Fig. 2.6.

to reach it as well as the required control efforts, and resultingly, the smaller are the profits. At some point, then, it might be better to give little or no price reductions and accept that this product will ultimately fail, than to reduce the price so that no profits are made for the considered prefered time period.

Starting near the DNSS curve depicted in Fig. 2.7 one would choose following pricing strategies: If the number of users would be so small initially that the product is bound to fail, then one would give some rather small price discounts as long as the number of users is not too small. The purpose of this strategy is to use the little, but still existing influence of these customers to find a few additional customers contributing to the small profits a company can make under such conditions. If the number of users is large enough, that the second equilibrium can be reached, one would give price discounts as big as possible as long as the number of customers and their impact on others to imitate them is small in order to provide additional incentives for potential customers to start adopting the product. However, as the number of users and their impact on others is large enough, one would not need to

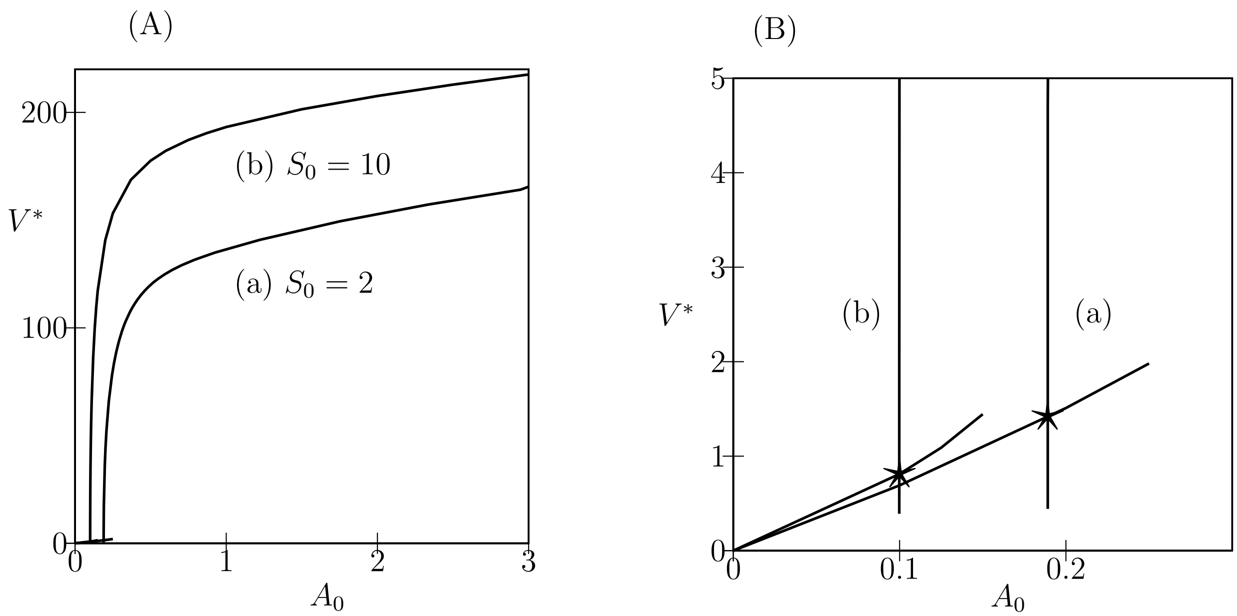

Figure 2.8: Value of the objective function for different initial values of users and susceptibles (a) $S_0 = 2$ and (b) $S_0 = 10$. Panel (B) is a magnification of panel (A) showing the DNSS points, denoted by a $\star$.

give large price reductions anymore and could exploit the growing influence of the customers on others until the steady state is reached. The optimal control spendings also depends on the number of susceptibles. If there are many of them one would give larger incentives to them to adopt the product, because any control efforts would be more efficient then.

Note that on trajectories ending at the steady state with many users the value of both costates are very big on points close to the DNSS curve. This reflects that with little more users and susceptibles the overall outcome would be much better since it would be easier and faster to gain additional users then. Of course the costate corresponding to the number of users is much larger than the one corresponding to the susceptibles, which shows again that additional user is better for a decision maker, because he directly contributes to the profits of a company.

Remark on the Numerical Calculations *A DNSS point can be found by solving the following problem: Find two points with the same state values lying on the stable manifolds of two different steady states, where the values of the Hamiltonians (and therefore also of the objective functions) are equal. In order to find such a point numerically, one can calculate the solution paths for different initial state values lying on a line connecting the two steady states. For each initial point there might exisit two candidates for an optimal solu-*

tion, each leading to a different steady state. Which one is optimal depends on where the value of the Hamiltonian is larger. Then, at some point, one might find a DNSS point where the value of the Hamiltonian of both solutions are equal. Once having found such a DNSS point, one can use it as initial solution for the search of the other points of the DNSS curve.

Fig. 2.8 reveals not only that the value of the objective function becomes greater the bigger the number of existing users is, but also that the profits of a company are higher if the initial number of potential customers, i.e. the market for a certain product, is larger. The reason for the first is that on the one hand the profits increase if there are more paying customers, on the other hand less control efforts are required for positively influcencing people in their product adoption decision. A higher number of susceptibles means that more people can be convinced starting a certain product over time. The DNSS point also depends not only on the number of existing users. If there are many susceptibles, it is "easier" for existing customers to exercise any influence on anyone, therefore less initial users are required to make a product successful if the number of consumers is larger. Although the profits made on the DNSS point with many susceptibles are smaller than on the other DNSS point, one would still make more profit for the same initial number of users if there were little susceptibles for the previously described reasons.

When ending in a steady state with no users, profits are obviously lower than if there are finally many users. Note that unlike in the case where only one steady state served as candidate for an optimal solution, now, if there are initially now users, there will be absolutely no profit. This is because the number of users will never increase.

The shape and the location of the DNSS curve also depends on the parameters used, e.g. if the fraction of innovators is larger, then less users are required to make a product successful. Similar is valid for the impact of existing users b. However, if a becomes too large, the steady state with no users will disappear and if b becomes too small (and a is not too big) the steady state with many users vanishes. On the other hand if the outflow rate of users μ becomes larger, then it becomes harder to reach the steady state with many users, as they have due to their smaller size less influence on the susceptibles.

2.4 Limit Cycle

Unlike in the previous cases, an optimal solution does not necessarily always have to end in a fixed point, but can also converge towards a limit cycle if

Table 2.6: Parameter values for the limit cycle

r	π	k	δ	μ	a	b	β	ω	α
0.04	1	1	0.05	0.5	0.02	0.0268	1	0.75	1.75

the dynamical system and the used parameters allow this. It is rather easy to find cycles if the quitting rate of users μ is high, because then basically the outflow of customers can not sufficiently compensated by the inflow to users to ensure ending in a steady state.

Remark on the Numerical Calculations *Limit cycles can be found, e.g., with the help of the Matcont toolbox; cmp. Dhooge et al. (2003) and Dhooge et al. (2006). The usual technique is to find a Hopf bifurcation of a steady state, i.e., where the real parts of the eigenvalues of the Jacobian become zero for a certain parameter. Having located one cycle, this can be continued by changing a parameter. Note that a limit cycle has to fullfill the boundary condition, that a trajectory must reach its starting point again after the time period Θ, and some phase condition to pin down a specific representation of the limit cycle since each trajectory starting at a point of the limit cycle represents the same geometrical object.*

For the calculation of the monodromy matrix, which determines the stability properties of the cycle, I refer the reader to, e.g., Grass et al. (2008).

The cycle depicted in panel (A) of Fig. 2.9 has already served as some introductory example in Chapt. 1. This section here will provide some more detailed information. The parameters of this cycle can be found in Tab. 2.6. Parameter b was chosen so that the cycle is the largest found (in terms of the distance between the maximum and minimum values of the state variables), if the other parameters remain the same. The value of the states of the steady state found for these parameters are shown in Tab. 2.7, this steady state however does not serve as a candidate where an optimal solution can end, it is unstable. The eigenvalues of the monodromy matrix of the cycle are $\xi_1 = 142.067$, $\xi_2 = 2.9537$, $\xi_3 = 0.0208$, and $\xi_4 = 1$, since one eigenvalue of the monodromy matrix is smaller than one and one has the value one, the cycle has a two-dimensional stable manifold. Any solution lying on it leads to the cycle. The period of such a cycle is $\Theta = 27.0763$.

One can differentiate between four phases of a cycle, in the first one, denoted as (I) in the phase portrait and the corresponding time paths (Fig.

Table 2.7: Steady state values for the cycle

$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
8.4522	1.1548	1.6077	3.7258	0.1609	0	0

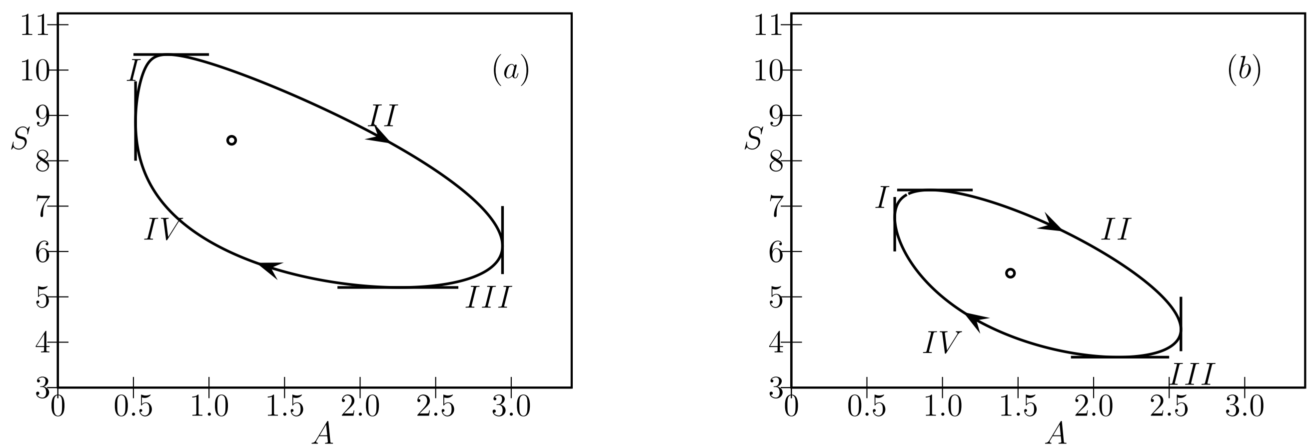

Figure 2.9: Cyle in the price reduction model. Panel (a) shows a cycle occuring for the parameters listed in Tab. 2.6 with parameter $b = 0.0268$, in (b) parameter b was set to 0.05. $\circ$ are steady states which are not candidates for where an optimal solution can end, they are however relevant for the limit cycle.

2.10), both the number of actual and potential customer rises. As the influence of existing users is not particularly strong it is initially a waste of resources to give any price reductions then as can bee seen in Fig. 2.10. This changes, however, as the number of susceptibles and therefore the potential future paying customers rises as well as the number of existing customers due to the rising number of innovators. A company has then a larger incentive to promote its product with the help of a price reduction. After some time, in the second phase (II), because of the larger flow from the susceptible to the user state due to the stronger impact of the control instrument and the existing users and because of its constant inflow, the number of susceptibles can not grow anymore and even starts to decrease. Those lost susceptibles become new customers adding to the profits of a company by spending money on the product and by serving as example to others, leading to a further increase of users. In this phase it first makes sense to spend much for efforts making the product more attractive to potential customers, but as their in-

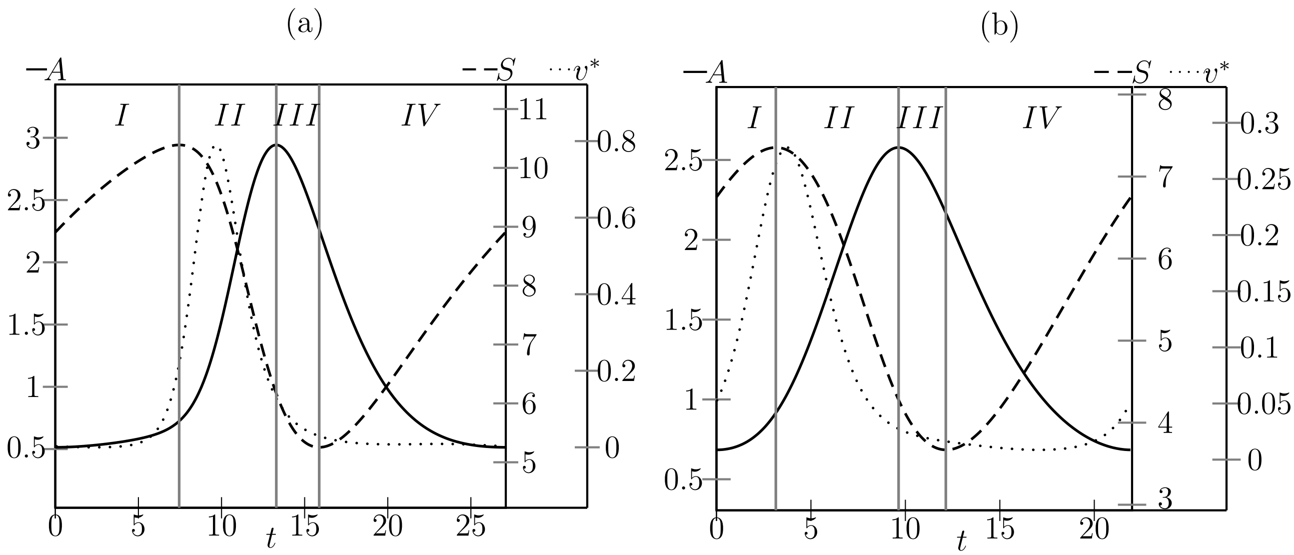

Figure 2.10: Timepath of the cycles of Fig. 2.9, where (a) $b = 0.0268$ and (b) $b = 0.05$.

fluence grows stronger and stronger it simply is not neccessary any more. However, after a while the pool of susceptibles gets so low that the number of users also starts to fall. The reason for this is that there are simply not enough susceptibles left who could, by deciding to adopt the product, compensate the fast decrease of consumers due to the particularly large number of users quitting the system in phase (III). Due to the lack of people who can be convinced to become users it also does not make sense to give large price reductions, because it simply would not be effective. As the number of users decreases and the flow from one state to the other along with it, the number of potential can slowly recover (phase IV).

The reason why the cycling now occurs is that the number of users decreases so fast in phase III and IV of the cycle, that, while the impact of control spendings would be too small to change much, the number of susceptibles can not recover sufficiently fast to compensate the higher outflow and stabilize it at some fixed level. On the other hand when the number of users grows that happens too fast to allow ending with a fixed number of users and susceptibles.

Figs. 2.9 and 2.10 also reveal the impact of parameter b on the shape of the optimal solution: In the first case (panel (A)) b is assumed to be 0.0268 in the second case to be 0.05. Interestingly, the number of users and

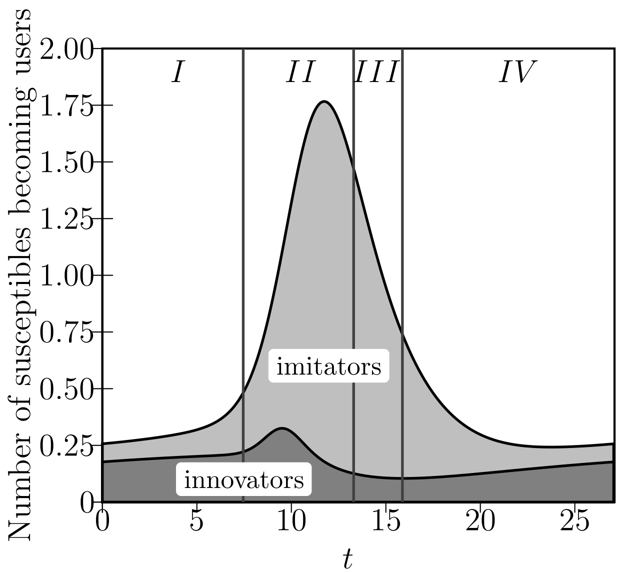

Figure 2.11: Number of susceptibles starting to use a certain product at different times of the limit cycle

susceptibles can grow larger if this parameter is small[4], and the time period of the cycle is larger in the first case with $\Theta = 27.0763$, compared to the second case with $\Theta = 21.934$. If the fraction of imitators $b$is small, initiation has to be at least partially supported by giving higher price discounts, making the product more attractive. Yet it takes longer to build up some consumer base and if the number of users is low, the number of susceptibles can better recover. If, however, one would lower parameter b even further, the control spendings would become less effective and because of the lower initiation less people would become users over time.

Fig. 2.11 shows how the flow from susceptibles to users develops over time on the cycle shown in Fig. 2.9 (A), distinguishing between the share of innovators and imitators of this flow. It can be seen that if the number of users is small in phase I of the cycle, but the number of susceptibles is relatively large, the share of innovators becoming users is larger than the share of imitators. The reason for this is, obviously, that the existing number of users is so small that they do not have much influence on the product adoption decision of others. This means, that in this phase the product is a good, that is mainly used by people who are attracted to it independent of others. However, after a while the pool of potential customers becomes so large that

[4]However, if this parameter would be smaller the cycle would be so too due to the less efficient customer acquisition.

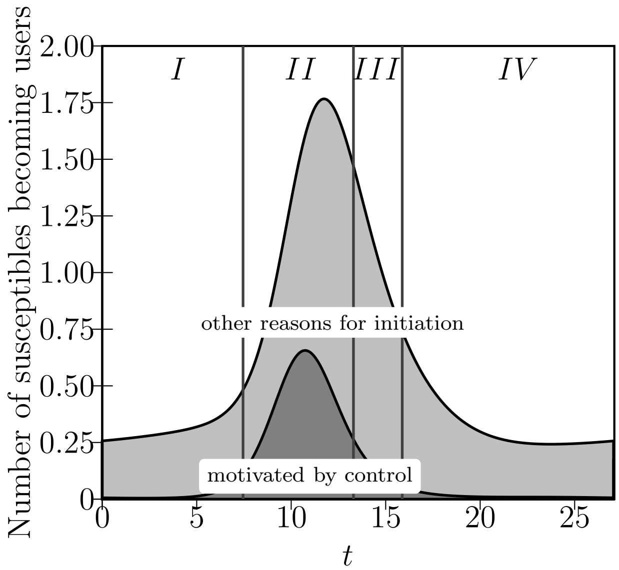

Figure 2.12: Number of susceptibles starting to use a certain product at different times of the limit cycle (not) under direct influence of the control

it makes sense to give larger price reductions in order to exploit the larger market size. As the control effort rises in the second phase, the number of imitators grows, as well as the number of innovators. Yet, the growth of imitators is larger, since the growing number of users means that they now have a stronger impact. Here the product becomes popular, however, as the number of potential consumers is exhausted, there are not enough people to sustain the high level of product usage, and the flow from susceptibles to users declines. The number of innovators decreases simultaneously with the decreasing application of the control (but also due to the smaller number of susceptibles). Because of the still existing influence of the users, the decrease of imitators takes longer to start. This decline of imitators in the initiation lasts from the second phase to the fourth, however, the number of innovators can increase again in the fourth phase due to an increasing number of susceptibles.

Fig. 2.12 shows which share of the flow from susceptibles to users starts using the product because of the incentives caused by control application, i.e. a lower price. We can see that the rise and decline of the proportion of people starting to use a product because of the control efforts happens always earlier than if they start adopting the product for other reasons. This reflects that the control efforts are mainly required to make the product rather popular and build some kind of consumer base, who can exercise some

Table 2.8: Parameter values for the DNSS & limit cycle case

r	π	k	δ	μ	a	b	β	ω	α
0.04	1	1	0.05	0.3	0.00	0.03	0.25	0.75	1.75

further influence on potential adopters so that when their number is large enough, no further convincing is needed. It can also be seen that if the number of users as well as the number of susceptibles is small it is rather useless to give any price reductions, since it would not lead to the sufficient creation of additional users to compensate the lower profits. The fraction of people convinced by marketing measures should be rather small in such phases.

2.5 DNSS Curve & Limit Cycle

For the parameters given in Tab. 2.8, it is possible to find a situation where, depending on the initial number of potential and actual customers, the decision maker might have the choice of either letting the product fail or to choose a solution path leading to a limit cycle, meaning that the number of susceptibles and users will always oscillate. There are again no innovators, i.e. parameter a is zero. Note that like in the previous section parameter μ, i.e. the quitting rate from the consumers is rather large and that parameter β was only decreased[5] in order to avoid the violation of control contstraints on the limit cycle, which would complicate the numerical computations, however, it would also be possible to find such a case with a larger β.

Table 2.9 shows again that there are two relevant steady states can be found, the first is a candidate for an optimal solution, with no users and many susceptibles, the second is not because of its instability. Close to this first steady state again it does not make any sense to give price reductions, because it would hardly motivate anyone to start using the product. People then are only attracted to the product if other people use it and this is not the case here.

Fig. 2.13 now shows the occuring DNSS curve and the limit cycle. If one

[5]Parameter β weights the impact of the control on the initiation. If it becomes smaller that would mean that the application of the control becomes less effective. As such, the higher number of users attracted by the lower price, which are won under higher efforts, cannot compensate the lower profits caused by the lower price.

Table 2.9: Steady state values for the parameters shown in Tab. 2.8

$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
20	0	0	2.9412	0	0	0
5.0342	2.4943	4.3841	7.0385	$0.4954 * 10^{-3}$	0	0

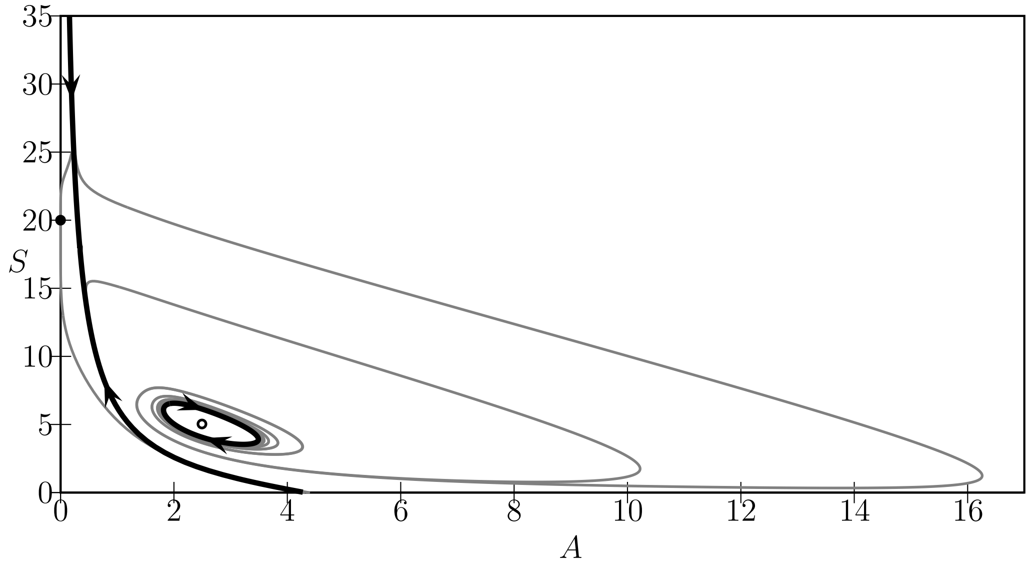

Figure 2.13: Phase portrait for parameters where a limit cycle and a steady state are seperated by a DNSS curve.

would start on the left side of this DNSS curve, it would be optimal to let the product fail, because on the one hand close to the DNSS curve the influence of the exisiting users on the potential customers would be too small to have any real impact, i.e. no one would serve as a positive example demonstrating the capabilities of the product, or if it is a network good, there would be not much utility of the product if the number of people owning it would be too small. A decision maker would not give large price reductions particularly if the number of initial users was very small, since these efforts would essentially only take away the profits. He would give some price reductions, however, if there were at least some users of the product in the beginning in order to strengthen their impact and win at least some additional customers who contribute to the profits of the company.

If starting on the right side of the DNSS curve the optimal solution would be cyclical. Fig. 2.14 depicts the timepath of a trajectory lying on the limit

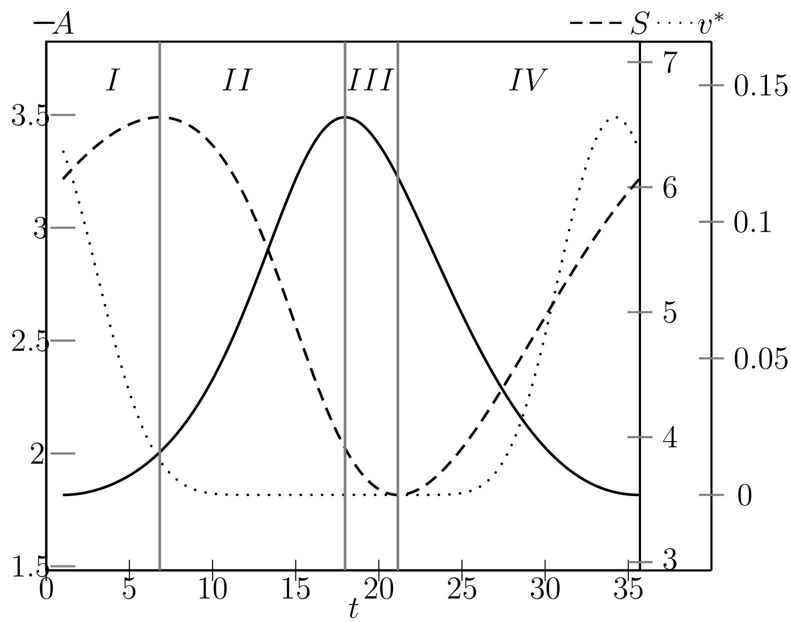

Figure 2.14: Timepath of the cycle in the DNSS & limit cycle case

cycle: Then, if the initial number of users is rather small or the number of susceptibles particularly high, one would give large price discounts in order to give additional incentives to people to start buying the product. After some time, however, the flow from the first to the second state becomes so large that the number of potential customers starts to decrease due to its constant inflow. One can reduce the control efforts then due to the large impact of the customers. As in the previous section, after a while the people starting to adopt the product becomes so small, that the low customer loyality cannot be compensated anymore and the number of users decreases. Then, however, the number of susceptibles can recover again, so that there are again enough people that can be influenced by the customers. One can give some additional incentive by reducing the price, that the number of users can rise again.

Starting exactly on the DNSS curve a decision maker would have the choice between these two solution possibilities, because again the time needed to build up a larger customer base would be so long and the price reductions so large, that it would be equally good for the company to do nothing or at least not much, exploit the profits caused by the few existing users and let the product fail.

Fig. 2.15 shows how the value of the optimized objective function depends on the initial number of potential and actual customers of a certain good. Again it can be seen that it is better to have many actual and/or potential

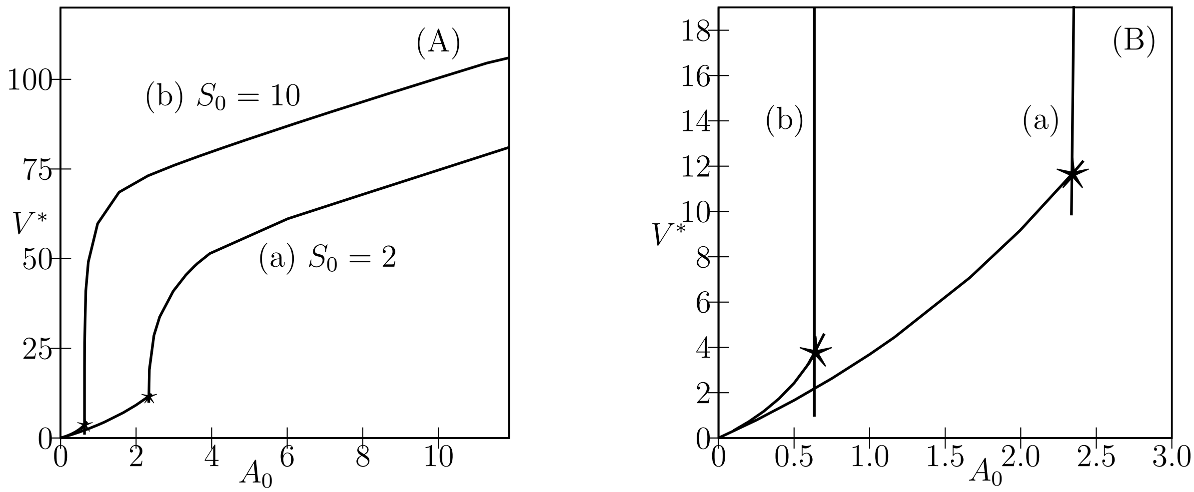

Figure 2.15: Value of the objective function for different initial values of users and susceptibles in the case where the DNSS curve seperates a steady state with no users and a limit cycle where (a) $S_0 = 2$ and (b) $S_0 = 10$. Panel (B) is a magnification of panel (A) showing the DNSS points, denoted by a $\star$.

users in the beginning - having a large customer base means after all that the product has some attractiveness because many other people use it and on the other hand a large market potential is good, because there are many people who can be reached with such marketing measures, leading over time to a larger number of customers. It can be seen that if the initial number of users is small and the product is bound to fail, the profits of the company are not particularly high.

This changes, however, rapidly when the number of initial users reaches a certain level; then, if the time needed to create a profitable customer base is not too long and control efforts are not too high, one could make larger profits. The stars in Fig. 2.15 again show the DNSS points for different initial values of the potential customers. It can be seen that the objective value is always larger if there are many of them as well as that if their number is large one needs less initial users to be able to gain a cyclical development of the products, because the users have a larger audience to which they make the product seem attractive.

Chapter 3

A Two-Stage Marketing Model Approach

In the previous model it was always assumed that the system evolves according to the given system and the same parameters. However, this might not always be very realistic. Sometimes, shocks might occur to the system, e.g., caused by the market entrance of a competitor, that can alter the system behavior significantly. A decision maker might or might not have under certain conditions influence on the occurrence and impact of such a shock to a certain degree. In order to find out how a change of the system and/or parameters influences the optimal solutions one can use the methods of multi-stage modeling, a short description of them can be found at the end of Appendix A. In the following sections the previously described marketing model will be extended by introducting two stages, which differ in the parameter values used.

3.1 Market Entrance of Competitors in the Price Reduction Model

The basic idea behind the price reduction model is that a decision maker, who wants to influence the sales/distribution of a certain product, can give price discounts in order to make this product more attractive to potential customers. Of course, any reduction of the price/profits has negative consequences on the revenues caused by each customer, but since this lower price gives some additional incentive to buy the product, who might considerably strengthen the influence of the consumers on potential buyers, the negative effect might be compensated. The decision maker's control instrument in this model is the percentage reduction of the average profit per user (e.g.

achieved by a certain price reduction).

Here we want to consider a two-stage version of the model, which is based on Caulkins (2007). In the first stage the decision maker is a monopolist, meaning not only that he is the only one providing such a good, but also that there are no alternatives on the market that a consumer could use instead. He does not face competition, because it is assumed that he either is protected by a patent or governmental regulation, which serve to protect the interests of either the the producer of the good or the consumer of the good. The purpose behind such measures is that by enforcing some kind of regulation a government becomes able to exert some influence on the price, on the supply of the good, on the provided infrastructure, etc. Patents allow its owner exclusive rights regarding production, distribution, sale, import and use of this inventive, new good or technology and are provided to companies in order to protect a newly developed product (in order to reward research work) for a certain time period, which varies according the country and/or agency by which it is granted. As a consequence the company faces no pressure regarding the price by competition, and the (exogenously determined) constant market price can be rather high leading to a larger average profit per user π.

It is assumed that after a certain time t_s the monopoly breaks up and suddenly the former monopolist is confronted with perfect competition, which can occur when the barriers to market entry are low, meaning that it is rather easy for competitors to produce and/or sell a similar good. Reasons why someone has to give up a monopoly would either be patent expiration or an opening of the market due to some governmental organisation.

Stage 2 reflects the impact of competitors on the system. In this simple case the two stages only differ in their parameters, particularly in the price/profit, initiation parameters (the share of innovators and imitators) and the quitting rate. It depends on the nature of the product and the market how the parameters change. Usually a company would loose, due to the alternative products of the competitors, consumers, attractiveness and direct revenues because of a potentially lower market price. However, competitors do not only have to be associated with negative impacts, Kotler & Armstrong (2008) mention that competition can also have several advantages, such as an increased total demand, shared costs for product development, a help for legitimizing a new technology, a higher product differentiation, a lower risk of getting trouble with antitrust authorities, improve bargaining power against labor unions and regulators etc. A monopolist might then at some point have some incentive to give up monopoly. It is assumed that no costs are attached to switching between stages.

3.1.1 The Model

The company faces following optimization problem with the stages i, where $i = 1, 2$. Note that now the parameters of the different stages are emphasized by different subindices i or the number of the corresponding stage, respectively.

$$\max_v \int_0^{t_s} e^{-rt}(\pi_1 A(1-v))\mathrm{d}t + \int_{t_s}^{\infty} e^{-rt}(\pi_2 A(1-v))\mathrm{d}t$$

$$\text{s.t. } \dot{S} = k - \delta S - f_i(A)Sg(v) \tag{3.1}$$

$$\dot{A} = f_i(A)Sg(v) - \mu_i A \tag{3.2}$$

$$0 \le v \le v_{\max} \le 1$$

It must be the decision maker's objective to maximize his profits now over both stages, where the application of the control variable leads to decreased profits per user, but also to an increased flow from susceptibles to users and, resultingly, to an increase of paying customers.

The functions describing initiation for stage i are now $f_i(A) = a_i + b_i A^\alpha$. It is assumed for simplicity that the efficiency of application of the control does not change, and function $g(v) = 1 + \beta v^\omega$, where $\omega < 1$ in order to fulfill the Legendre-Clebsch condition if no constraints are active (see App. C).

The Hamiltonian function of each stage is given by

$$\mathcal{H}_i = \pi_i A(1-v) + \lambda_1(k - \delta S - f_i(A)g(v)S) + \lambda_2(f_i(A)g(v)S - \mu A), \qquad i = 1, 2. \tag{3.3}$$

and the Lagrangian function by $\mathcal{L}_i = \mathcal{H}_i + \nu_1 v + \nu_2(v_{\max} - v)$.

The optimal control for stage i can be derived with the help of

$$\mathcal{L}_v = -\pi_i A + (\lambda_2 - \lambda_1)\omega\beta v_i^{\omega-1} f_i(A)S + \nu_1 - \nu_2 = 0 \tag{3.4}$$

and the Legendre-Clebsch condition is fulfilled if no control constraints are violated in stage i only if

$$\mathcal{L}_{vv} = (\lambda_2 - \lambda_1)\omega(\omega-1)\beta v_i^{\omega-2} f_i(A)S \le 0. \tag{3.5}$$

Again the complementary slackness conditions have to be fullfilled which are

$$\nu_1 v = 0, \qquad \nu_2(v_{\max} - v) = 0, \qquad \text{and} \qquad \nu_1, \nu_2 \ge 0$$

for both stages.

We know from the matching conditions that it is optimal to switch from the first to the second stage if the value of the Hamiltonian[1] function of both stages are equal, i.e. $\mathcal{H}_1 = \mathcal{H}_2$. If one Hamiltonian is larger than the other one would optimally always choose being in the stage with the higher value, however, if this is not possible, e.g. when a patent expires one has no other choice but to accept the lower value.

Finally, the costate equations of the two stages have following shape in stage i:

$$\begin{aligned} \dot{\lambda}_1 &= (r+\delta)\lambda_1 + (\lambda_1 - \lambda_2) f_i(A) g(v) \\ \dot{\lambda}_2 &= (r+\mu_i)\lambda_2 + (\lambda_1 - \lambda_2) f_{iA} S g(v) - \pi_i(1-v) \end{aligned}$$

Remark on the Numerical Calculations *While the OCMat toolbox already provides the possibility to initialize and find the needed analytical results, most of the existing functions do not yet include the possibility of dealing with more than one stage. However, by creating and initializing two models, each describing one stage of the problem, one can already use the toolbox for dealing with such problems.*

3.1.2 Changes of the System Behavior Due to Entering Another Stage

In the following section we shall see how the change of which parameters affects the system and the optimal solution when switching between the two stages. This is done by analyzing which parts of the state equations are affected by a parameter change, by trying to analytically find out what happens to the optimal control at the switching point and by looking at the value of the Hamiltonians of the different stages.

The State Equations

A decrease of the initiation parameters means that the flow from susceptibles to users will become smaller if this decrease is not sufficiently compensated by additional application of the control instrument. By a decreasing consumer loyalty, reflected by a larger quitting rate μ of the consumers, the actual number of users becomes smaller or at least grows slower, not only directly

[1] Therefore also of the Lagrangian function, since the values of both functions are equal due to the complementary slackness conditions, which also have to be fulfilled in both stages.

due to this higher outflow, but also indirectly through the lower initiation caused by a smaller number of users.

The Optimal Control

The state equations and therefore the development of the states is not only directly affected by the change of the parameters occurring in the state (and costate) equations, the change of the optimal control must also be considered. Therefore, let us take a look at the optimal control at the switching point when no control constraint is active, i.e. $0 < v < 1$ and $\nu_1 = \nu_2 = 0$. By solving (3.4) it can be found that

$$v_i^* = (\frac{(\lambda_2 - \lambda_1) f_i(A) S \beta \omega}{\pi_i A})^{\frac{1}{1-\omega}}, \qquad i = 1, 2. \tag{3.6}$$

It can be seen that, unlike the other changing parameters, the quitting rate of the users μ does not directly influence the optimal control. Assuming now that the parameters of the initiation function a and b change to $100\chi_f$ percent of its original value, then we find that

$$f_2(A) = \chi_f f_1(A).$$

Now, the average profit per user decreases to $100\chi_\pi$ percent of the original value, meaning

$$\pi_2 = \chi_\pi \pi_1.$$

Inserting this into Equ. (3.6) in order to express the optimal control of stage 2 v_2^* as a function of the optimal control of stage 1 v_1^*, we find that

$$v_2^* = \left(\frac{\chi_f}{\chi_\pi}\right)^{\frac{1}{1-\omega}} v_1^*,$$

where due to $0 \leq \omega \leq 1$, the exponent $\frac{1}{1-\omega}$ is always larger than one. It can be seen that the relation between χ_f and χ_π determines the change of the control at the switching point in the following way under the assumption that both the price and initiation decrease in the second stage, e.g., due to the market entrance of a competitor:

- $\chi_f > \chi_\pi$: The control will jump to a higher level at the switching point if the decrease of initiation is smaller than the decrease of the profits. This means that a decision maker has to do more in order to compensate the lower profits of the second stage. The same also applies if initiation increases and the profits fall in the second stage or if the increase of initiation exceeds the increase of the profits. Note that a higher initiation also means that control efforts become more efficient.

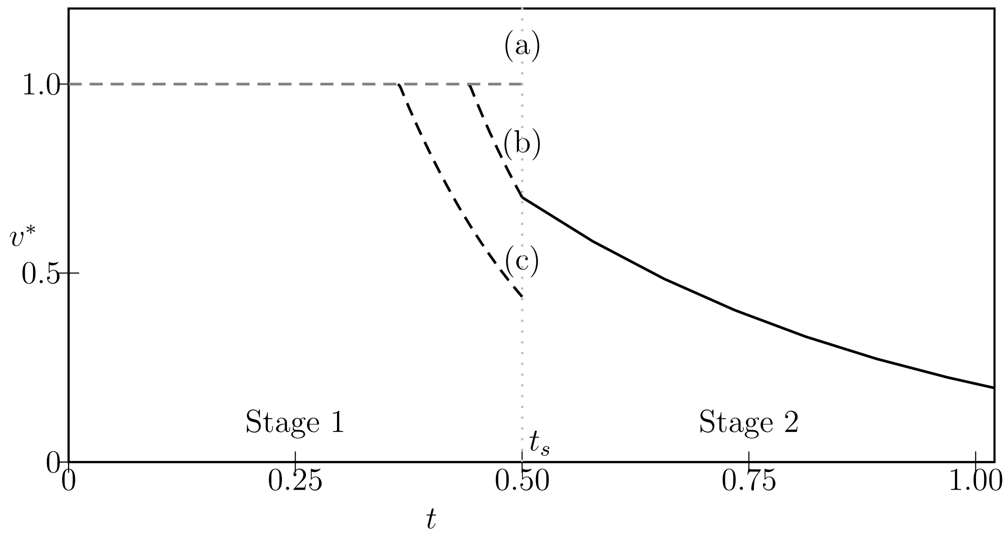

Figure 3.1: Jump of the optimal control at the switching point where the relation of the change of initiation and average profit is in case (a) $\chi_f < \chi_\pi$ in (b) $\chi_f = \chi_\pi$ and in (c) $\chi_f > \chi_\pi$. A dotted part of the optimal solution is located in the first and the solid part in the second stage.

- $\chi_f < \chi_\pi$: Here the control will also jump, but into the opposite direction as before: the initiation decreases percentage-wise more than the price. This would mean the application of the control becomes less effective and any other strategy than lowering the control would lead to a loss of profits that could not be compensated by the increased initiation.

- $\chi_f = \chi_\pi$: The control does not change at the switching point - i.e. it evolves continuously. Since neither the the impact of the change of profits nor of the change of initiation exceeds the other, it is not necessary to adapt the control in case of a switch from the first to the second stage.

Fig. 3.1 illustrates how a the parameter change affects the control: Here it is assumed that stage 2 has always the same parameters as illustrated in Tab. 3.1 and in stage 1 $f_2(A) = 0.5 f_1(A)$. However, three different prices are considered in the first stage. We will see, that a decision maker would basically choose similar strategies, i.e. give large price reductions if the number of users is small and do less and less as their number decreases.

The starting points of the three trajectories are rather close, yet not completely similar as the initial values of the state variables are chosen in

Table 3.1: Parameter values of stage 2 for the comparison of the jump of the control; parameters different in stage 1 are $a_1 = 0.02$, $b_1 = 0.05$, $\mu_1 = 0.0976$ and (a) $\pi_1 = 1.8$, (b) $\pi_1 = 2$ and (c) $\pi_1 = 2.25$ and $\chi_f = 0.5$

r	π_2	k	δ	μ_2	a_2	b_2	β	ω
0.04	1	1	0.05	0.12	0.01	0.025	1	0.75

a way that part of the trajectories are exactly the same in the second stage regarding state and costate values as well as optimal control application.

Remark on the Numerical Calculations *Since the numerical calculations are done backwards in time starting at the steady state of stage 2, this is done to ensure to have the same switching point for all three cases to have some comparability.*

The initial points are such that there are little users and some susceptibles, therefore similar optimal control strategies have to be chosen in all three cases, i.e. give initial large price reductions as large as possible in order to attract potential customers and then reduce efforts to exploit the higher profits. However, the timing of when starting to do this differs *before* it is optimal to switch between the stages and will be explained in the following:

(a) $\pi_1 = 1.8$ (or $\chi_\pi = 0.56$) In this first case the percentage reduction of the initiation is larger than the percentage reduction of price. That means at the switching point the control will jump down: A decision maker will put more control efforts into the first stage. The reason for this is while the initiation is the same in all three cases the profits are the lowest here. That means that one would have to do more to gain users who contribute to the profits. Then, when switching to the second stage, profits suddenly become larger in relation to the initiation - in order to exploit these bigger profits one would not have to do as much as before.

(b) $\pi_1 = 2$ (or $\chi_\pi = 0.5$) Here the price of the first stage is twice as big as in the second stage. Then, as initiation and price decrease for the same percentage, the optimal control can evolve continuously at the switching point.

(c) $\pi_1 = 2.25$ (or $\chi_\pi = 0.44$) Thirdly, the percentage decrease of the price is larger than of the initiation, resulting in an upwards jump of the control at the switching point: Because of the highest price in the first stage, a decision maker would try to take advantage of it then and would not give as much price reductions as in the other cases. Then, in the second stage, the application of the control becomes more efficient compared to exploiting the profits.

What is particularly interesting is that it can be seen that the relation of the change of the two components of the model on which control has a impact, namely initiation and the profits, determines how the optimal strategy has to be adapted when switching between the stages.

The Hamiltonian

When switching optimally from the first stage to the second, the values of the Hamiltonians of both stages have to be equal, cp. Makris (2001), Grass et al. (2008). This does not have to be the case if the switching time is exogenously given. By using (3.3), the difference between the two Hamiltonians of the two stages can be written as

$$\begin{aligned} X &= \mathcal{H}_1 - \mathcal{H}_2 = \\ &= (\pi_1 - \pi_2)A(1-v) + (f_1(A) - f_2(A))(\lambda_2 - \lambda_1)Sg(v) + (\mu_2 - \mu_1)A \end{aligned} \tag{3.7}$$

It is optimal to switch from the first stage to the second when $X = 0$, optimal to be in stage 1 if $X > 0$ and in stage 2 if $X < 0$. Using (3.7) we find that $X = 0$ if

$$\frac{\pi_1 - \pi_2}{f_1(A) - f_2(A)} = -\frac{(\lambda_2 - \lambda_1)Sg(v) + (\mu_2 - \mu_1)A}{A(1-v)} \tag{3.8}$$

is fulfilled. Expressing now again the second stage parameters and initiation function as function of the first stage parameters, i.e. $f_2(A) = \chi_f f_1(A)$, $\pi_2 = \chi_\pi \pi_1$ and $\mu_2 = \chi_\mu \mu_1$, (3.8) can be rewritten as

$$\frac{(1-\chi_\pi)\pi_1}{(1-\chi_f)f_1(A)} = -\frac{(\lambda_2 - \lambda_1)Sg(v) + (\chi_\mu - 1)\mu_1 A}{A(1-v)} \tag{3.9}$$

The assumption that the initiation function and changed parameters have to be greater than zero means that also $\chi_f, \chi_\pi, \chi_\mu \geq 0$.

Let us now consider the following cases that will be described later in more detail:

(-) Initiation and the profits decrease in the second stage (e.g. by market entrance of a competitor), i.e. $0 \leq \chi_f, \chi_\pi \leq 1$. In this case (3.9) cannot be fullfilled if in this second stage the quitting rate of existing users increases or remains the same ($\chi_\mu \geq 1$), as the fraction on left side is always positive (or zero) and the fraction on the left side due to the sign before it is always negative since $S, A, \pi_1, f_1(A), \mu_1, g(v), (1-v), (\lambda_2 - \lambda_1) \geq 0$. The economic interpretation why it does not make sense to switch from the first stage to the second is pretty obvious: A decision maker would simply not have any advantages from the change caused by the switching between the stages and looking again at (3.7) we find that in this case indeed $X \geq 0$.

If, however, the quitting rate gets smaller in the second stage for whatever reason, it is possible that this decrease, if it is large enough, makes the second stage again so attractive that (3.9) can be fulfilled, meaning the lower initiation and profits can get compensated by a higher number of users caused by this lower quitting rate. Then it might be possible to optimally switch from one stage to the other.

(-) Vice versa, initiation and profits increase in the second stage (e.g. when competitors leave the market). Since now the fraction on the left hand side of (3.9) is again always positive, it is only possible that the right hand side is positive if the quitting rate of the second stage is high enough with $\chi_\mu < 1$. If that was not the case, i.e. there is no advantage of ever being in the first stage, (3.7) reveals that in this case $X < 0$ and it would be always better to immediately switch to the second stage.

(-) Initiation decreases and profits increase in the second stage (e.g. due to network effects, different pricing strategy, etc.), or vice versa. Then the fraction on the left side is always negative, meaning that unless the quitting rate of users becomes so much better (i.e. smaller) that it would be always optimal to be in stage 2, there might be some point at which switching from one stage to the other is optimal: While one stage provides advantages by a higher initiation, the other one stands out due to a higher price. Which one is effectively better depends on the value of the states, costates and optimal control as well as on the parameters describing profits, initiation and customer loyalty and how they change as we will see later.

3.2 Case 1: Patent Expiration: Dealing with an Exogenously Given Switching Point

In the following section it will be seen how an exogenously given switching point can affect the optimal solution.

3.2.1 Model Assumptions

In the first stage the decision maker is the only provider of a certain good, in the second stage he cannot maintain his monopoly and has to switch to perfect competition.

It is assumed that a decision maker does not have any advantage from giving up his status as monopolist, i.e. the admittance of competition leads to a lower market price, meaning the former monopolist would have to cut his profits to remain fit for competition. Another reason for lower profits is, that since the good is non-durable and the consumer has to frequently spent money to be able to use/consume it, this user of the good has an alternative product which he might purchase every now and then, leading to smaller numbers of sales per customer and therefore lowering the average profit per user per time. Competition might also have negative effects on initiation, since some of the susceptibles might find the other product more attractive and would start to use this instead. Although a lower price would make the product more attractive, this additional attractiveness is assumed not to be able to exceed the lower motivation to buy such a product because of the availability of alternatives. There might also be more consumers who completely stop using the former monopolist's product and only use the other good, meaning that competition leads to a lower consumer loyalty and therefore a higher quitting rate of users. It is assumed that users of the competitors product are still potential consumers of the former monopolist's product and, as already stated, that it is possible to use both products more or less at the same time.

Although without any advantages, the monopolist might be forced to enter competition due to expiration of a patent at a switching time which is exogenously given. Of course he also has the option of giving up the patent earlier if this appears to be advantageous.

The parameters change in the following way:

$$\boxed{\pi_1 > \pi_2,\ \mu_1 \leq \mu_2,\ a_1 \geq a_2,\ b_1 \geq b_2 \text{ with } f_i(A) = a_i + b_i A^{\alpha},\ i = 1, 2}$$

Due to the described change in parameters (the price and initiation become smaller and the outflow from users becomes larger) the value of the Hamiltonian in the second stage is always smaller than in the first. Not surprisingly, this means that a decision maker would never, if having the choice, give up his monopoly and enter competition as he would only have disadvantages by the competitors, who would take away parts of the clientel and lower the price due to the higher supply.

But what would be his strategy if he is forced to give up monopoly? If there are initially only a few users it is optimal to give price reductions as

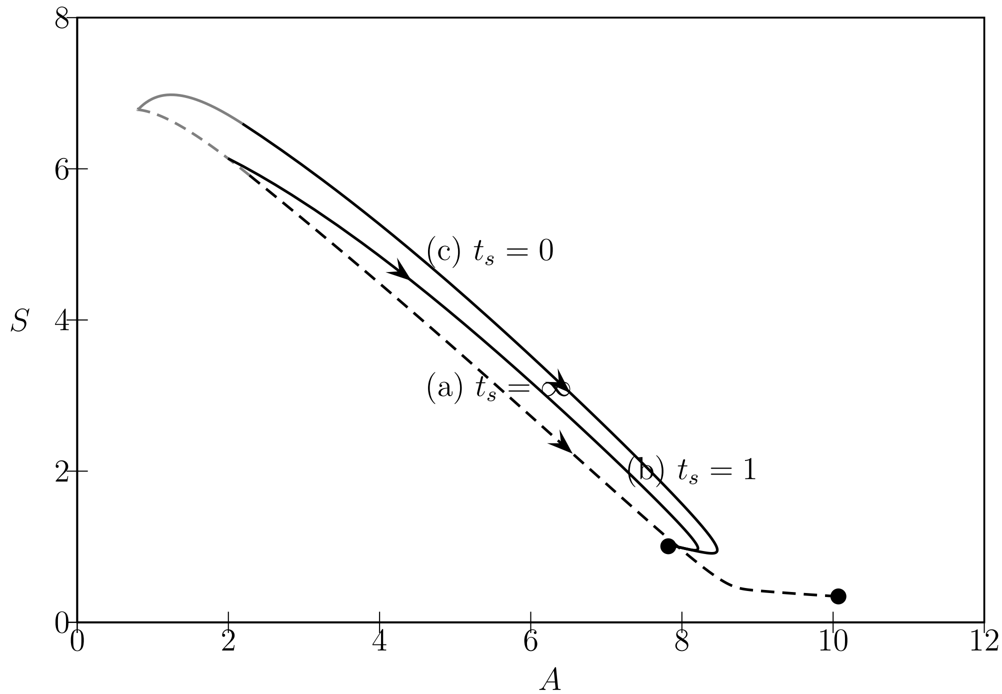

Figure 3.2: Phase portrait of case 1 of the price reduction multi-stage model. The black lines depict trajectories - the dotted parts show the part of the optimal solution lying in stage 1, he solid parts are in stage 2. Trajectory (a) remains always in stage 1, (c) switches immediately to stage 2 and (b) switches after a certain time period ($t_s = 1$). The gray parts of the trajectories show that on this part of it, the control constraint is active and $v^* = v_{\max} = 1$.

large as possible in stage one, so that when competition takes place, he has due to imitation effects the advantage of offering the more established and therefore more attractive product to the customers. However, if competition would not take place the decision maker would take a similar approach in order to build a customer base and then take advantage of it. Of course, the optimal solution would differ a bit regarding the magnitude of the states and the control. It can be seen that (in the patent expiration case with only one steady state) the number of customers is, due to the smaller initiation and the higher quitting rate, smaller in the steady state compared to the steady state if one would always remain monopolist, yet one would try to compensate this higher outflow by giving slightly higher price discounts.

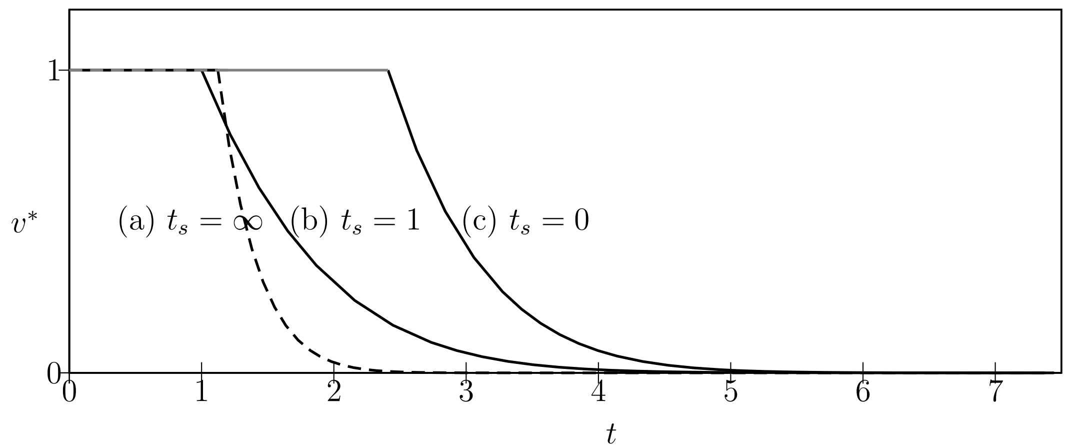

Figure 3.3: Time path of the optimal control corresponding to the trajectories of Fig. 3.2.

3.2.2 Comparison of Monopoly and Competition

One long-run steady state

Now assume that initially there are many susceptibles and only a few users. The optimal strategy of the decision maker is always to try to gain these susceptibles as users and therefore he would initially give price reductions as large as possible as can be seen in Fig. 3.3. As the number of users becomes larger, so does their influence on potential consumers and the decision maker can exploit this and can reduce his price reducing efforts. The number of susceptibles then decreases and the number of users increases until either the steady state is reached. Another possibiliy is that the pool of potential consumers is so much exhausted that due to the smaller initiation (since there are not enough susceptibles who could become users and the constant inflow is not enough to supply as much susceptibles as necessary to allow further growth of the users) and the relatively large outflow of users (due to the percentage outflow rate and the large number of users) the number of users starts to decrease and the susceptibles can grow again until their number gets to its steady state value.

However, if having the choice, the question would arise whether it is better to remain monopolist forever, to remain monopolist until one is forced to enter competition at an exogenously given time t_s or to enter competition as soon as possible.

Obviously the best option would always be to remain monopolist and to deny market entrance to any possible competitor. Then, profits would not suffer under a lower market price and consumers would not have the choice of

Table 3.2: Parameter values of the two stages of the price reduction model in the patent expiration case with one steady state

	r	π_i	k	δ	μ_i	a_i	b_i	β	ω
$i = 1$	0.04	2	1	0.05	0.0976	0.02	0.05	1	0.75
$i = 2$	0.04	1	1	0.05	0.12	0.01	0.025	1	0.75

Table 3.3: Steady state values if (a) one would never leave stage 1 (i.e. $t_s = \infty$ and (b) if one would change at some point to the second stage.

	$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
Case (a)	1.01	7.91	12.87	14.09	0	0	0
Case (b)	1.01	7.91	9.65	10.57	0	0	0

using another company's product, leading consequentially to a higher number of users in the steady state due to the higher initiation. However, this might not always be possible, e.g., when a patent expires and/or barriers to market entry get lower for some other reasons.

Fig. 3.2 shows three different trajectories, depicting the case (a) where one would always be monopolist, i.e. $t_s = \infty$, (b) where one would be monopolist until one is forced at $t_s = 1$ to enter competition and (c) where one would immediately switch to competition ($t_s = 0$). Since on trajectory (a) one would never enter the second stage, it leads to a different equilibrium than the other two trajectories. Comparing these two steady states (see Tab. 3.3) we see that if one would never switch between the stages one would end in a steady state with a higher number of users and less susceptibles. In both cases not much control spending is required in the steady state.

Comparing the values of the objective functions (which are in the calculated case (a) 182.26, (b) 175.11 and (c) 170.26), it can be seen that of course it would be optimal to remain monopolist forever, but if having the choice between exploiting the better parameters of stage 1 for time $t_s = 1$ and immediately switching to the second stage, it better to remain monopolist as long as possible.

Comparing the three different optimal control strategies by looking at Fig.

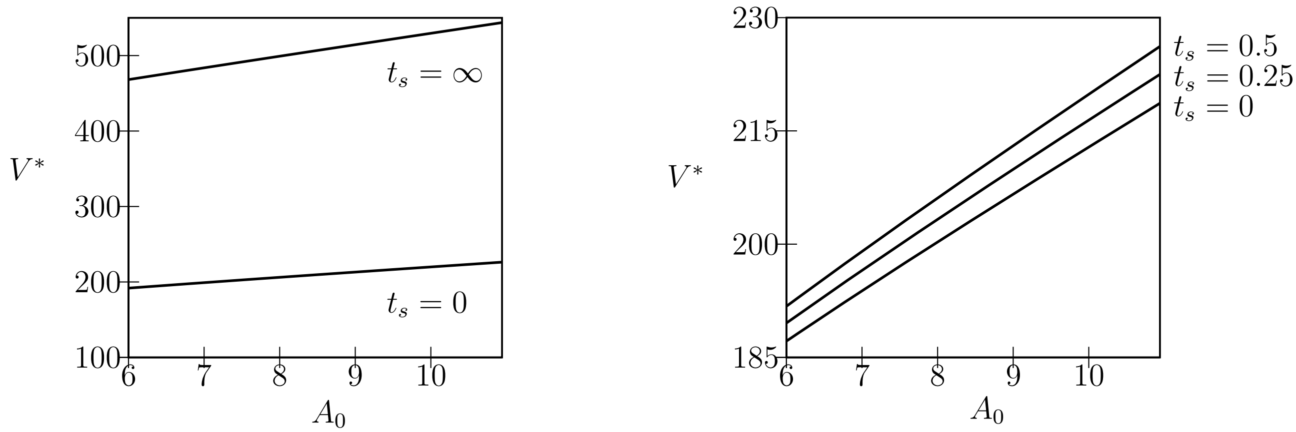

Figure 3.4: Value of the objective function depending on the initial number of users for different exogenously given switching times.

(3.3), we see that the highest control efforts are required in the case where the company always has to face competition. Interestingly, the decision maker can stop giving the maximum price discounts sooner in the case where one switches from the first to the second stage at $t_s = 1$ compared to the case where one is always monopolist. The reason for this might be that in the second stage application of the control is less effective than in the first stage due to the worse initiation parameters, so one would be able to start earlier to exploit the impact of the existing users. One would, however, decrease the amount of the control slower in this case than in the monopolist case, since it would be still necessary to compensate this lower initiation flow and the higher quitting rate of users despite the decreased efficiency.

Fig. 3.4 shows how the value of the objective function depends on the initial number of users (S_0 is assumed to be fixed at 2) and on the switching time between the stages. It can be seen that it would be the best solution by far if the company could remain monopolist forever under the given parameters. Also not surprisingly, it is better if there are initially many users, which not only means higher revenues caused by higher sales but also that there are more people who can exert some influence on the potential consumers and less price reductions are required.

Comparing now the case without competition to the ones where the company cannot remain monopolist forever, it can be seen in Fig. 3.4, that one is alway better off the longer the time is that one can stay without any competitor. Interestingly, the smaller t_s gets, the less steep is the line describing the objective value of different initial user numbers. Of course, also for the previously given reasons, it is always better to start with many users even in an always-competition case, but the less steep line reflects the worse parameters / price / profits in the competition situation, meaning due to the lower initiation less people can become consumers and they do not lead to an equal profit as in the monopoly case.

Remark on the Numerical Calculations *In order to calculate the value of objective function for different numbers of initial consumers and a fixed number of susceptibles one has to find the trajectories fulfilling the conditions of the following boundary value problem: (1) The number of susceptibles (and to a certain degree the number of users) has to have a certain value, (2) starting at the point one has to remain for the time period t_s in the first stage after which one has to switch to the second one and (3) the optimal solution consisting of parts in each stage has to end in the equilibrium in the second stage. Having found these points one can calculate the value of the objective function (see Appendix A).*

DNSS curve

In Chap. 2 we saw that the system can behave differently for different parameters and it is possible that more than one steady state becomes a candidate for an optimal solution. This can also happen in the multistage case.

Let us now assume again that the fraction of innovators among the people who decide to adopt a certain product is zero[2] in both stages (i.e. $a_i = 0$ for $i = 1, 2$). This means that no one would ever buy such a product if there is no other person who already owns it, indicating either that the product is worthless if there are no other users (e.g., a network good such as a telephone) or that it only makes sense to own such a good if it is trendy enough (e.g., fashion good) or if people rely very much on previous experience when it comes to product adoption (e.g. when it is uncertain that the new product is advantageous for them.)

In the one-stage version of this model in Chap. 2 it was shown that depending on the initial number of susceptibles and users the optimal solution would either end in a steady state with many or with no users. The question then arises again what impact does the entrance of a competitor have, if competition means a lower initiation, a lower market price and a higher quitting rate of users.

Tab. 3.4 shows a set of parameters for which a DNSS curve can occur in the multistage case. The corresponding equilibria can be found in Tab. 3.5.

[2] The following behavior of the system can also occur if this parameter is very small.

Table 3.4: Parameter values of the two stages of the price reduction model in the patent expiration case - one steady state

	r	π_i	k	δ	μ_i	a_i	b_i	β	ω
$i=1$	0.04	2	1	0.05	0.0976	0.00	0.05	1	0.75
$i=2$	0.04	1	1	0.05	0.12	0.00	0.025	1	0.75

Table 3.5: Steady state values if (a) one would never leave stage 1 (i.e. $t_s=\infty$ and (b) if one would change at some point to the second stage in the DNSS case.

	$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
Case (a)	20	0	0	14.09	14.53	0	0
Case (a)	0.35	10.07	14.65	15.11	$0.83 * 1^{-7}$	0	0
Case (b)	20	0	0	6.25	0	0	0
Case (b)	1.01	7.91	6.44	7.07	$0.98 * 1^{-5}$	0	0

It can be seen that no matter whether one would remain in the first stage or switch at some time to the second stage, a steady state with no users a certain number of susceptibles, which does not vary for the different stages might be reached. The reason for this is that the number of susceptibles depends only on the parameters k and δ (describing the inflow and quitting rate of the susceptibles) which are assumed not to be changed by the market entrance of the competitor. However, the value of the costate corresponding to the number of users changes, showing that an additional user would be of higher value in the first stage, i.e. when he causes more profits, influences more people and remains a loyal customer for a longer period of time. Since there is no chance of ever winning any users, it does not make sense to do anything. Of course, there are due to the more favorable parameters again more users in the steady state where one always remains in the first stage.

Fig. 3.5 shows that the value of the objective function depends on the number of initial users as well as on the switching time. The number of initial susceptibles is assumed to be fixed at $S_0 = 2$. It can be seen that if the initial number of users is very small, their influence is not large enough to convince

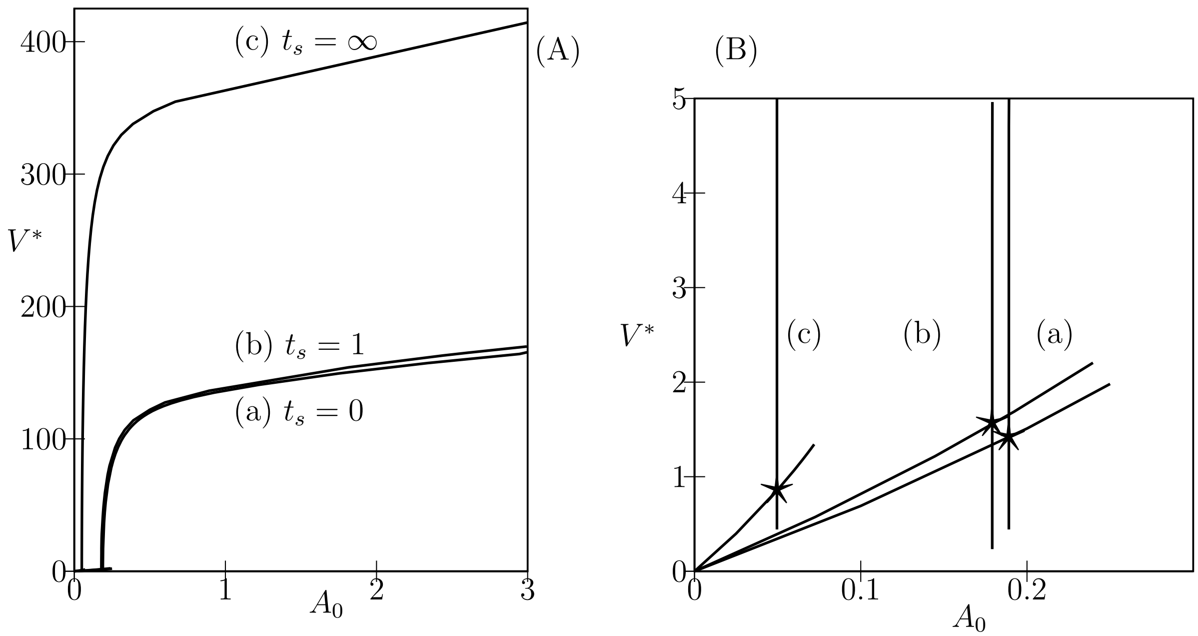

Figure 3.5: Value of the objective function for different initial values of users and different switching times (a) $t_s = 0$, (b) $t_s = 1$ and (c) $t_s = \infty$. Panel (B) is a magnification of panel (A) showing the DNSS points, denoted by a $\star$.

many potential consumers to imitate them in their purchase decision and one will always end in a steady state with no users. At a certain point the decision maker suddenly has the decision whether he wants to let the product fail or finally end near a steady state with many users. Usually one would assume, that if having the choice, between no or many users in the end, one would always prefer the solution with many paying consumers, however, as shown in Chapt. 2 starting at the DNSS point, the time needed to gain a large consumer base is too long and the control efforts too high to be really attractive for the decision maker.

In Fig. 3.5 it can be seen that the number of users in a DNSS point depends on the switching time: If one would never enter the second stage, the number of users required to make a product successful does not have to be as large as in the other cases, the number of initial users on the DNSS point is the smallest. This low number of existing customers leads to a smaller profit compared to the other DNSS points, however, the profits for this particular initial number of users is still larger than in any other case. If one is forced to enter the second stage after a certain period (here assumed to be $t_s = 1$) the profits of the company suffer, but, also due to the lower initiation of the second stage, the DNSS curve moves to the right, i.e. more initial users are required to end in the steady state with many users. As before the worst scenario is if having to switch immediately to competition. In this case the value of the objective function at the DNSS points is worse than in the previous case (DNSS point for $t_s = 1$), even though there are more initial users. The additional profits caused by the larger number of consumers is not as large as the higher gains caused by the higher market price of the first stage in the case of an exogenously given switching point.

Remark on the Numerical Calculations *Similarly to before, one has to find initial points where one of the states is fixed at a certain value, furthermore they have to lie on the stable manifold of a steady state and the emerging trajectories have to be for certain periods of time in either stage. Now, depending on the switching time, one has to consider four different steady states, which are candidates for optimal solutions. It now has to be considered, when doing the numerical calculations, that it is not always possible to find points for every initial states leading to every steady states due to the shapes of the stable manifolds.*

3.3 Case 2: An Aggressively Acting Monopolist: When Switching Points Can Be Optimally Determined

The question arises what happens if monopoly is not protected by a patent and barriers to market entry are still low? While the previous section suggested that in such a case the monopolist would always have to deal with the bad parameters of competition, what happens if the decision maker wants to keep out any possible competitor even if this is associated with additional costs in order to win more customers.

As such let us now consider a case in which the average profit per user increases and initiation decreases in the second stage. There might be several possible reasons for such a change:

(I) Etro (2007) reports that a monopolist sometimes might have a greater incentive to act more competitive in order to maintain its monopoly compared to markets with no obvious dominant market leader, particularly when the barriers to market entry are low. If they would be high, the monopolist would less likely be challenged by a competitor and, thus could act less actively.

There are different ways by which a monopolist might keep competitors out of the market, one would be by providing a product so attractive that no one else could compete. An example here would be that this is achieved by innovation regarding product improvements or further adaptions of the product to the customers' needs, or by offering a paricularly high-quality product- related and customer-friendly service. When a company spends much for such measures, it becomes harder for possible competitors to threaten the monopolist, particularly if the taken measure, e.g., frequent product improvements by innovation, is a key aspect in the product adoption decision and competitors could not offer a similar price when trying to achieve a sufficient product development in order to be taken seriously by the consumers as an alternative provider of the good. However, large spendings for innovation would lead to higher costs and therefore to smaller profits per user and a company might have the incentive at some point to give up this strategy due to these costs. Then, while profits might increase in a second stage, initiation might suffer due to alternative products of competitors, who now might risk to enter the market, as well as due to lower product attractiveness because of the slower product advancements.

(II) An similar approach, which might be a bit more problematic (because of possible conflicts with anti-trust laws) is that a monopolist could try keeping prices low in order to either signal possible competitors that it would not be profitable to enter the market or to make it impossible for anyone to compete. However, according to Motta (2004) it is not always clear whether such low prices are simply competitive and therefore advantageous for the consumer or exclusionary behavior. The problem for authorities with such a strategy is that although consumers might have a short-time advantage, they would be worse off in the long run, suffering from a higher price due to the lack of competition.

In our model it is assumed that in first stage the monopolist keeps the price very low to keep out any competitor and exploit the fact that he is the only provider of the good, leading to a high initiation and customer loyalty. In a second stage he does not care about this anymore and charge a higher normal price leading to higher profits per user. This would negatively influence initiation and the quitting rate of users not only because of possible alternative products for the customer, but a higher price would also reduce the attractiveness of the product (and might make the market even more attractive to other competitors).

(III) As previously described a price reduction might not always be seen as something positive nor does a price increase be always received as something negative. Leaving now aside monopoly and competition, one could use another economic interpretation of a two-stage model with such a parameter change. Assume a company wants to basically redefine its image and, e.g., tries to appeal only to a wealthier clientel or faces the economic need to increase the profits / price. While the company before only had the possibility to reduce temporarily the basic price allowing price discounts, it might also be interested in seeing what happens if the price and therefore the average profit per user is (exogenously) increased.

It is assumed that a permanent change of the basic price has a different effect on the dynamics of the (potential) consumers than this frequent adaption of the price by giving discounts. As before, the consumers see the basic price as some kind of reference price and price reductions only as promotional incentives to make the product more attractive. However, a second reference price gets a larger role, namely the price a product is supposed to cost, based on personal reasoning. Because of this, a smaller fraction of the susceptibles are able or willing to pay the suggested basic price if no price reductions are given meaning that initiation would fall and even if the company would give temporary price reductions to reach the previous price level that would not necessarily mean that the number of people starting to use a product would be as big as before. An increase in the quitting rate of users makes also sense in such a case since the users of the good are regularly contributing to the profits of the company, i.e. they buy it frequently or have to make use of some service (e.g. maintenance) and might not be willing or able to afford doing this as long as before anymore.

Therefore, we differentiate now between two stages: In the first the price is low, but initiation and customer loyalty (reflected by the quitting rate of users) is high; in the second stage the price is higher, but initiation and customer loyalty is lower.

Then the parameters change in the following way

$$\boxed{\pi_1 < \pi_2,\ \mu_1 \leq \mu_2,\ a_1 \geq a_2,\ b_1 \geq b_2 \text{ with } f_i(A) = a_i + b_i A^\alpha,\ i = 1, 2}$$

Looking at the control variable at the switching points we can see that in this case the optimal control will always jump to a lower value in the second stage: A decision maker would give larger discounts in the stage with the lower price and higher initiation, on the one hand because giving price

Table 3.6: Parameter values of the two stages of the price reduction model in the aggressively acting monopolist case

	r	π_i	k	δ	μ_i	a	b_i	β	ω
$i=1$	0.04	1	1	0.05	0.0976	0.02	0.025	1	0.75
$i=2$	0.04	2	1	0.05	0.12	0.01	0.025	1	0.75

reductions is more effective in this stage since more people can be convinced to start using the product. On the other hand a decision maker simply might want to exploit the higher profits of the second stage.

Fig. 3.6 shows a phase portrait of such a case with the parameter values shown in Tab. 3.6. The black lines with the arrows show two trajectories. If these line is dotted that means it is optimal at this part of the optimal solution to be in stage 1 with the lower price but higher initiation, if it is solid the higher price but lower initiation, i.e. stage 2, is preferable. The gray line shows where it is optimal to switch between the stages - being on the left side of this curve, i.e. there are not many users of the product, would mean that a company should accept a lower basic price in order to build up a consumer base. This can serve, when crossing the gray line and switching to the other stage, as example for potential consumers, compensating the lower initiation. If the initial value of the states would be on the right side of the curve one would immediately switch to the second stage. The lower of the two drawn trajectories represents some kind of threshold, its initial value is chosen in a way that one would switch at the beginning to the second stage. If, however, there were only very, very little less users in the beginning, the strategy of the decision maker would be to remain in the first stage for a while, give price reductions and, when the number of users and susceptibles is large enough, switch to the other stage.

The interpretation in the monopoly-competition case would be as follows: Unlike before the switching time from monopoly to perfect competition can be determined optimally: If the number of users is small it is better to accept the smaller profits per user, but exploit the position as monopolist and keep the competitors out - since there is no competition, all of the susceptibles who decide to purchase such a product add to the consumer stock of the company. Then if there are enough users and the market position is strong enough, it becomes optimal to switch to competition and exploit the higher price together with the good position in the market, which compensates the

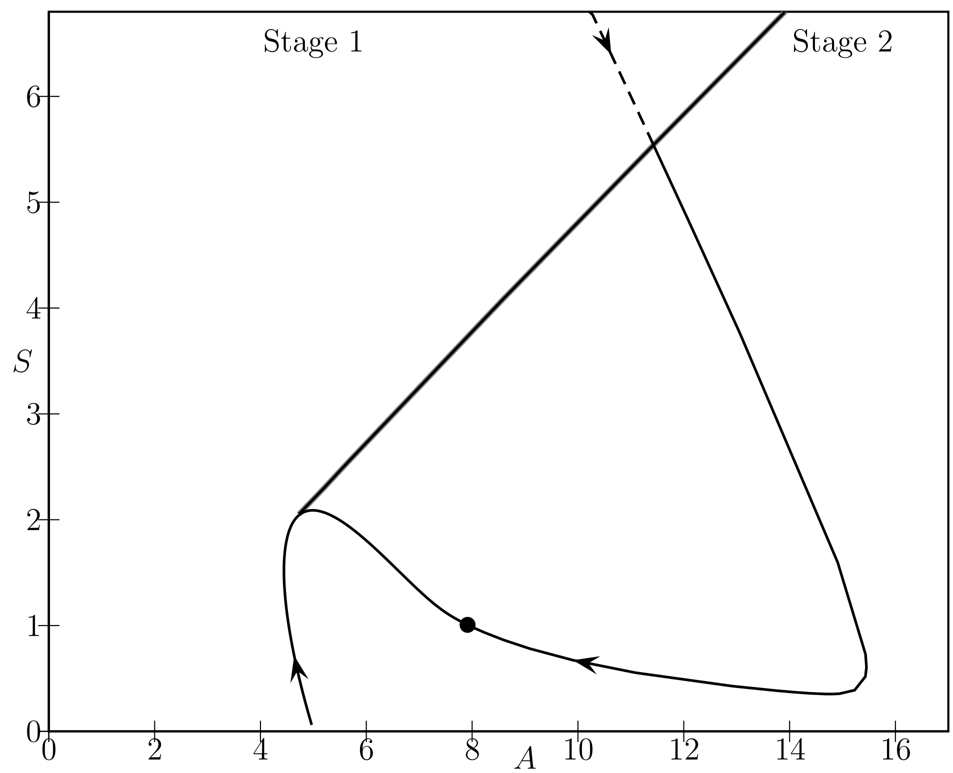

Figure 3.6: Phase portrait of case 2 of the price reduction multi-stage model. The black lines depict trajectories - the dotted parts mean that at that points it is optimal to remain in stage 1, on the solid parts it is optimal to be in stage 2. The gray line shows where it is optimal to switch between the stages.

lower initiation. The optimal strategy would again be to give large price reductions if the number of users is low, but only very small or none at all if their number is large. It is optimal to give larger reductions in the first stage, because it helps to win consumers - then in the second stage the company can exploit the profits and give less profits, the optimal value of the control in the switching point will jump down.

Such a strategy might be certainly advantageous for the decision maker, exploiting the higher initiation when the customer base is low and the higher profits if there are many users of the product. At at some point $t_s \geq 0$ a decision maker would always enter the second stage with the higher profits. This would mean for the customers, who might have only become attached to the product due to the lower price of the first stage, that they have to pay more for the product at some point. While they might have a higher utility of the product in the second stage due to the higher number of users, they might be worse off of than in a case where the monopolist would not act as aggressively in the first stage. Then a monopolist could not keep out competition as long and prices might get lower than they are in the second stage at some point, because a monopolist might not be able to gain such a strong market position that he could charge this higher price.

Similarly, when considering the consequences of an increase of the basic price of a monopolist, would only adapt the basic price, i.e. switch to the second stage, if the market position is strong enough that sufficient people would accept the price increase and the influence of these people is strong enough to motivate others to start adopting the product.

Remark on the Numerical Calculations *One can find the optimal switching points in this particular case by solving following problem: Find a point on the stable manifold of a second stage equilibrium, where the value of the Hamiltonian of the second stage is equal to the value of the Hamiltonian of the first stage. To find an optimal solution starting at a certain point, one has to include the following degenerated cases, where the switching time is either zero or infinity: If the value of the Hamiltonian function of the second stage is, for all points on the trajectory emerging from this starting point, always greater than the corresponding value of the first stage, the trajectory must always follow the dynamics and parameters of the second stage. If this is not the case, the second part of the optimal trajectory has to be in the second stage starting at an optimal switching point, any part of this solution with a lower time value has to be located in the first stage.*

The Influence of the Parameters on the Optimal Switching Points

By determining the curve consisting of the optimal switching points, one can find information when and how a decision maker has to change into the other stage. However, the question arises what happens if the parameters change in a different way between the stages. Let us assume that in a second case the average profits per user reflecting the market price changes less than before ($\pi_2 = 1.5$ instead of $\pi_2 = 2$).

Table 3.7 shows how the steady state changes if parameter π_2 is changed: While it does not alter the values of the states, it has an effect on the costates, which become smaller if the price is lower, meaning that an additional unit of either a susceptible or user is less worth if the profits made through them are lower.

Which role such a different parameter plays for the optimal switching points, can be seen in Fig. 3.7. The upper gray line shows the previous

Table 3.7: Steady state values for different average profit per users in stage 2

	$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
Case $\pi_2 = 2$	1.01	7.91	12.87	14.09	0	0	0
Case $\pi_2 = 1.5$	1.01	7.91	9.65	10.57	0	0	0

switching curve while the lower one shows the switching curve for the new, smaller average profits per user. The black line show the last trajectory where one would always prefer the second stage - starting with a little less users one would remain for a while in the first stage until the number of susceptibles and users is large enough that it becomes optimal to exploit the higher basic price and accept the lower initiation.

Comparing the two switching curves, it can be seen that, if the price increase is not so big in the second stage, the switching curve moves to the right and becomes less steep, meaning that at the optimal switching curve either have to be

(a) more users and/or

(b) less susceptibles.

The reason for (a) is that because of the lower profits the incentive to accept the lower initiation for the higher profit is not as large as before. In a way the decision maker would compensate the lower profit in this parameter case by trying to find a larger pool of customers before changing to the other stage. Closely related to this is the reason why there are less susceptibles: Switching at a later point in time, the larger number of users has to come from somewhere, and the only possible source are the susceptibles, who become smaller by this later change.

For similar reasons the last trajectory where one would always remain in the second stage shifts to the right: The incentive to be in the first stage and exploit the better parameters is now greater than before and only if the number of users and susceptibles is large enough, one should start exploiting the higher price.

Other changes of the parameter values also have some effect on the switching points: If a parameter is more advantageous to the decision maker in one stage (and in the other one remains the same or does not change so much into this favorable direction), it is optimal to remain longer in this stage and the switching curve shifts.

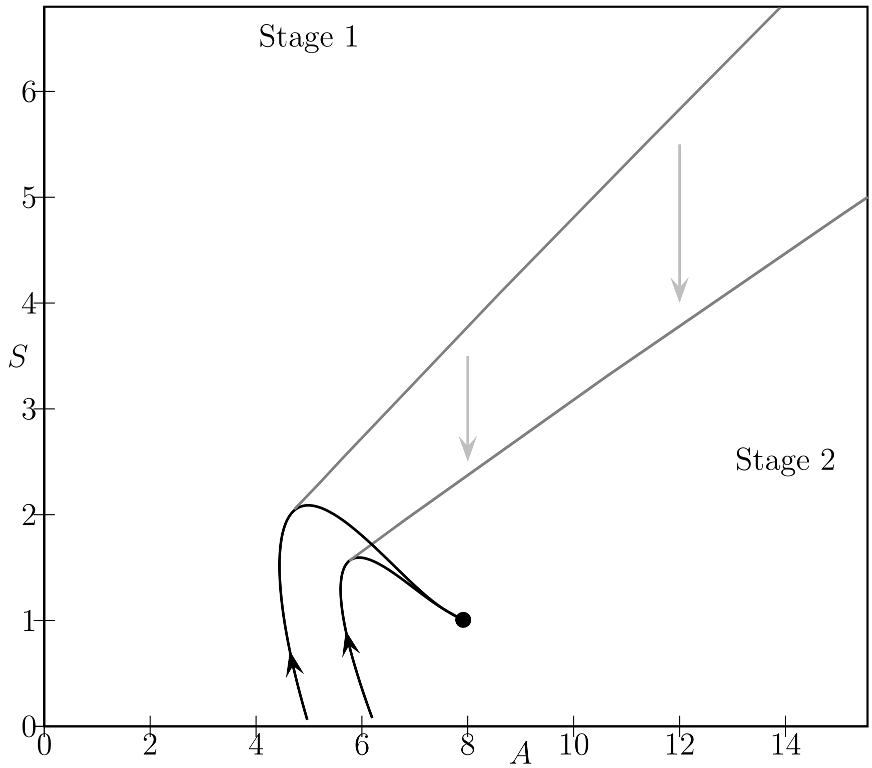

Figure 3.7: Phase portrait comparing the optimal switching points between the two stages for different average profits per user.

Concluding it can be said that a decision maker has to adapt his optimal strategy according to how the parameters change as well as the values of the state and costate variables at the switching point, both in a case where it is exogeneously given as well as when the switching point can be optimally determined.

Let us now move to another field, where one can use diffusion models to gain more knowledge about the development of the number of potential and actual consumers of a certain good, namely drug policy. Unlike before, a minimization problem will be considered where a decision maker seeks to keep total harm arising from drug usage as small as possible.

Part II

Harm Reduction

The harm caused by the usage of illicit drugs is not only present to people directly affected, but impacts society itself on a social and economic level. Costs, affecting not only the drug users, arise for instance by necessary additional health care, through the loss of productivity, by the transmission of diseases like HIV or hepatitis, by additionally required safety measures, etc. (see ONDP, 2004; MacCoun et al., 1996). This does not mean the damage arising to the users themselves is neglectible: It is them firsthand who get infected or who might loose their job and/or family, it is them who face legal prosecution and it is them who might get into acquaintance with criminal networks (cf. MacCoun et al., 1996).
Since the number of people using various drugs is not neglectible in many countries (cp. UNODC, 2006) it must be a policy maker's objective to reduce the harm caused by them. The classical approach, in terms of drug policy goals, towards this problem is to say that it is best to reduce the number of users, so that no harm can arise through them. (cf. Caulkins & Reuter, 1997, MacCoun, 1998, but also ONDP, 2006). Typical drug policies are prevention, law enforcement and treatment, how to apply these (optimally) has already been the topic of many works (e.g. Behrens et al., 2000, Winkler et al., 2004, Everingham & Rydell, 1994, to name just a few).
Yet it is impossible to completely eliminate the whole drug problem (Caulkins & Reuter, 1997) and even worse: these policy excaberate also some drug related harms , for example concerning the involvement in illegal activities, e.g. higher law enforcement leads to more arrests and punishments, and increases the price of the drugs, which in consequence might lead in some cases to a higher rate of drug related crimes. Therefore it might make sense to explicitly consider another desirable drug strategy goal: to reduce the harm caused by drug usage itself. This can be done in many ways: There are for example needle exchange programs, trying to decrease the infection rate of HIV or hepatitis among drug users, legalization programs for certain drugs in order to avoid to make users of these drugs criminals and even to decriminalize them, to authorize certain vendors to sell the drug, in order to be able to control the quality and the price of the drugs. (cf. MacCoun et al., 1996; Lenton & Single, 1998; MacCoun, 1998).
Although creating such programs certainly makes sense, the harm reduction approach to the drug problem is far more controversial than other drug policies, ranging from people claiming that drug users should have to face the consequences of their behavior to the argument that the "wrong message is sent" (MacCoun, 1998). It is definitely not the goal of harm reduction to justify or even encourage drug use, but to lessen the harm arising from it, which is, despite some prejudices against it, not only for the good of the users, but also for society. Yet it is plausible harm reduction might indeed

make people believe to a certain extend, that it is safer to take drugs (which is of course in the interest of drug users) and therefore are more easily willing to do so. (cf. MacCoun, 1998; Caulkins, 2005).
The harm reduction model investigated in this work is based on a model introduced in Caulkins (2005). Due to the dynamic character of drug problems (Caulkins, 2001; Reuter, 2001) it is the aim of the model to see how the optimal use of harm reduction varies over time and how the number of users and susceptibles change.
As mentioned two groups of people will be considered in this model: of course the drug users, but also people susceptible to drug use. The reason for this is that not all people are equally susceptible to the use of certain drugs, because of their age, their religion, their social environment, etc. Similar to the number of users, the number of potential users is subject to certain dynamics. The inclusion of susceptibles can lead to a better understanding of why the number of drug users evolve the way it does (cf. Caulkins, 2004). Similar inclusions of susceptibles have already been done in various models, e.g. Rossi (2001), Caulkins (2004) or Wallner (2008), yet this approach is rather new.
Another issue, that will be considered in the analysed model here, is the question how the susceptibles are influenced by the users in their decision on whether to take drugs. Such social interactions have been addressed e.g. in Glaeser et al. (2002), Kleiman (1993) and Himmelstein (1983).
In the next section an optimal control model considering harm reduction will be described and analysed and different parameters will be considered to gain knowledge about the general system behavior. The model was first introduced by Caulkins (2005) extending Caulkins (2004), has been analyzed in a one-state version in Wallner (2005) and numerically validated with a slightly different objective function and function describing the impact of the control on the initiation in Wallner (2008).

Chapter 4

Harm Reduction Models

Another application for which the underlying dynamical system is useful lies in the field of drug policy. Existing users of a drug often provide incentives for non-users to start taking a certain drug, e.g., by serving as bad example or by exercising a certain pressure, may it be done consciously or not. In such a case it also makes sense to distinguish between potential and actual consumers of a certain drug, since not all members of a society might be equally attracted to drug usage, maybe because of age, education, religion, etc. The social interaction between these two groups, leading to the diffusion of a certain drug or not, is a crucial reason for why people start taking drugs. Like in the marketing model, it makes sense to distinguish between innovators and imitators. Again it is possible for a decision maker to influence the development of these two states by the application of some control instrument, which is harm reduction. Yet, unlike the marketing model a decision maker does not seek to optimize the profit arising for a certain company, but to find the best possible way basically for society to influence the existing drug situation.

As previously mentioned the control instrument is now harm reduction, which is applied to lessen the arising harm caused by people who consume a certain drug. While applying such a control measure leads to a smaller total harm to society, if it safer to use a certain drug it might lead to an increased attractiveness to potential consumers and therefore an increased flow from susceptibles to users.

4.1 The Objective Functions

In the following sections, two different versions of the model will be considered, each differs in its objective function. While the first and the second case differ in the assumption whether harm reduction measures influence the existing users' drug usage.

4.1.1 Base Case

While before it has been optimal to maximize the profit of a company, it must be now the decision maker's objective to minimize the total harm arising by drug usage. Defining total harm as harm per users times the number of users, as suggested in MacCoun (1998), total harm is either rather small if the number of users is low or if the harm per user is not too large. The objective function is now given by

$$\min_{0 \leq v \leq v_{\max}} \int_0^\infty e^{-rt}(A(1-v) + cv^2)\mathrm{d}t, \tag{4.1}$$

where parameter r is again the discount rate and A the number of users. The control variable v now describes the percentage reduction of harm, as such it has to be larger than zero and smaller as a certain value $v_{\max} \leq 1$ which describes the maximum percentage of harm that can be reduced. Parameter c describes the costs of the application of the control. It is assumed that, since a decision maker would use the more cost efficient measures first, the total costs of such interventions are quadratic.

The dynamical system is given by

$$\begin{aligned} \dot{S} &= k - \delta S - f(A)Sg(v) && (4.2) \\ \dot{A} &= f(A)Sg(v) - \mu A. && (4.3) \end{aligned}$$

S denotes again the number of susceptibles, k is the number of persons who join this group of people per time unit, δ is the outflow rate of the potential consumers, μ the quitting rate of drug users.

The influence of the control on the initiation of users is given by the function $g(v)$. As previously described, if a policy maker decides to implement harm reduction measures, it means that drug users are less exposed to the accompanying dangers of illicit substance abuse. Due to this more people might be tempted to start taking drugs. It therefore might make sense to use

$$g(v) = 1 + \beta v \tag{4.4}$$

This particular form of the function was introduced in Wallner (2005) as an alteration of a function suggested in Caulkins (2005). It means that there is a flow from susceptibles to users, which is not directly influenced by the impact of the harm reducing measures, but this flow gets increased by β times v percent, due to the previously described effect, that less harm increases the attractiveness of taking drugs. Parameter β weights the impact of the control. Note that, unlike in the marketing model, this function now depends linearly on the control variable.

When considering such a model, especially when including the susceptibles, is important to address the question why do people start taking drugs? In many cases the motivation for taking drugs or any other illegal behavior can be found in social interactions (cf. Caulkins, 2004; Glaeser et al., 2002; Himmelstein, 1983; Kleiman, 1993). These motivation can be caused for example by wanting to impress or imitate certain people or word-of-mouth advertisement done by the existing drug users. Therefore if the number of users is large, many susceptibles will find themselves in contact with users and might feel a stronger pressure or at least influence regarding their decision on whether to take drugs.
It therefore makes sense to use the initiation function $f(A) = a + bA^\alpha$ as suggested in Caulkins (2004) and Caulkins (2005), with a describing the percentage of susceptibles who decides to take drugs without being influenced by anyone else (e.g. for self-medication), this fraction of people are the so-called innovators. bA^α gives the influence of the users on the susceptibles (the so-called imitators), with the parameters b and α describing the strength of the impact of the number of users on the number of susceptibles. α is assumed to be greater than one meaning that the more people take a certain drug the stronger becomes the pressure of the susceptibles to fit in and imitate this behavior.

Application of Pontryagin's Maximum Principle

The Hamiltonian function than can be written as

$$\mathcal{H} = A(1-v) + cv^2 + \lambda_1(k - \delta S - f(A)Sg(v)) + \lambda_2(f(A)Sg(v) - \mu A) \quad (4.5)$$

Since we are dealing again an optimal control problem with control constraints (see, e.g., Grass et al., 2008; Feichtinger & Hartl, 1986; Léonard & Long, 1992), we have to consider the Lagrangian function, which is given by $\mathcal{L} = \mathcal{H} + \nu_1 v + \nu_2(v_{\max} - v)$

Applying now Pontryagin's Maximum Principle, we find that

$$\mathcal{L}_v = -A + 2cv + (\lambda_2 - \lambda_1)f(A)Sg_v + \nu_1 - \nu_2 = 0. \quad (4.6)$$

Due to (4.4) we find that $g_v = \beta$, which can be inserted into (4.6) in order to express the optimal control as

$$v^* = \frac{A + \beta(\lambda_1 - \lambda_2)f(A)S - \nu_1 + \nu_2}{2c}. \tag{4.7}$$

We can see here that if the number of users is large it is optimal to do much in order to fight the harm caused by them. As we will see later, the second costate is greater than the first one, which means that an additional user is worse than an additional susceptible. Then, if initiation or the impact of the control on it is big or it is very undesirable at this point to have additional users, expressed by a large λ_2, one should not do much harm reduction measures in order to avoid this additional flow caused by control application. Also, if the costs are large, one should not do too much, because the lower harm would not justify these costs. Since dealing with a minimization problem now, the Legendre-Clebsch condition is fulfilled if

$$\mathcal{L}_{vv} = 2c + (\lambda_2 - \lambda_1)f(A)Sg_{vv} \geq 0. \tag{4.8}$$

Using (4.4), the second derivation of the function $g(v)$ with respect to the control is zero, meaning that the previously given condition is always fulfilled since $c \geq 0$. Again, only the necessary conditions for optimality are fulfilled and only extremals are found, however, they can be assumed to be optimal because of the economic interpretation of the candidates for an optimal solution.

The Lagrange-multipliers ν_1, ν_2 are due to the complementary slackness conditions such that

$$\begin{aligned} &\nu_1 v = 0 \qquad \nu_2(v_{\max} - v) = 0 \\ &\nu_1, \nu_2 \leq 0. \end{aligned}$$

The costate equations are then given as

$$\dot{\lambda}_1 = (r + \delta)\lambda_1 + (\lambda_1 - \lambda_2)f(A)g(v) \tag{4.9}$$

$$\dot{\lambda}_2 = (r + \mu)\lambda_2 - (1 - v) + (\lambda_1 - \lambda_2)f_A S g(v) \tag{4.10}$$

Like before the sign of the costates is positive, however now it denotes the maximum amount that a decision maker would be willing to pay for one unit less of these states, because unlike the marketing model, a decision maker neither wants to have people, who cause harm to society by drug usage, nor people who are attracted to this behavior. Note that the first costate equation describing the development of the shadow price of the susceptibles depends in the same way on the same parameters and functions[1] as the first

[1] Their specific shapes are of course different.

costate equation denoting the shadow price of the potential customers in the marketing model. This is because the objective function does not depend on the number of susceptibles in both cases and the state equations are basically the same (except the specific form of the used functions). In the numerical calculations of the subsequent sections, the second costate will be again larger than the first one, this time because it is obviously better to have potential than actual consumers of a certain drug. With the help of (4.9) it can be seen that in a steady state $\lambda_2 > \lambda_1$ because of the non-negativity of r, δ, $f(A)$ and $g(v)$.

Steady States

Crucial to the behavior of the system are the occurring steady states and their stability properties. They can be found by setting the state and costate equations (4.2)-(4.3) and (4.9)-(4.10) to zero. As a consequence, when starting in these points, the values of the states, costates and consequentially of the optimal control and Lagrange multipliers will never change.

By setting $\dot{S} = 0$ and $\dot{A} = 0$ we find with the help of $\dot{S} + \dot{A} = k - \delta S - \mu A$

$$\hat{S} = \frac{k - \mu \hat{A}}{\delta} \tag{4.11}$$

This holds for all steady states. We can see here that in an equilibrium with many users, there won't be many susceptibles and vice versa. If the inflow to the susceptibles k is high there will be more susceptibles in the steady state than if these parameters would be low. If many susceptibles chose to leave the system without ever becoming a user, or in other words if δ is high, or if the quitting rate μ of the users is big, the number of susceptibles will be smaller in the equilibrium that if their quitting rate was be low.

By setting (4.9) to zero we find that

$$\hat{\lambda}_2 = \hat{\lambda}_1 \frac{r + \delta + f(A)g(v)}{f(A)g(v)},$$

and since the discount rate r and the quitting rate of susceptibles δ are both larger than zero, this means that in a steady state the second costate is always larger than the first one, representing that it is worse to have an additional drug user than a potential consumer.

When using the previously specified functions $g(v) = 1 + \beta v$ and $f(A) = a + bA^{\alpha}$ and assuming that parameter a, i.e. the fraction of innovators among the susceptibles, is zero, it is possible to analytically find a steady state which can be seen in Tab. 4.1 and to find the eigenvalues of the Jacobian matrix

Table 4.1: Steady state values for $a = 0$

$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
$\frac{k}{\delta}$	0	0	$\frac{1}{r+\mu}$	0	0	0

as $\xi_1 = -\delta$, $\xi_2 = -\mu$, $\xi_3 = r + \delta$, $\xi_4 = r + \mu$. As such it is a saddle point with a two-dimensional stable manifold and a candidate for an optimal solution.

The explanation of this steady states is as follows: There is no one among the susceptibles willing to take drugs if there is no one there, who gives them any reason to imitate them in their behavior. Therefore if there are no users serving as examples, there will never be any. Since there are no users it is not necessary to reduce the harm caused by the drugs, because simply no users means no harm to society. A more detailed interpretation of this steady state will be given later.

4.1.2 An Exponential Objective Function

Previously it has been assumed that the harm reduction measures lead to a linear decrease of total harm per user until some maximum value $v_{\max}$ is reached. Yet, this may not be entirely realistic. Let us now redefine the optimal control instrument v as the percentage reduction of *reducible* harm. (Before it was assumed that the maximum amount of reducible harm is determined by $v_{\max}$.) An argument, why the efficiency of harm reduction measures might be convexly decreasing, is a negative impact on the handling of drug usage by the consumers themselves: While it was previously assumed that harm reduction measures make drug usage more attractive for non-users, some drug users might also lower their concerns regarding the damages caused by their bad habit, and they might be induced to use the drug more frequently - leading again to a higher total harm of the drug. Of course, such a behavior might not necessarily occur for all drugs and their users, because a heavily addicted user would probably care little about how safe a more frequent drug usage is even if no harm reduction is done at all. Also, if it is assumed that possible harm reducing measures are not equally efficient, a decision maker would primarily use the more efficient tools, i.e. those who do not lead to an increased usage of the product.

Although v describes the percentage of reducible harm, it might still be possible that the maximum admissible value of the control is smaller than 1, particularly if the decision maker is restricted in his choice for harm reducing instruments, by arguments of political nature. E.g., some people might not be particularly intrigued by the thought of handing out clean needles to heroin users (although it would lower the risk of HIV or hepatitis C transmissions) or by a significant reduction of legal actions against drug consumers. Still that would be possible measures that would certainly decrease the harm and costs arising for society.

The optimization problem then is given by

$$\min_{0 \leq v \leq v_{\max}} \int_0^\infty e^{-rt}(A(h + (1-h)e^{-mv}) + cv^2)dt \tag{4.12}$$

Parameter h describes the maximum possible harm left per user if the control is fully applied, as such it has to be smaller or equal to one, and parameter m serves to better weight the impact of harm reduction on the objective function, c reflects again the costs of harm reduction measures.

The state equations including the initiation function as well as the function that describes the impact of the control on the flow from the susceptibles to the users remain the same as in the base case model.

4.1.3 Comparing the Objective Functions

When analyzing an optimal control problem the question arises how does the used assumption regarding the objective functions influence the optimal solution. Therefore let us take a closer look at both of the used objective functions, which describe the total harm caused by drug usage and the costs arising by harm reduction measures. Taking a closer look at the functions we see that for $v = 0$ and $h = 0$

$$A(1-v) + cv^2 \qquad \text{and} \qquad A(h + (1-h)e^{-mv}) + cv^2$$

return the same value, i.e. A.

The Taylor approximation (of 2nd degree) at $v = 0$ of $A(h+(1-h)e^{-mv})$ is

$$A(1-(1-h)mv) + O(v^2)$$

which is very similar to $A(1-v)$ of the prior objective function, especially for the parameter values $h = 0$ and $m = 1$. If one would look at the objective functions for larger values of v this approximation would not describe the exponential function as good as before anymore. This implies that if the control spendings are small, then it would not make a big difference which objective function is used and therefore which assumptions are made regarding the impact of the control on total harm arising to society. Also if the number of existing consumers is rather small it does not make a big difference which objective function is used since the arising harm is so small then that control measures do not influence many people in their decision to start taking drugs anyway and the impact of control spendings would be rather similar. However, it must be considered, when choosing the objective function, that the number of users as well as the optimal control expenditures might change over time, and while at some points in time it might appear that both ways to describe the impact of the control on total harm are fitting for a given problem, this does not have to be the case as time goes on.

In order to be able to compare the solutions of both problems, we want to find parameter values to keep the objective functions as similar as possible. Therefore we want to find the parameters m and h for which the difference of $h+(1-h)e^{-mv}$ to $(1-v)$ is as small as possible and therefore have

$$\min_{m,h}(1-v-(h+(1-h)e^{-mv}))^2$$

which gives

$$h=0, m(v)=-\frac{\ln(1-v)}{v}$$

The mean value of $m(v)$ for $0 \leq v \leq 1$ would then be determined by

$$\int_0^1 m(v)\mathrm{d}v = 1.6449,$$

a parameter value which will be used together with $h=0$ in the subsequent sections.

Fig. 4.1 shows harm per user as assumed by the two objective functions for $A=1$ for the different admissible values of the control variable. There it can be seen that, of course, total harm is linearly decreasing for the basic objective function. The exponential objective function captures that for small application of the control instruments harm decreases stronger, which occurs when a decision maker decides to use the more efficient measures first. Fig. 4.1 also reflects that further control application becomes less and less effective because on the one hand users might increase their usage of the drug.

Application of Pontryagin's Maximum Principle

The Hamiltonian function now is given by

$$\mathcal{H} = A(h+(1-h)e^{-mv})+cv^2+\lambda_1(k-\delta S-f(A)Sg(v))+\lambda_2(f(A)Sg(v)-\mu A), \tag{4.13}$$

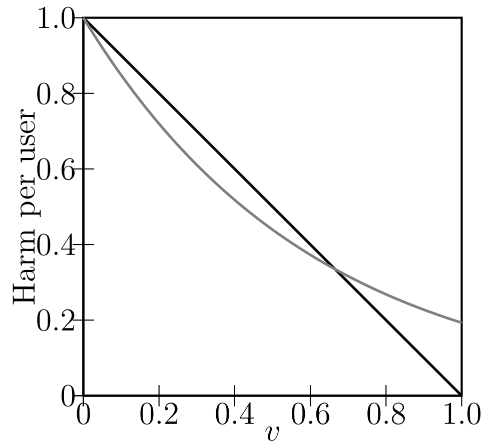

Figure 4.1: Harm per user using the base case objective function (black line) and the exponential objective function (gray line)

but due to having to include the control constraints $v \geq 0$ and $v \leq v_{\max}$ one has to consider the Lagrangian function which is given as $\mathcal{L} = \mathcal{H} + \nu_1 v + \nu_2(v_{\max} - v)$.

The necessary condition for optimality

$$\mathcal{L}_v = -A(1-h)me^{-mv} + 2cv + (\lambda_2 - \lambda_1)\beta f(A)S + \nu_1 - \nu_2 = 0 \tag{4.14}$$

has to be fulfilled, however, it is not as easy as before to express the optimal control with the help of this condition. Yet, it is possible to express one of the costates using this condition and later substituting it into the dynamical system consisting of the two state equation, the costate equation of the other adjoint variable and an equation describing the development of the optimal control over time[2].

Remark on the Numerical Calculations *The OCMat toolbox is also able to handle cases where it is not possible to express the optimal control explicitly given by the previously derived condition. In that case a costate dynamics is replaced by the control dynamics. All these transformations of the canonical system into the state-control space are done automatically by the toolbox.*

The Legendre-Clebsch condition also has to be fulfilled, which happens if

$$\mathcal{L}_{vv} = A(1-h)m^2e^{-mv} + 2c \geq 0.$$

The parameters used in this condition are such that none of elements of the expression can be smaller than zero, therefore this condition is always fulfilled.

[2]See Appendix.

Table 4.2: Parameter values for the single steady state case

r	c	k	δ	μ	a	b	β	α	h	m
0.04	8.247	1	0.05	0.0976	0.02	0.01	0.1732	1.75	0	1.6449

The costate equations now are given by

$$\dot{\lambda}_1 = (r+\delta)\lambda_1 + (\lambda_1 - \lambda_2) f(A) g(v) \quad (4.15)$$

$$\dot{\lambda}_2 = (r+\mu)\lambda_2 - (h + (1-h)e^{-mv}) + (\lambda_1 - \lambda_2) f_A S g(v). \quad (4.16)$$

Again we find that due to (4.15) in a steady state the first costate is smaller than the second, again meaning that a consumer of a drug is less desirable to have than a potential drug user. Note that the first costate equation (4.15) is the same as in the base case. The reason for this is that S does not occur in the different objective function, while the state equations where it occurs are the same. If no innovators feel attracted to the drug, i.e. parameter a is equal to zero, it is again possible to find one steady state with no users analytically, the values of the states, costates, control and Lagrange multipliers are the same as in the base case shown in Tab. 4.1.

4.2 One Long-Run Steady State

Like in the marketing for certain combinations of parameter values only one steady state can be found as admissible candidate for an optimal solution. One example for such parameters is shown in Tab. 4.2, the same parameters are used for both of the objective functions. These parameters are not chosen entirely randomly, they have been derived and applied to similar models dealing with harm reduction in the cocaine problem of the U.S. in Caulkins (2004) and Wallner (2005), however, in this combination they probably do not fit well to any drug problem.

There is only one steady state, as well for the base case as for the exponential objective function. These steady states are shown in Tab. 4.3. In both cases the number of users is rather large. Unlike in the marketing model, where one would do as good as nothing and exploit the impact of the users, one would now optimally apply the control measure in the steady state. The large influence of existing drug consumers is something undesirable. Using harm reducing measures would increase the flow from susceptibles to users,

Table 4.3: Admissible steady state values for the parameters given in Tab. 4.2 if the base case and the exponential objective functions are used

	$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
Base Case	1.59	9.43	3.24	3.75	0.5674	0	0
Exponential Case	1.62	9.42	3.60	4.17	0.4459	0	0

application of the control would decrease the total harm arising. In such a case harm reducing measures compensate the additional harm caused by their application.

Remark on the Numerical Calculations *Numerical calculations are done again with the help of the OCMat toolbox. However, in order to receive correct results, one has to consider that when control constraints become active the Lagrange multipliers have to be non-positive dealing now with a minimization problem. In order to ensure correct switching to the system where the control takes its maximum or minimum admissible value, OCMat needs positive Lagrange multipliers for calculations. This is handled automatically by specifying in the initialization file of the model that a minimization problem has to be dealt with. Another possibility is to formulate the minimization as a maximization problem, i.e. multiply the objective function with -1 before initialization of the model.*

Due to the similar parameters the steady state values are close for both objective functions, the biggest difference is the value of the second costate, which is larger for the exponential objective function. The reason for this is that λ_2 describes the maximum price a decision maker would pay for a user less. Since the assumption is made that application of the control does only increase the flow from the first to the second state, but also increases usage of existing user, the total harm of an additional user is worse for the exponential objective function. Note, however, that starting on this point, the value of the objective function would be a little bit smaller for the exponential case due to the slightly smaller number of users and the resultingly lower control application in the steady state.

Now the optimal strategy would look as follows: If the number of users is initially small, any efforts in reducing the harm would only lead to costs and a direct and indirect[3] increase of initiation, which cannot be compensated

[3]Caused by the imitators following the example of the people who are attracted to the drug due to the lower harm.

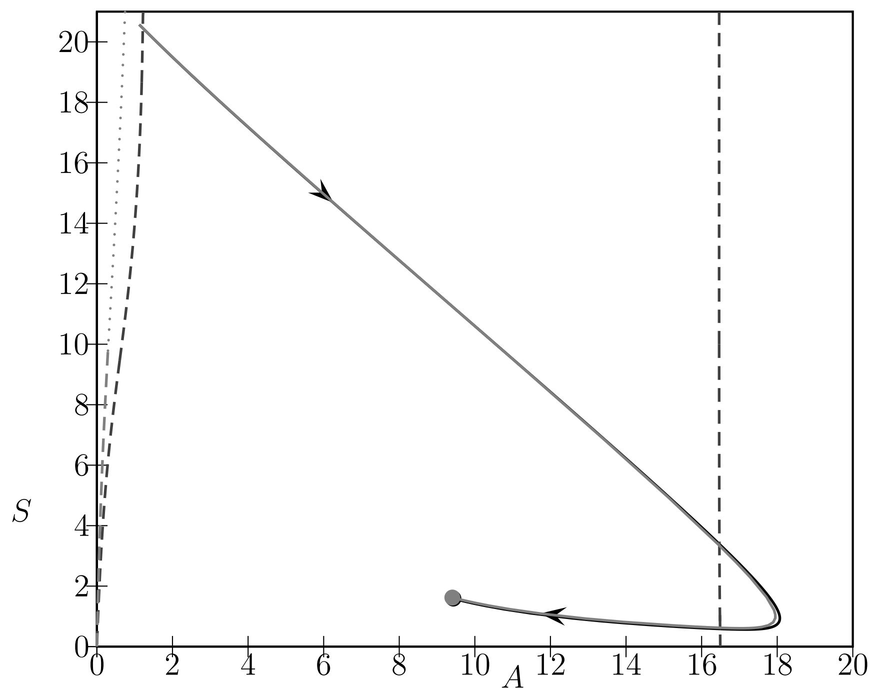

Figure 4.2: Phase portrait of the single steady state case of the harm reduction model for the base case (black line) and the exponential case (gray line) objective function. The dashed and dotted lines show where control constraints become (in)active in the base and the exponential case respectively: If the number of users is small one should not apply any harm reduction measures, however if their number is large, the maximum amount of control spendings is required. The curve showing where the control variable reaches its maximum admissible value is not shown (for lack of space).

by the reduction of the total harm in the objective function. Therefore, it would be basically optimal to ignore what happens in such a situation, the harm arising by drug users would not be worth consideration. However, due to the influence of the existing users and the fraction of innovators among the susceptibles, the number of users will rise after some time. Then, if their number and their impact is large enough, the number of susceptibles will decrease, because too many of them would decide to start using the drug.

If the number of susceptibles is initially large and the number of users not too big, a decision maker has to be careful regarding his control efforts: As the message that it is safer to use a certain drug reaches more potential users, one would keep harm reduction measures as low as possible in order to avoid a fast escalation of the problem.

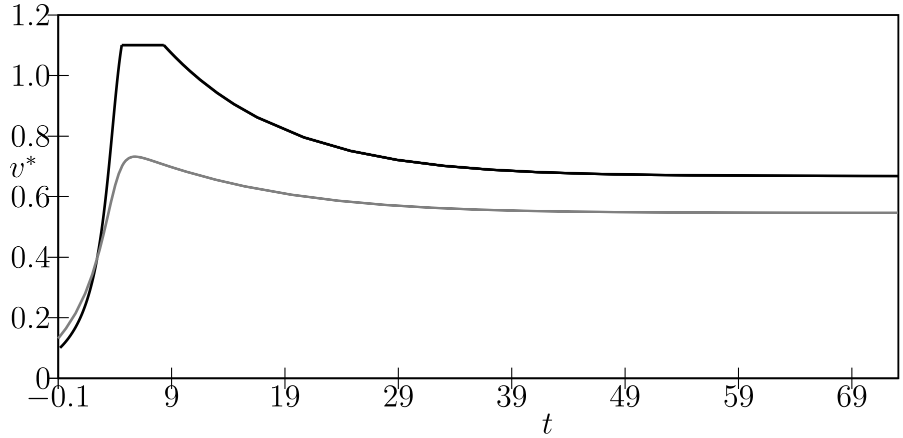

Figure 4.3: Timepath of the optimal control corresponding to the phase portrait Fig. 4.2.

Fig. 4.2 shows how an optimal solution starting at an initial point (A_0, S_0) with many susceptibles but little actual drug users might look like for the two different objective functions. While the paths are rather similar when the sizes of user and susceptible groups are rather small, this changes for a larger number of users. Fig. 4.3 reveals the different optimal control strategies: In the beginning it is optimal to do nothing (the according control constraint becomes active). If the harm per user is described more adequately by the base case objective function. In the other case, i.e. when using the exponential objective function, one could do a little bit, simply because harm reduction is more efficient then if applied only a little. In both cases the number of users increases, while the number of susceptibles starts to fall. The control spendings in the base case soon exceed the spendings in the exponential case, because a decision maker would be rather hesitant to give larger price reductions if this would mean a larger drug usage caused by the existing users and therefore the measures become less and less efficient. This is also the reason why after a certain time one would do the maximum amount of harm reduction possible in the base case, but not in the exponential case. Resultingly, due to the higher incentive to potential users, the number of drug users increases slightly more in the base case. However, after some time there are not enough potential users left who could follow the example of existing users, and therefore the number of drug consumers starts to fall until the steady state is reached. The initially fast increase of users because of the many imitators and the following decrease of their number because of growth limitations due to the only constant inflow to susceptibles is the reason why there is an overshoot and the steady state is not approached

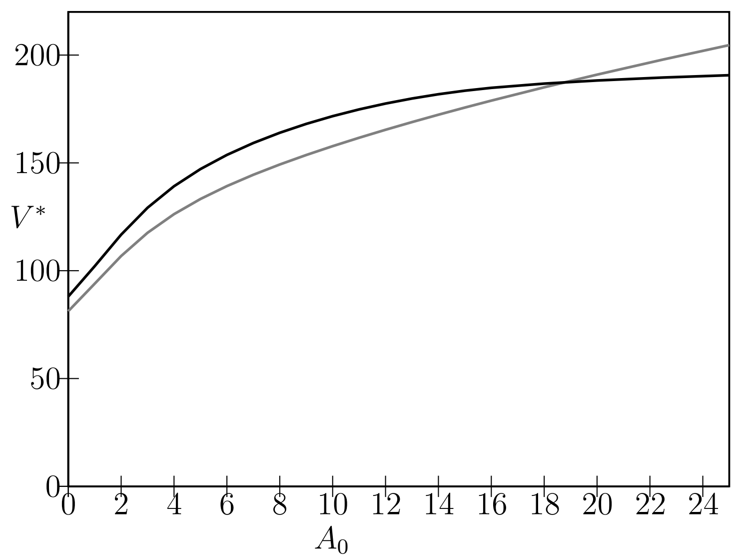

Figure 4.4: Value of the objective function for different initial values of users and susceptibles for the base case (black lines) and the exponential case (gray lines) for the single steady state case.

directly. Note that the unfavorable impact of harm reduction on initiation is not so strong that it can prevent the decrease of the drug users; if there are hardly any susceptibles left who find motivation in such harm reducing measures, the negative impact is too small to make much of a difference. Because of the lower control spendings the steady state in the exponential case consists of little less users and more susceptibles.

As their motivation to initiate drug usage does not depend on others, it is primarily the innovators who become drug users if the number of existing users is low. However, if the user group starts to grow due to these people, the influence of people taking drugs on people who do not becomes larger and larger and after some time the drug problem becomes rather large. Later we will see what happens if there are no innovators, which is particularly interesting if the number of existing users is rather small.

Fig. 4.4 shows the value of the objective function for different initial values of users, where the number of initial susceptibles is fixed at $S_0 = 2$ for both objective functions. It can be seen that if there are little users initially less harm arises in the exponential case. The reason for this is that for the most part of the optimal solution control spendings are small, because the arising harm, is because of the lower number of users, small too. As we assume in the exponential case that the higher control efforts lead to additional harm caused by increased usage of drug use, small amounts of harm reduction measures are more efficient. However, this changes if the initial number of users is large: Then it was assumed in the base case that all of the harm can be eliminated, but not in the exponential case. This means that in the exponential case the harm caused by a large number of users contributes to the size of the value of the objective function, while this is not the case using the base case objective function. There the actual harm by users is zero and only the costs for the control application arise on parts of the solution where it is optimal to do the maximum amount of harm reduction.

The Influence of Parameters on the Steady State

There is only one steady state for the given parameters. But what happens if these parameters change? The question arises how any change of the given parameters affect the steady state and therefore the optimal solution. Depending on the parameters the system will change in the following way:

- b If the impact of existing users increases, then there will be more users in the steady state. There are two reasons for this: On the one hand the increased flow directly leads to a larger number of users and, on the other hand because of the higher number of users more control spendings are required increasing the incentive of susceptibles to start taking drugs. Of course then the number of susceptibles becomes smaller. A decrease of the impact of existing users on the other hand leads to a higher number of susceptibles and a lower number of users and less control spendings in the steady state (and therefore would be desirable for the decision maker, since there is less harm caused by users due to the lower initiation and decreased costs due to the smaller spendings on harm reduction). Note, however, that if b increases, the growth of the number of users is limited by the number of available susceptibles and their constant inflow rate. Also note that not only the steady state is affected by the parameter change but also the optimal solution. For example more control efforts are necessary if the number of susceptibles is low (but not too low) and the number of users is high. If this initiation parameter is big, more susceptibles will become drug users and the shape of the optimal solution is slightly different: Due to the larger initiation the number of users can still grow when for a lower parameter b the number of users would already fall because there are not enough susceptibles becoming users to support further growth.

- a The fraction of innovators among the susceptibles, who start to adopt drug usage independent of others, also plays an important role for the steady state value. A larger parameter a means that more people start using the drug, which again has to a higher direct and indirect impact on the number of users, because of the additional initiation and their influence on potential consumers. However, if the number of existing users is very small, it is almost only the innovators who become users. Therefore if their fraction is very small or even zero, the system behavior changes and a second steady state becomes relevant, which we will consider more closely in the next section.

β If the impact of the control on initiation becomes stronger, i.e. users are more inclined to take drugs because of harm reduction measures, the number of users in the steady state will increase and the number of susceptibles will fall. Because of the larger number of users one has to increase control spendings in order to adequately deal with the arising harm. This, however, would lead to an additional increase of the user group.

c If the costs for harm reduction measures change, then the steady state would change too since a decision maker has to rethink his optimal strategy if it gets more or less expensive. If harm reduction becomes cheaper one would spend more a little more for such measures in order to reduce the harm. Yet, one would not do too much more because of the corresponding increase of initiation. Therefore, if the costs sink there would be more users and less susceptibles in the steady state. On the other hand if the costs for harm reduction rise, one would do less leading to a smaller number of users in the steady state. The impact on the optimal solution is such that if costs are large one would do less harm reduction.

μ A larger quitting rate of the users means that there due to a higher outflow, there will be less users and more susceptibles in the steady states. As such, less control efforts are required and the number of users becomes even smaller. If, however, the quitting rate becomes smaller, users stay true to their undesirable habit for a longer time, meaning that there are more people causing harm by their behavior and by their influence on others. Then more harm reduction is needed leading to a further increase of users.

Table 4.4: Parameter values for the weak DNSS case

r	c	k	δ	μ	a	b	β	α	h	m
0.04	8.247	1	0.05	0.0976	0.0	0.01	0.1732	1.75	0	1.6449

4.3 Weak DNSS Curve

If one sets the fraction of innovators in the initiation function to zero, there are as in the marketing model two steady states serving as candidates for an optimal solution. Their stable manifolds are now separated by a weak DNSS curve, meaning that unlike in the marketing model a decision maker does not have the choice between different optimal solutions, at least not for the used parameters.

This case has already been described in Chapter 1.3.1 as an introductory example. There only the results of base case objective function were described. However, as before results are pretty similar to the exponential case. Particularly if the number of users is small, it does not make much of a difference what function is used. If there are many consumers of a drug, a decision maker would apply more harm reduction measures under the base case assumption. This would lead to a higher initiation and therefore a larger user group.

The optimal strategy would be as follows: On the left side of the weak DNSS curve one would be careful not to give potential users additional incentives to start taking drugs, because that might unnecessarily lead to an escalation of the problem. Also, if there are hardly any consumers it is not required to do much, since the lower harm would not justify the costs. On the right side of the weak DNSS curve, if the number of users is small one should not apply any control measures, because there is not too much harm arising and additional users would only lead to a faster escalation of the current drug situation. Then, if the number of consumers increases, a decision maker should increase his control efforts in order to fight the harm. A really big user group might make the maximum amount of harm reduction measures necessary. As mentioned in the exponential case a decision maker would be more careful with the applied measures, since it is assumed that harm cannot be entirely erased and any effort would make drug usage more attractive to existing users.

The reason why it makes sense that there is “only‘” a weak DNSS curve

Table 4.5: Steady state values for the parameters shown in Tab. 4.4 if the base case and the exponential objective functions are used. The third and the sixth are no candidates for an optimal solution; they are relevant for the location of the weak DNSS curve.

	$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
Base Case	20	0.0	0.0	7.2674	0.0	0	0
Base Case	1.66	9.4	3.29	3.82	0.5651	0	0
Base Case	19.22	0.40	-0.73	-32.73	0.04	0	0
Exponential Case	20	0.0	0.0	7.2674	0.0	0	0
Exponential Case	1.69	9.38	3.64	4.25	0.4447	0	0
Exponential Case	19.21	0.40	-0.69	-31.34	0.05	0	0

here, is that no matter whether one starts closely on the left or the right side of the curve, it would not make any sense to do any harm reduction. Closely on the right side of the curve one would not take any harm reducing measures, because the number of users is so small that the compensated harm would not justify the costs created by the application of the control measures and only lead to a faster escalation of the drug problem. On the left side of the DNSS curve, the main reason, why it does not make sense to do any harm reduction, are not so much the costs, but the mentioned increased flow from susceptibles to users. The additional users would not only contribute directly to the total harm arising but lead to a further increase of initiation and then the steady state with no users could not be reached anymore for the same initial level of users and susceptibles. The corresponding phase portrait has already been shown in the introduction (see Figs. 1.2 and 1.3).

Looking at Fig. 4.5 we find that the value of the optimized objective function depends on the initial values of the state variables. If the initial number of drug users is big the arising harm is larger than in a case with little initial users. Yet, if the drug problem reaches a certain size total harm does not become much larger anymore by an additional initial drug user in the base case. The reason for this is that one would apply the maximum possible amount of harm reduction. The control instrument is the percentage reduction of the harm per user and is limited to a maximum value, which is assumed here for simplicity as $v_{\max} = 1$, meaning one could reduce any harm caused by users. The costs are assumed only to depend on the harm reduction measures and not on the number of people to which they are applied. As such,

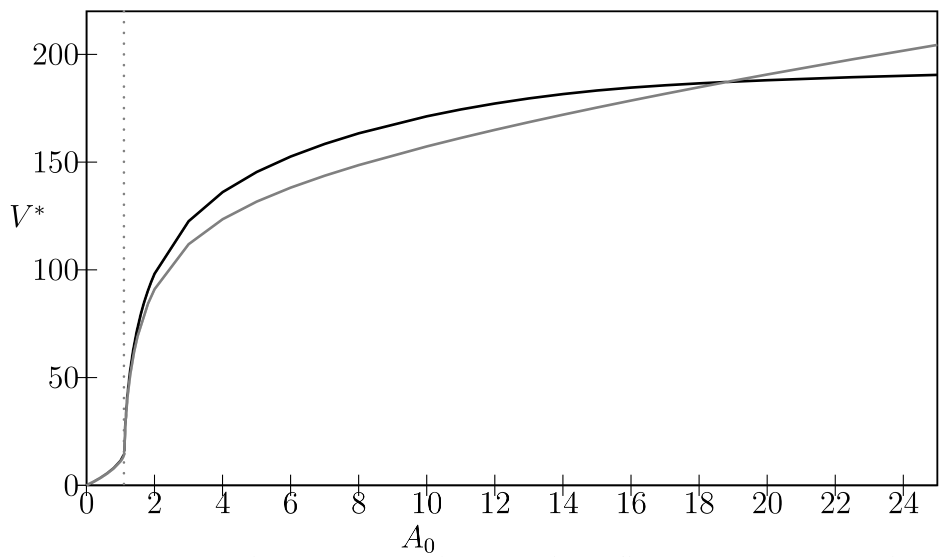

Figure 4.5: Value of the objective function for different initial values of users and susceptibles for the base case (black lines) and the objective function (gray lines) when a weak DNSS curve (shown by the dotted line) occurs.

no harm would arise theoretically to society if one would take all possible measures in the beginning for a large number of initial users, only the cost would have an unfavorable impact on the objective function. Additional users would only increase the harm by making the decision maker apply the maximum control measures for a longer period of time. This is not as "expensive" as if each additional user would increase the harm by his mere presence as is the case when using the exponential objective function.

The value of the objective function depends on which side of the weak DNSS curve one starts: Of course if one has the chance to let the drug problem disappear it is smaller than if the drug problem escalates and leads to the steady state with a high number of drug consumers. It can be seen again on the right side of the weak DNSS curve total harm rapidly increases if the number of initial users changes only slightly. Similar to the marketing model, the reason for this is that the time needed for the problem to develop such a size to be of a larger concern is very long close to the DNSS curve and so not much harm arises over the relevant time period. As such the value particularly of the second costate is very large closely to the weak DNSS curve: An additional user is very disadvantageous then.

Comparing the values of the two objective functions, we see that if the initial number of users is rather small the values of the objective functions are rather similar. This changes, however, for a larger initial group of drug consumers. For a small number of users total harm is larger if using the base case objective functions, the reason for this is that the initial amount of measures taken is not considered as efficient in reducing the harm per user and over the course of the solution path more people will become users. However, if the number of drug users is rather large this changes: Since harm per user is considered as being not totally erasable in the exponential case, harm still increases proportionally to the number of users if the maximum control is applied. Also costs for control application arise. In the base case the assumption that one can eliminate any harm caused by users, if one would do this there would only be the costs for the measures and no harm as such which increase proportionally to the number of users.

Depending on which side of this weak DNSS curve one starts, the drug problem either escalates or disappears. While a decision maker cannot control this outcome using harm reduction measures, in such a case he might consider, if desired, the application of some other, maybe one-time measures in order to manipulate the number of users and/or susceptibles to get to the politically more convenient side of the weak DNSS curve. Note, however, that such measures are not included in this model and if one did so one would also have to consider that there might be costs attached.

Where this weak DNSS curve can be found depends again on the parameters. If the impact of existing users reflected by parameter b becomes larger, then less users are needed for the problem to escalate. Because of this the curve also becomes more steep, i.e. a low number of susceptibles cannot prevent the escalation of the problem as much as before due to a smaller initiation. Parameter a, i.e. the fraction of innovators among the susceptibles, has a similar impact on the weak DNSS curve. On the other hand if parameter μ increases the outflow from the user state becomes larger and due to this the influence of existing users and therefore initiation becomes smaller. Thus, more initial users are required that the problem escalates and the DNSS curve becomes less steep, meaning that if the number of susceptibles is not big enough the problem cannot become too big. If control application becomes cheaper, i.e. the costs c decrease, then a decision maker has higher incentives to use his control instrument. Then a decision maker would apply harm reduction measures already for a lower number of users. If, however, he has the possibilty to avoid the escalation of the problem, he would still do nothing in order not to provide additional incentives for potential user to start taking drugs. If the outflow rate of the susceptibles δ becomes larger, there are more users necessary for an escalation of the problem because there are less susceptibles who can be motivated to become users. Similar effects on the DNSS curve has a decrease of the constant inflow k to susceptibles.

Table 4.6: Parameter values for the limit cycle case

r	c	k	δ	μ	a	b	β	α	h	m
0.04	3	1	0.05	0.35	0.02	0.0272	0.1732	1.75	0	1.6449

Table 4.7: Steady state values in the limit cycle case if the base case and the exponential objective functions are used.

	$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
Base Case	5.9609	2.0056	2.3046	4.0660	0.3003	0	0
Exponential Case	5.9563	2.0062	1.9969	3.5214	0.3041	0	0

4.4 Limit Cycle

Another possible result is that of the limit cycle. Tab. 4.6 shows the parameters for which a limit cycle is the only admissible candidate for an optimal solution for the base as well as the exponential case. Tab. 4.7 shows the relevant steady state, however, as they are instable they are no candidates for optimal solutions. The eigenvalues of the monodromy matrix of the calculated limit cycles are both such that two of them are greater than one, one is smaller and one is equal to one. This means the dimension of the stable manifold of the limit cycles is two.

A large outflow from users (e.g., caused by a large death rate or efficient treatment), for instance, can be the crucial reason for the occurence of cyclical behavior. Starting with some susceptibles and only a few users (phase I), the number of susceptibles can rise due to the constant inflow to the susceptibles and the relatively low flow to the users. The number of consumers can grow due to the rising flow from susceptibles and relatively low number of users quitting drug usage. If the number of users grows stronger and stronger, the constant inflow to S is not sufficient anymore to allow a further increase of this state. The number of susceptibles starts to fall, while the consumer group grows (phase II). That means that in the state equation describing the development of the susceptibles $f(A)Sg(v) + \delta S$, i.e. the components describing the outflow from this state, exceed the inflow k. However, after

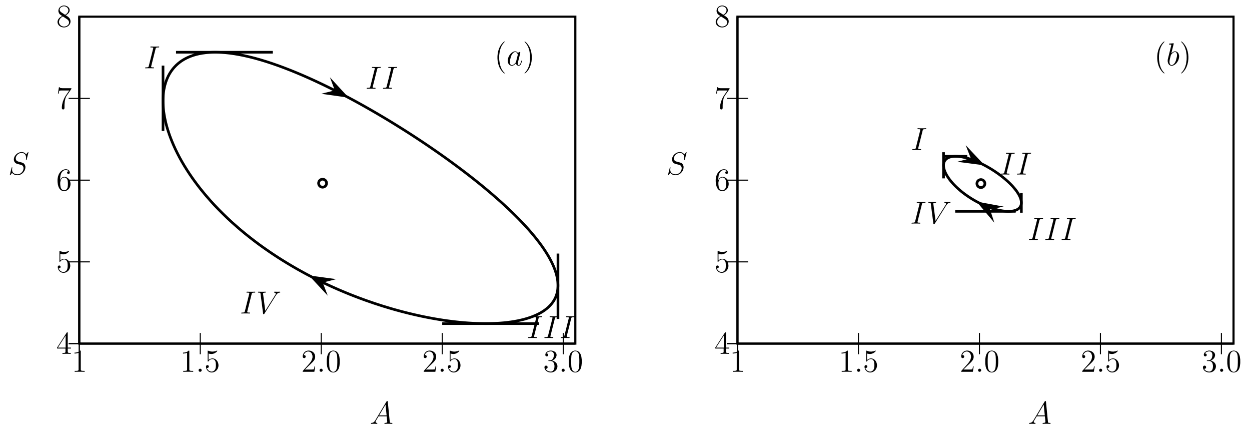

Figure 4.6: Limit cycle for the two different objective functions, where (a) corresponds to the base case and (b) to the exponential case. Steady states which are no candidates for where an optimal solution can end in, but are relevant for the cycles, are shown by $\circ$.

a while the number of potential drug users will become so small that there are not many people left who could be convinced to start taking drugs and due to the large outflow rate, the number of drug consumers starts to fall too (phase III). This happens if in the state equation $\dot{A}$ the outflow μA becomes greater than the inflow $f(A)Sg(v)$. If there are not many people left who influence others in their drug adoption decision, the number of susceptibles will increase again due to the constant inflow, while the number of consumers might still fall (phase IV).

Initially it is not optimal to do much, the lower harm caused by this measures cannot compensate the costs. To fight the harm caused by additional users when the problem becomes larger, one would do much to decrease total harm. Then one has to accept the larger number of drug users and the arising costs. With a falling number of users one can again reduce the control efforts.

Fig. 4.6 depicts limit cycles for the parameters shown in Tab. 4.6. Fig. 4.7 shows the corresponding time paths of the state variables and the control. Comparing these two cycles we find that, while their period is rather similar with $\Theta_{base} = 34.81$ and $\Theta_{exp} = 34.64$, using the base case objective function, the oscillations will be larger. The reason for this is that initiation is slower in the exponential case due to the lower number of susceptibles when the number of users is small, and the smaller control spendings when the number of users is large. This means that initiation is smaller and takes approximately the

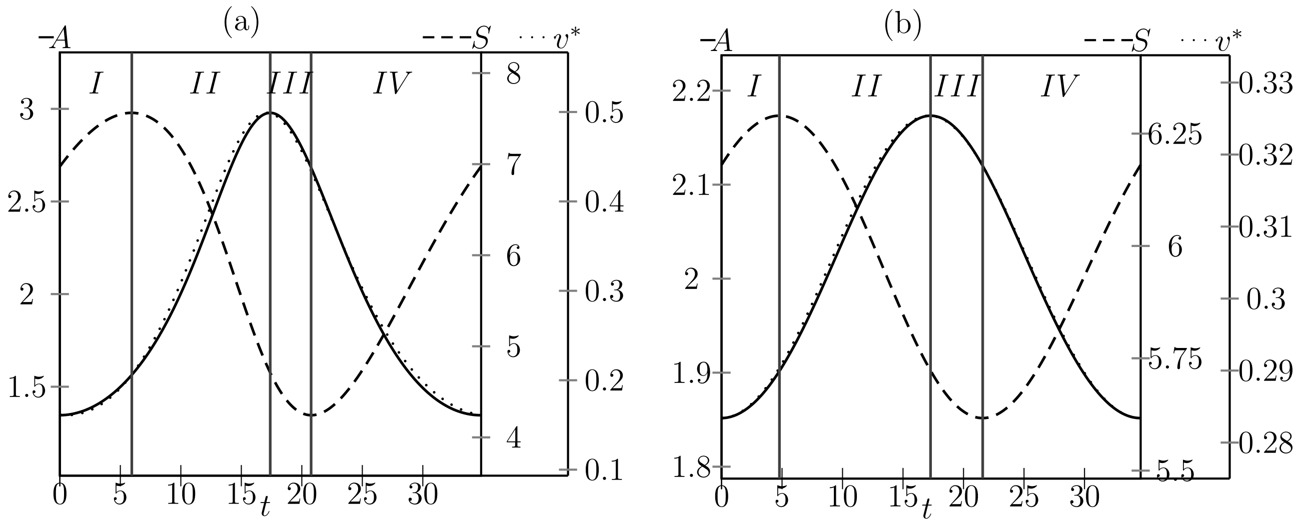

Figure 4.7: Timepath of the control corresponding to the phaseportrait Fig. 4.6.

same amount of time as in the base case with the larger oscillations.

In the exponential case, the reason why one would not change one's optimal control application much over the course of the cycle is that small harm reduction measures are more efficient in terms of reducing more harm per user. As such, one would be more willing to do more harm reduction if the number of users is rather small and would not increase the control spendings much even if the number of users rises. Since the additional flow, which would be particular large if much harm reduction is made when the number of users is large, is then quite small, the oscillations will be smaller. Fig. 4.6 shows the larger application of the control measures over the course of the cycle in the base case. However, this additional flow from the potential to the actual consumers of a drug as well as the higher number of susceptibles increases initiation so much that more people would start taking drugs, making again to higher control efforts necessary.

Remark on the Numerical Calculations *Again the cycles were found with the help of the OCMat and Matcont toolboxes. Note while for both objective function limit cycles could be found for the same parameters this is not always the case. This can be seen calculating the relevant steady states for different values for a certain parameter and by looking at the parameter value where the Hopf bifurcation occurs.*

Now consider that the initial number of users is small and there are a few susceptibles. In phase I both the number of users and susceptibles

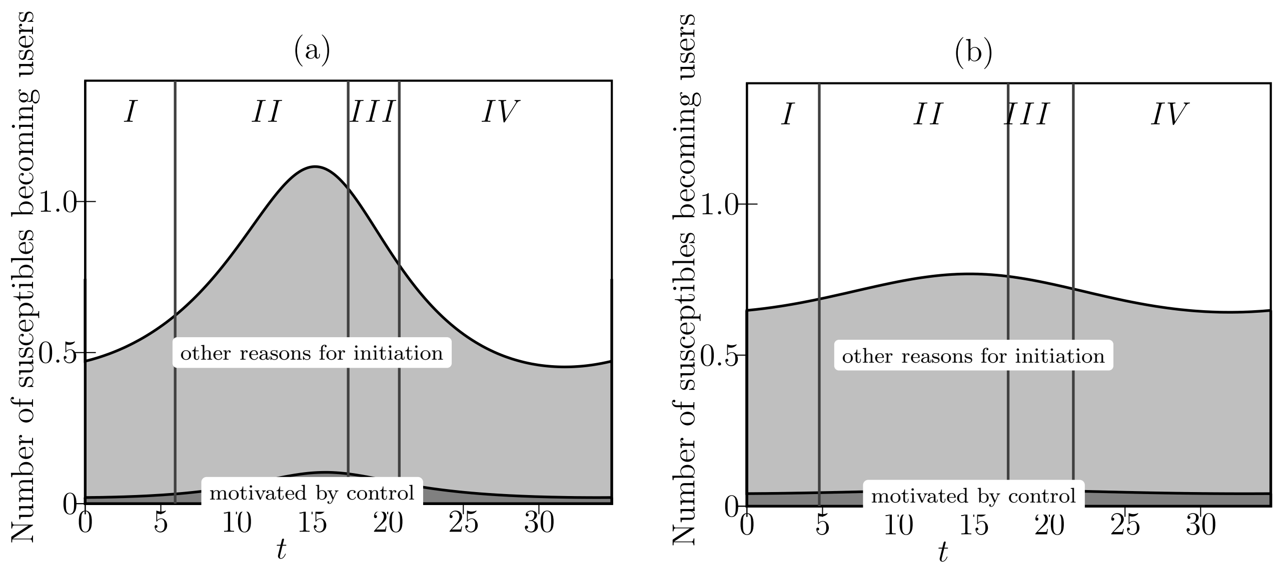

Figure 4.8: Number of people who decide to start drug users because of increased attractiveness of the drug due to lower harm and people who are at most only indirectly influenced by this for other reasons in the base case (a) and the exponential case (b)

can grow. It can be seen that the initial number of users on the cycle is smaller for the base case, while the number of susceptibles is much larger there. Resultingly, the amount of harm reduction would be much larger in the exponential case, because on the one hand one would reach more users and on the other hand the danger of affecting too many susceptibles in their adoption decision is rather small. Yet, one would also do some harm reduction measures if the problem is most adequately captured by the basic objective function to decrease the harm caused by the few existing users. Due to the smaller number of existing users phase I lasts a little bit longer in the base case.

As the number of users rises even further in the second phase it is necessary in both cases to raise the control efforts. Since their growth is stronger if using the basic objective function so must be the increase of harm reduction measures. Then, in the third, but also in the fourth phase, the number of users starts to fall because there are not enough susceptibles left to permit further growth. Interestingly, in the base case one would decrease control efforts before the number of drug users falls in order to prevent additional initiation caused by a higher attractiveness of the drug, while in the expo-

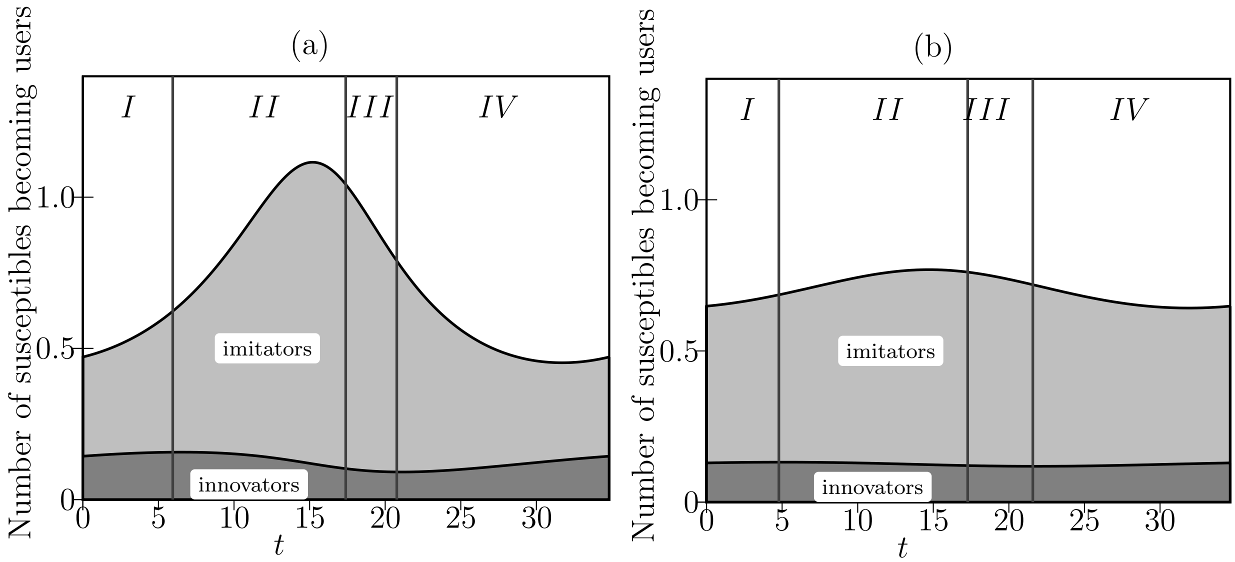

Figure 4.9: Number of innovators and imitiators among people who start drug usage in the base case (a) and in the exponential case (b)

nential case one decreases control efforts basically at the same time as the users' number decreases. If one considers the base case, the number of users as well as their oscillation is rather big on the cycle and one would have to be particularly careful even then not to rise an additional flow that would worsen the result. In the last phase the number of susceptibles can recover due to the lower influence of the existing users.

Fig. 4.8 shows the fraction of people who decide to start taking drugs because of lower risks due to harm reduction measures. It can be seen that the oscillations are not particularly large in the exponential case, the reason for this are the rather small changes of the control spendings over the course of a cycle. In the base case the increase of the people motivated by the lower risks is larger than in the other case, which of course results of the higher control efforts due to the bigger group of users. In both cases the number of people motivated by a lower harm are the largest when initiation reaches its maximum. This is because one would put most efforts into fighting the harm when the drug problem is big even though that would lead to higher incentives for non-users to start taking drugs.

Similarly the change of the fraction of innovators is only very small in the exponential case as can be seen in Fig. 4.9. This is not only because of the little changes in users and susceptibles over the course of a cycle, but also

Table 4.8: Parameter values for the weak DNSS with a limit cycle case

r	c	k	δ	μ	a	b	β	α
0.04	5	1	0.05	0.32	0.0	0.03968	0.1732	1.75

because when the number of susceptibles is rather large, control application and therefore the additional initiation is pretty small. On the other hand, when the additional flow caused by harm reduction measures is rather large, the number of susceptibles is rather low and there are simply not many people left who could decide to start taking drugs independently of others. The fraction of imitators grows and falls with an increasing and decreasing number of drug users. This also occurs in the base case. What can be rather well seen in this case is that the fraction of innovators decreases and rises together with the number of potential drug users. In both cases it can be seen that the largest part of the people who decide to start taking drugs are mostly imitators and the fraction of innovators is rather small. This depends, of course, entirely on the used parameters.

4.5 Weak DNSS Curve with Limit Cycle

The next case is that both, a weak DNSS curve and a limit cycle occur. The parameters needed to gain such a result are again such that the fraction of innovators, who start taking drugs independently of others, is zero or very small and the outflow from the user-state is rather large. Then starting on the left side of the weak DNSS curve the impact of the existing users would not be large enough to be of any concern and one would always end in a steady state with no users. However, on the right side of the curve the optimal solution would oscillate, i.e. never reach a point with a fixed number of users and susceptibles.

The parameter used are shown in Tab. 4.8 and the corresponding phase portrait can be seen in Fig. 4.10, which depicts the steady states, the cycle and the weak DNSS curve. Tab. 4.9 shows on the one hand the steady state with no users, which has saddle properties and is a candidate for an optimal solution and the for the limit cycle relevant steady state, which is unstable. The third steady state is also no candidate for an optimal solution, however, it is relevant for the location of the weak DNSS curve.

Table 4.9: Steady state values in the weak DNSS curve with limit cycle case for the base case. The last two are not candidates for where an optimal solution can end, but they still have some influence on the system behavior.

$\hat{S}$	$\hat{A}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{v}^*$	$\hat{\nu}_1$	$\hat{\nu}_2$
20	0.0	0.0	2.7789	0	0	0
3.8743	2.5196	2.7943	4.0027	0.2358	0	0
17.795	0.3445	−0.375	−5.8269	0.0448	0	0

Remark on the Numerical Calculations *This section only focuses on the base case. In the exponential case a cycle was found for the used parameters, however control constraints were violated. Locating a limit cycle with active control constraints is not possible with the current version of the OCMat toolbox. However, theoretically this is possible by solving the following boundary value problem: Find a trajectory starting and ending at the same point, consisting of parts where the control constraints are active and inactive. The only practical problem is to provide an initial solution.*

The curve is again "only" a weak DNSS curve as closely to the curve the optimal strategy would always be to do nothing. On the left side of the curve harm reduction would do no good, because harm is low anyway. Any action would only increase the incentive of the susceptibles to start taking drugs, which might lead to an escalation of the problem. Due to the low harm it also optimal on the right side close to the curve to do nothing, because the costs and the impact on the potential consumers would not be justified by the lower harm.

On the right side of the cycle a limit cycle occurs: If the number of users is small, both their number and the number of susceptibles can grow due to the constant inflow and the rather small flow from S to A due to the rather low influence of existing drug users on potential consumers. After some time the number of users and their impact on susceptibles becomes so large that the potential user group starts to become smaller, while there is still growth of the actual consumer group. Due to the larger outflow and the smaller group of susceptibles their number starts to fall after some time too. When the impact of existing users becomes smaller, the number of susceptibles can grow again.

Fig. 4.11 shows the timepaths of the two trajectories that were depicted in Fig. 4.10. One of these optimal solutions ends in a steady state with no users

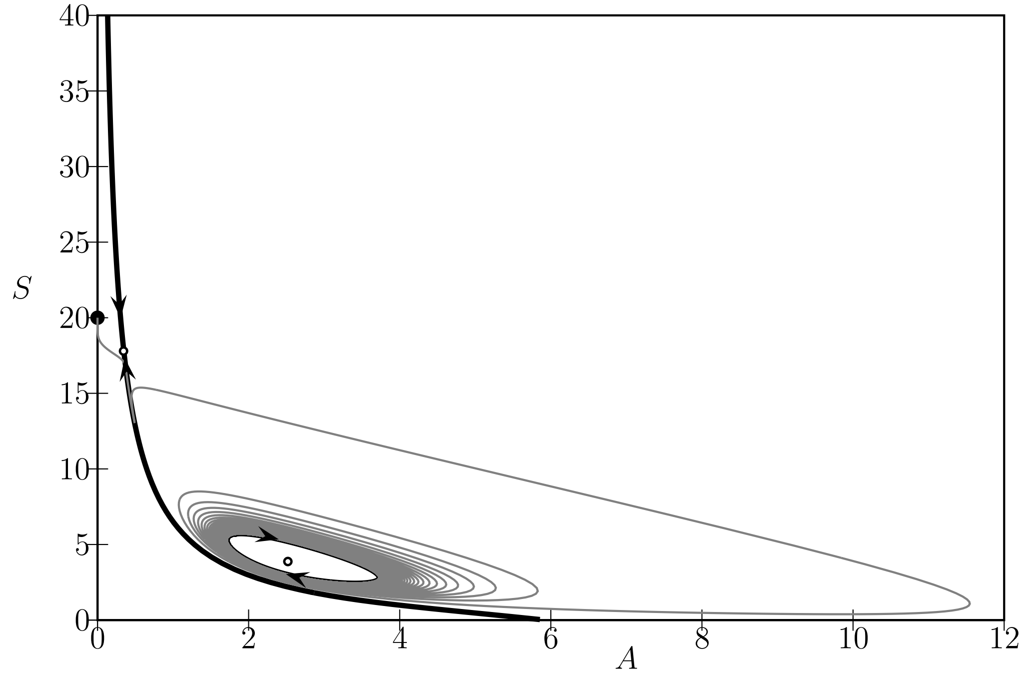

Figure 4.10: Phase portrait depicting the case where both a weak DNSS curve and a limit cycle occur. The grey lines show some trajectories ending either at the circle or at the steady state with no users.

and many susceptibles and the other one oscillates towards the limit cycle. While it is not optimal to do anything in the first case, in the second case it can be seen that the amount of optimal control spendings varies depending on the actual number of users and susceptibles and even at certain points maximum and minimum harm reduction are optimal to be applied: If the number of users is low then it is optimal to do not much or even nothing. If the harm arising through a large number of consumers becomes big, then would start fighting the harm.

Remark on the Numerical Calculations *In this specific case where control constraints become active and inactive, finding an optimal solution converging to a limit cycle is very time consuming when solving the corresponding boundary value. Therefore the solution paths ending at the limit cycle are calculated by solving an initial value problem. The initial condition is given by starting at a point on the linearized stable manifold of the limit cycle. Practically this is done by determining the stable eigenvectors of the monodromy matrix. Then the ODE is solved backwards in time, analogous to the case of a steady state.*

Note that again more total harm over time arises if the initial number of users is large, due to the direct harm caused by them and due to the costs arising through harm reduction measures. Also if the number of susceptibles is large total harm is bigger as if their number was low because more people are tempted to start taking drugs. If the initial number of users and susceptibles is sufficiently small, however, and the steady state with no users can be reached total harm is smaller, because there is no long-term menace caused by the harm of drug usage.

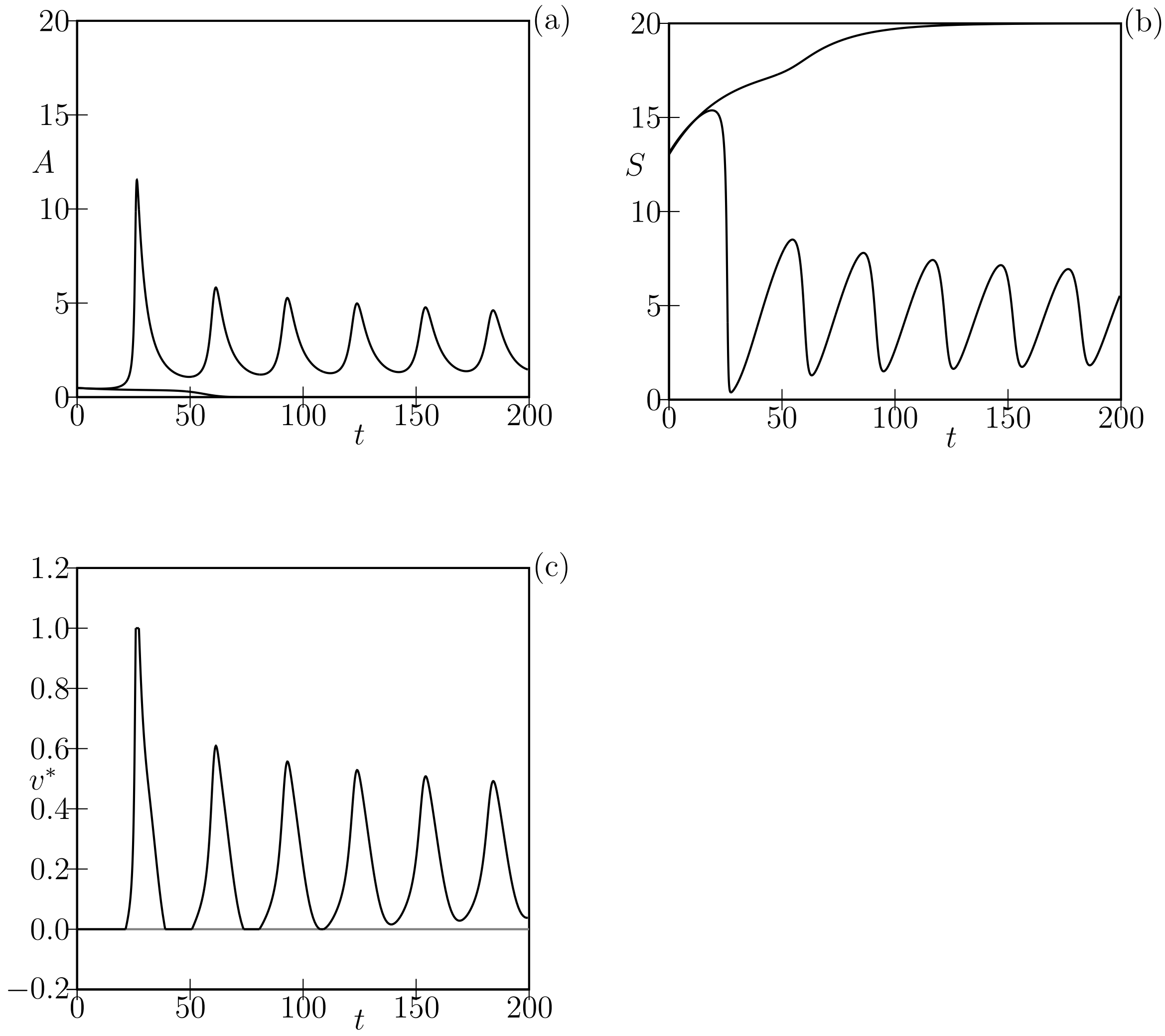

Figure 4.11: Time paths of the two trajectories depicted in Fig. 4.10, where panel (a) shows the development of the number of users (b) of the susceptibles and (c) of the optimal control.

Chapter 5

Conclusion

The aim of this thesis was to show that the used rather simple state dynamics can be helpful in providing valuable insights about how one can optimally influence the spreading of a certain good among the members of society for different applications.

In the field of marketing a decision maker's objective was to maximize his profits, his tool to influence his outcome was giving promotional price reductions, which has a negative impact on the company's profits.

Not surprisingly, we saw, that the optimal strategy heavily depends on the initial number of potential and actual customers of a product: If the number of users of a certain good is rather small in the beginning, there cannot be much of a for the decision maker advantageous interaction between this group and the susceptibles. Therefore large price reductions should be given then in order to make the product more attractive. If, due to the lower price, the customer base grows, a decision maker can reduce his control efforts, because the incentive to start using the product due to its popularity increases accordingly. If, however, the market potential, i.e. the number of susceptibles, is rather large, one should also give price reductions, because such measures are very efficient then as a large audience can be reached.

Another thing that was seen is that rather complex behavior of the system is possible, i.e. one might be able to find an optimal control strategy with an oscillating number of potential and actual customers with this simple model. Depending on the used parameters, there are several ways how an optimal solution might look like, sometimes a decision maker would always reach a high number of customers in the end, sometimes he would basically have no other choice than to let the product fail, if acting optimally. Sometimes, he even might have the choice between making the product successful under high efforts and doing nothing, even though that means the product will find hardly any buyers effectively. The most complex behavior found was that a decision maker might have the choice for certain parameters between letting the product fail and an oscillating number of users and susceptibles.

As an extension of this basic marketing model, a two-stage version was introduced in order to see how a sudden change of certain parameters, triggered by some event, influences the optimal solution and the spreading of a product. It was assumed that in the first stage the company was a monopolist, but has to face competition in the second.

Of course, it is necessary to adapt the optimal strategy according to the parameters of the different stages. It was seen, that the optimal control might jump at the switching point, the direction there depends on the relation of the change of initiation and the average profit per user reflecting the market price.

If all relevant parameters worsen in case of the market entry of competitors, a decision maker would have no incentive to admit competition. However, when he has no choice of doing so, e.g., when a patent expires, it is always better for him to stay monopolist as long as possible. It is easer for him then to gain and to hold customers, who can create even higher profits due to a higher market price.

Each stage must have advantages if a decision maker considers to optimally switch from one stage to the other. This can occur for instance if a decision maker acts particularly aggressive in the first stage in order to keep competitors out of the market. Such a strategy would mean for customers, if their initial number was low, that they first would get attracted to a product because of a low standard price (and possible price reductions). Then, if the usage product becomes popular, they would have to pay a higher price for their remaining time as users in order to be able to benefit from the advantages of a higher user group.

The optimal switching point itself, however, depends on the parameters used in the different stages and how much they change, respectively. If for instance, initiation gets much worse in the second stage, one would remain longer in the first stage for an initially small number of users.

The second field of application for this model is drug policy and, more specifically, harm reduction. There, the optimal solution again depends on the starting point. If the initial number of users is low it is always optimal not to do much harm reduction: The cost of such measures and the additional harm caused by people who start taking the drug because of the lower risk would exceed the lower harm arising to society. If the potential drug market, i.e. the number of susceptibles for a certain drug, is large, one also has to be careful regarding harm reduction measures. The additional attractiveness of the drug due to harm reduction can very easily lead to a faster escalation of the drug problem. However, if the initial number of users is large, then a decision maker has to accept the costs of control application and the additional initiation in order to be able to adequately respond to the large harm arising.

For certain parameters, i.e. when the number of people who start taking drugs independently of others is zero or very small, the drug problem might disappear after some time if the initial group of drug users is rather small. Then the influence of drug users is not big enough to make many people imitate them and their number will decrease. In such a case one has to be particularly careful about the amount of harm reduction measures taken, because the additional number of users might lead to such an increase of the number of customers, that due to their rising influence the problem might escalate after some time.

The optimal solution depends on the objective function used. Comparing the results, using two different objective functions it was seen, that if the drug problem is small, it does not make much difference which one of the two described is used: The strategy concerning the size of the applied measures and the development of the number of susceptibles and users is not much different then as not much harm arises anyway. However, this changes with an increased number of users. Then one has to be more careful in choosing an objective function fitting to one's problem (which in this case differ by the assumptions made regarding the impact of control application on the harm per user).

It was also seen that optimal strategy and the behavior of the system depend on the used parameters, oscillations of the number of susceptibles and drug users might also occur for the harm reduction models. Then the decision maker would have to face sometimes a larger, sometimes a smaller drug problem. Another interesting outcome was that a weak DNSS curve and a limit cycle can exist simultaneously: Then depending on the initial value of the state one would either be able to erase the drug problem or finally have an oscillating number of potential and actual consumers.

Of course there are many extensions possible for the models described here, most important of all the next step would be to parameterize them and see whether it is possible to explain and influence the spreading of real goods with this simple model.

In the marketing case, for example, one could add a backflow from users to susceptibles, i.e. a certain fraction of people who stop using a product might be still interested in using the product, or include another state to adequately deal with technological progress or consider the influence that users of a different product might have on the initiation decision. Another possible extension for the marketing model would be to consider a differential game, where, e.g., one single competitor enters the market. For both, marketing and harm reduction it might make sense to consider the age distribution of the susceptibles or split up the user group, e.g., into light and heavy users.

While these are just a few possible extensions, I think this thesis made clear that it is possible to find some very interesting results using this particular dynamics describing the development of the number of potential and actual adopters of a good and that it certainly would make sense to do further investigations.

Part III

Appendices

Appendix A

A Short Guide to Optimal Control Theory

In order to improve the readability of this work for people, who are not familiar with optimal control theory, a short compendium of the used terminology is given in the following. More detailed information can be found in optimal control literature, such as Grass et al. (2008), Feichtinger & Hartl (1986), Léonard & Long (1992), and many others.

Optimal control problem: Often a decision maker faces an optimization problem that has a *dynamic* character, i.e. it evolves over time. Therefore, when choosing his optimal strategy he has to include this development over time and consider the underlying *dynamical system*. It is assumed that the control problems here evolve continuously and are autonomous. An optimal control problem is called *autonomous* if its occurring functions do not explicitly depend on time. Optimal control theory provides the tools to find an optimal solution of an optimal control problem.

Objective function: A decision maker's objective is to maximize or minimize his profits or losses, respectively, over time. For this purpose, for the models in this work, we have to consider an *objective functional* of the form $V = \int_0^T e^{-rt} g(x(t), u(t)) \mathrm{d}t$, where the *discount rate* $r$reflects the decision maker's time preference regarding whether he or she puts more emphasis on the near or the remote future. T denotes the *time horizon* over which the decision maker wants to maximize or minimize, and is assumed to be infinite here. The function $g(x(t), u(t))$ describes the profits or losses at each instant of time, $x(t)$ denoting the state and $u(t)$ the control variable.

State variable: At each instant of time, a dynamical system is in a certain state. The state variables $x(t) \in \mathbb{R}^n$ characterize the system behavior, meaning that they display the key aspects of the system, that a decision maker might want to control. The size of a state cannot directly be manipulated by a decision maker, only by virtue of the control. The development of a state variable over time as well as how it is influenced by the current state and control at time t is described by its *state equation / dynamics* $\dot{x}(t) = f(x(t), u(t))$, where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $\dot{x} = \mathrm{d}x/\mathrm{d}t$.

Control variable: The control variable $u(t) \in \Omega \subseteq \mathbb{R}^m$ (with control region Ω determined by control constraints $c(u(t)) \geq 0$) serves as the decision maker's instrument to influence the system's development over time. *Control constraints* ($c(u(t)) \geq 0$ with $c : \mathbb{R}^m \to \mathbb{R}^l$) determine whether a control value is admissible, meaning there can be restrictions among which values of the control a decision maker is allowed to choose - here in this work it is assumed that the control has to satisfy $0 \leq u \leq u_{\max}$. The control can be determined optimally with the help of *Pontryagin's Maximum Principle*. Particular difficulties can arise, if a control constraint is violated/becomes *active*, then the control has to take its *boundary* (maximum or minimum) value. If it is a linear control problem, a *bang-bang solution* occurs if the control switches from one boundary to another.

Hamiltonian: In optimal control problems, the Hamiltonian function is of particular importance. It is defined in its *current-value* notation as

$$\mathcal{H}(x(t), u(t), \lambda(t)) = \lambda_0 g(x(t), u(t)) + \lambda(t) f(x(t), u(t)), \tag{A.1}$$

where $\lambda(t) \in \mathbb{R}^n$ is called *costate* or *adjoint variable* (see later description). For the here presented models it is assumed that the problem is *normal*, i.e. $\lambda_0 = 1$. The Hamiltonian reflects that the control influences the optimal solution directly (via the objective function) and indirectly (as it affects the states via the state equations), and the costate variable serves to weight these two influential factors. If there are constraints (such as control constraints $c(u(t)) \geq 0$), it is necessary to consider the *Lagrange function*, which is defined as

$$\mathcal{L}(x(t), u(t), \lambda(t), \nu(t)) = \mathcal{H}(x(t), u(t), \lambda(t)) + \nu(t) c(u(t)), \tag{A.2}$$

where $\nu(t)$ is called *Lagrangian multiplier*.

With the help of the Hamiltonian (and therefore also of the Lagrangian) one can determine the value of the objective function starting at a certain point $x(0) = x_0$, where

$$\int_0^\infty e^{-rt} g(x(t), u(t)) \mathrm{d}t = \frac{1}{r} \mathcal{H}(x(0), u(0), \lambda(0)). \tag{A.3}$$

Pontryagin's Maximum Principle: Pontryagin's Maximum Principle[1] says that if $u^*(t)$ and $x^*(t)$ is an optimal solution of an optimal control problem, then there exists a continuous and piecewise continuously differentiable function $\lambda(\cdot)$, with $\lambda(t) \in \mathbb{R}^n$ satisfying[2] for all $t \geq 0$, where $u(\cdot)$ is continuous:

$$\mathcal{H}(x^*(t), u^*(t), \lambda(t)) = \max_{u \in \Omega} \mathcal{H}(x^*(t), u, \lambda(t)) \tag{A.4}$$

$$\mathcal{L}_u(x^*(t), u^*(t), \lambda(t), \nu(t)) = 0 \tag{A.5}$$

$$\dot{\lambda}(t) = r\lambda(t) - \mathcal{L}_x(x^*(t), u^*(t), \lambda(t)), \tag{A.6}$$

the complementary slackness conditions

$$\nu(t) \geq 0, \quad \nu(t)c(x^*(t), u^*(t)) = 0 \tag{A.7}$$

and the transversality condition

$$\lim_{t \to \infty} e^{-rt}\mathcal{H}(x^*(t), u^*, \lambda(t)) = 0, \tag{A.8}$$

where the asterisk denotes optimality.

Equ. (A.5) and (A.7) imply that for $0 < u < u_{\max}$ $\mathcal{H}_u(x^*(t), u^*(t), \lambda(t)) = 0$. Note also that (A.6) can be written as $\dot{\lambda}(t) = r\lambda(t) - \mathcal{H}_x(x^*(t), u^*(t), \lambda(t))$, with $\mathcal{L}_x = \mathcal{H}_x$ since the constraints used in this work do not depend on the state variables.

Legendre-Clebsch condition: A necessary condition for (A.4) in the interior of Ω is that the second derivative of the Hamiltonian with respect to the control must be in case of a maximization problem[3] non-positive[4], which is called Legendre-Clebsch condition. In a linear optimal control problem this condition is trivially satisfied with $H_{uu} = 0$, then, however, one has to consider the *generalized Legendre-Clebsch condition*[5].

Costate variable: Due to Pontryagin's Maximum Principle costate variables have to be included[6]. The economic interpretation of the costate variable is that of a *shadow price*: it measures the highest price a rational decision maker would be (hypothetically) willing to spent for an additional

[1] This "version" considers an autonomous control problem with control constraints and infinite time horizon, which is assumed to be normal.

[2] $\mathcal{L}_u = \partial\mathcal{L}/\partial u$ and $\mathcal{H}_x = \partial\mathcal{H}/\partial x$

[3] Any minimization problem can be transformed into a maximization problem by multiplying the objective function with -1, if this is not done, however, the second derivative has to be non-negative.

[4] In case of more than one control the Hessian matrix H_{uu} has to be negative semidefinite (in case of a maximization problem).

[5] See, e.g., Grass et al. (2008).

[6] These are analogous to the Lagrangean multipliers of static optimization problems.

unit of the corresponding state variable at time t (see, e.g., Grass et al., 2008). The development of the costate is described in the *costate equation* which can be found with the help of Pontryagin's Maximum Principle as $\dot{\lambda}(t) = r\lambda(t) - \mathcal{L}_x(x^*(t), u^*(t), \lambda(t))$.

Steady state: In a steady state (also called *equilibrium* or *fixed point*) both the state and costate dynamics are zero, meaning that there is no change in the states or costates as time evolves[7]. It can be shown that in optimal control problems with positive discount rate, steady states have to be *repelling* or *saddle points*. Such stability properties can be determined with the help of the eigenvalues of *Jacobian matrix*, which describes a linearization of the system at this point.

Steady states are of particular relevance since they determine the system behavior: A solution lying on the *stable manifold* of a saddle point, which is the set of points, where a trajectory starting on one of them leads toward this steady state, fulfills the *transversality condition* and therefore is a candidate for an optimal solution.

Limit cycle: Another possibility where an optimal solution can finally end in is a limit cycle: Then the values of state, costate and control variable will never be fixed but develop in such a way that starting at any point of this cycle, after some time period Θ, one will be again exactly on this point again. The stability of a cycle can be determined with the help of the eigenvalues of the *monodromy matrix*, and it can be shown that in an optimally controlled system it can only be either repelling or of saddle-type.

Numerical solution: As it is not possible to find the optimal solutions of this work analytically, intensive numerical computations have to be done. In this thesis there are two approaches used to numerically find an optimal solution: The first one is by solving an *initial value problem* (IVP), starting on the stable manifold near the steady state or cycle and simply calculating backwards in time with an ode solver. However, the drawback of this method is that it can be rather time consuming to find a solution starting at a certain point with sufficient precision. In order to avoid this problem, one can use another approach to calculate the optimal solution, that is by defining a *boundary value problem* (BVP). Within this approach the conditions at the initial and end time are stated and a solution is computed starting with an approximate guess.

[7] The equilibrium values of the states, costates and control will be highlighted with a hat (e.g., $\hat{x}$)

However, one must not forget to include possible control constraints within the calculations, in the initial value problem this can simply be done by checking the value of the control or more elegantly of the Lagrange multiplier (which becomes zero if a constraint becomes inactive and greater than zero if it becomes active) and switching at the appropriate time to the system where the control either is optimally determined or takes constantly its maximum or minimum admissible value. One can also use the information concerning the Lagrange multiplier and a constraint becoming active or inactive to as a condition for a boundary value problem. It is possible to formulate a BVP in a way such that it includes boundary conditions regarding both, certain starting points as well as control constraints (more information about numerical calculations regarding optimal control problems can be found in Grass et al., 2008, Chap.7).

One can use various programs such as MATLAB[8], Mathematica[9] and Maple[10]. While the last two products are certainly very useful when it comes to the formal analysis of optimal control problems, due to the better programming environment (which includes not only the mathematically required functions, but also allows to use an object oriented programming style) a Matlab toolbox OCMat was implemented to allow such numerical calculations as described. It can be found online at

`http://www.eos.tuwien.ac.at/OR/OCMat/index.html`

and will be described in more detail in Appendix B.

Bifurcation: When considering a dynamical system it is not only interesting to see how it evolves, but also how a change of parameters affects the outcome, particularly regarding the stability properties and the occurrence of steady states and limit cycles. A *bifurcation* is said to occur if a "small" change of a parameter leads to a "big" change of the behavior of a dynamical system. Since the optimal strategy of a decision maker often depends heavily on the underlying parameters, which might sometimes be, due to lack of accurate data, only roughly estimated, a *bifurcation analysis* of a system should not be omitted in order to find out how robust any found solution is against disturbances of parameter values. Looking at occurring bifurcations is also very useful when trying to find out which outcomes are possible under which premises in order to gain a more profound knowledge and understanding of the underlying system.

[8] A product created by The Mathworks, Inc.
[9] Developed by Wolfram Research.
[10] A product of Maplesoft.

A bifurcation of particular importance is the so-called *Andronov-Hopf bifurcation*, which is associated with a pair of purely imaginary eigenvalues of the Jacobian of a steady state, leading to a change of the stability of this relevant steady state and a limit cycle bifurcating from it. The Andronov-Hopf bifurcation is *supercritical* if the limit cycle is stable, *subcritical* if it is unstable. It can be shown (see Grass et al., 2008, Chap. 7.5, p. 349) that for an optimal control problem with a discount rate $r > 0$ the limit cycle can only be unstable, i.e. repelling or of saddle-type.

A very useful tool for locating bifurcations is the Matlab toolbox MATCONT (see Dhooge et al., 2003, 2006).

DNSS point: When only one steady state can be found, there is, under the assumption that the optimal solution lies on the stable manifold of a saddle point, only one possiblity where a solution can end. This changes, however, with the occurrence of multiple equilibria (which might be found by changing one or more parameter values and looking for bifurcations). Then, a particular interesting phenomenon that can occur is that of a DNSS point (or in this thesis, due to the two states, *DNSS curve* consisting of DNSS points). A DNSS point (also known as Skiba point) is both an *indifference point*[11] and a *threshold point*[12]. This means that at such a point a decision maker can choose between different solutions or strategies, respectively, leading often to completely different outcomes. For instance, in the marketing models described in this work a decision maker might, starting at such a curve be indifferent between spending much and win many consumers, or save the costs and do nothing, which would finally lead to a complete loss of customers.

Knowing that a DNSS point can only occur if there are two points found on the stable manifolds of the two different steady states, which share the same states and where the value of the Hamiltonian is equal (i.e. none of the solutions is better than the other), one can formulate a boundary value problem and solve it, which both can be automatically done by the OCMat toolbox (see Appendix B).

These points are called DNSS points in recognition of works by Dechert, Nishimura, Sethi and Skiba, who first studied the occurrence of such points for economic problems, i.e. for an optimal growth problem with a nonconvex production function (Dechert & Nishimura, 1983; Skiba, 1978) and for singular models with multiple optimal solutions (Sethi, 1977). Since then the

[11]A point is an *indifference point*, if there is more than one optimal solution path starting from it, where the solution paths differ on one point at least.

[12]A point is called a *threshold point* if there are two points in the close neighborhood, which are the starting points of two different solution paths leading to different long-run optimal solutions, i.e. steady states.

concept of DNSS points has found many applications in very different fields of economics, such as marketing, the economics of crime, or demographic problems - for more examples where DNSS points can be found see Grass et al. (2008).

Closely related to such points are the so-called *weak DNSS points*[13], which serve as some kind of *threshold* between two different solutions - yet, a decision maker does not have any choice about which optimal strategy to apply. Which optimal solution has to be chosen then is *history dependent*. This means that the optimal strategy depends more than ever on the starting point and it is not possible (at least not with the controls considered in the models) to manipulate the outcome in a way that the other strategy can be optimally applied. Such a curve might sometimes be inconvenient for a decision maker - especially when finding the starting point on the "wrong" side of the curve, e.g., in the drug context where, due to the influence of users on potential consumers, starting on one side of the curve would mean an optimal final outcome with no drug users, starting on the other side would mean an escalation of the problem with many users in the end[14]. However, one can find valuable insights about the reasons for the system behavior and the optimal control strategy by taking a closer look at such curves.

Since some optimal solutions share common states, looking at the different occurring steady states and the shape of their stable manifolds is a helpful start when trying to find a weak DNSS curve for the given parameter set.

This thesis will provide some examples for (weak) DNSS curves both in a marketing as well as a harm reduction context.

Multi-stage model: The assumption that the system always evolves continuously over time might work well for some models providing interesting insights into how an optimal solution might look like under certain conditions. Sometimes, however, a decision maker has to face sudden events (*shocks*) that might completely change the dynamical system or even his/her objectives, as well as the effect of the control on both, e.g. when a monopolist suddenly has to face competition. Of course, it becomes necessary to include the occurrence of such a shock into consideration when deciding about the optimal strategy and adapt the model accordingly. The *regimes* before and after this shock, where due to the different situations the objective function, the state (and therefore the costate) equations, the parameters can differ, are called the different *stages* of a model and the point in the state space, where the abrupt change from one stage to the other occurs, is referred to

[13]Weak DNSS points are threshold points, but not indifference points.

[14]Yet, a possible expansion of the model would be to include some instrument by which's (maybe one-time) application a decision maker could manipulate the starting point.

as *switching point*. The optimal control problem for a problem consisting of two stages (which are denoted by different subindices) then becomes

$$\begin{aligned} &\max_u \textstyle\int_0^{t_s} e^{-rt}(g_1(x(t),u(t))\mathrm{d}t + \int_{t_s}^{\infty} e^{-rt}(g_2(x(t),u(t))\mathrm{d}t \\ \text{s.t.} \qquad &\dot{x} = \begin{cases} f_1(x(t),u(t)) & \text{for } 0 \leq t \leq t_s \\ f_2(x(t),u(t)) & \text{for } t_s \leq t < \infty \end{cases} \\ &0 \leq u \leq u_{\max} \leq 1 \\ &x(0) = x_0 \end{aligned} \tag{A.9}$$

It is also possible to compare the values of the objective functions of an optimal solution by

$$\begin{aligned} &\int_0^{t_s} e^{-rt} g_1(x(t),u(t))\mathrm{d}t + \int_{t_s}^{\infty} e^{-rt} g_2(x(t),u(t))\mathrm{d}t = \\ &\frac{1}{r}(\mathcal{H}_1(x(0),u(0),\lambda(0)) + e^{-rt_s}(\mathcal{H}_2(x(t_s),u(t_s),\lambda(t_s)) - \\ &\qquad \mathcal{H}_1(x(t_s),u(t_s),\lambda(t_s)))) \end{aligned} \tag{A.10}$$

This leads to following implications for the optimal solutions:

(a) The switching time is optimally determined with the matching conditions. Then, since $\mathcal{H}_1(t_s) = \mathcal{H}_2(t_s)$ the value of the objective function is given by $V = \frac{1}{r}\mathcal{H}_1(0)$.

(b) It is never optimal to switch from stage 1 to stage 2 in finite time, i.e. $t_s \to \infty$ then $\lim_{t_s\to\infty} \mathcal{H}_1(t_s) = \mathcal{H}_2(t_s) = 0$ and $V = \frac{1}{r}\mathcal{H}_1(0)$.

(c) Another type of corner solution can occur where it is optimal to immediately switch from stage 1 to stage 2. Since consequentially $t_s = 0$, that means $e^{-rt_s} = 1$, $\mathcal{H}_1(t_s) = \mathcal{H}_1(0)$ and $\mathcal{H}_2(t_s) = \mathcal{H}_2(0)$; the value of the objective function is then given by $V = \frac{1}{r}\mathcal{H}_2(t_s)$.

Switching time and matching conditions: When considering two stage models it is not only interesting how one has to adapt the strategy for this kind of problem, but also when it is optimal to switch between those different regimes. In order to determine the optimal timing of the switch between the stages, one has to consider certain switching conditions. Since the dynamics of the models described here do not explicitly depend on the switching time and there is no salvage value, the matching conditions, first derived in Makris (2001) for infinite-horizon discounted two-stage optimal problems, are $\mathcal{H}_1(t_s) = \mathcal{H}_2(t_s)$ and $\lambda_1(t_s) = \lambda_2(t_s)$, meaning that on a switching point the value of the Hamiltonian has to be equal for both stages (if it would differ, one would of course choose the stage with the higher value of the Hamiltonian) and that the costate variable has to evolve continuously.

In some problems these conditions cannot be fulfilled. Then a decision maker would optimally choose a corner solution and would either immediately or never switch to the other stage. Yet, a switching time might be exogenously given; in that case a decision maker is forced to switch at a certain predetermined time to the other stage. E.g., when a patent expires a monopolist has to face competition.

One can calculate an optimal solution again either by formulating an initial value problem and calculate backwards in time. Yet, one has to consider either the values of the Hamiltonians, in order to optimally switch between the stages when the matching conditions are fulfilled, or the exogenously given switching time. One can also use the matching conditions as a condition for a BVP, and combine these with other conditions, such as locating a certain starting point or fulfilling certain constraints, in order to find an optimal solution of the problem.

Some remarks on the notation: In this work the time argument is omitted unless it is of particular importance. Optimality is highlighted by a asterisk and steady state values by a hat. Different stages are marked with different subindices. A letter as subindex denotes the derivation of the function with respect to the state, control, etc., e.g., $g_u = \partial g/\partial u$, except stated differently.

Appendix B

OCMat - A Toolbox Allowing the Computation of the Previous Results

OCMat is a toolbox developed in and for the MATLAB environment to provide an efficient tool to analyze optimal control problems. It is able to automatically generate files necessary for the handling of optimal control problems and provides functions to gain further knowledge about the corresponding solutions. The core of this toolbox is the formulation and solution of boundary value problems, allowing the computation of certain optimal solutions and was developed by Dieter Grass relying on techniques presented in Beyn et al. (2001); Pampel (2000); Steindl (1984) and Grass et al. (2008).

Participating in the development of this toolbox, in this thesis OCMat is used to do the numerically calculation necessary to gain the previously described results. The next sections will demonstrate, with the help of the price reduction model introduced in Chap. 2, how one can numerically analyze an optimal control model with the help of the toolbox by describing which function have to be used and give a rough picture about how they work.

B.1 Initialization of the Model

When wanting to analyze an optimal control problem with the OCMat toolbox the first step is to create a (MATLAB) file containing the necessary predetermined information about the model such as the objective function, the state equations and the control constraints. The initialization file[1] price.m[2] consists of following elements:

```
statedynamics=sym('[k-delta*x1-(a+b*x2^al)*x1*(1+beta*u1^0.75);
(a+b*x2^al)*x1*(1+beta*u1^0.75)-mu*x2]');
```

In this statement the state dynamics is symbolically described, containing the two state equations seperated by a semicolon, and therefore as a column vector. The state variables have to be denoted as `x1` (corresponding to the number of susceptibles S in the model) and `x2` (the number of users A) and the control variable as `u1`, which is the percentage reduction of the price v. The reason for this notation is that the toolbox must be able to automatically distinguish between parameters and these variables in order to allow the automatic generation of the files for as many models as possible.

```
objectivefunction=sym('x2*p*(1-u1)');
```

contains the symbolic expression for the (not yet discounted) objective function.

```
phasespace=[0 0];
```

With the help of phasespace one can determine in the current version of the toolbox in which space (state-costate space or state-costate-control space) one wants to calculate.

The reason why such an option is necessary is that it is not always possible to explicitly express the control by setting the first derivation of the Hamiltonian to zero - then it makes sense to express one of the costates by this equation, and calculate $\dot{u} = -(\mathcal{H}_{ux}\dot{x} + \mathcal{H}_{u\lambda}\dot{\lambda})/\mathcal{H}_{uu}$. This can be done automatically by the toolbox. By setting, e.g., phasespace=[0 1], one would replace the second costate by the first control variable.

The name of the variable / symbol representing the discount rate can be set by

```
discountvariable='r';
```

The purpose of the discount rate is to weight the decision makers time preferences, as such, the toolbox would then consider $V = \int_0^\infty e^{-rt} p(1-u1)\mathrm{d}t$ as objective function.

```
controlconstraint=sym('[u1-lb;ub-u1]');
```

[1]It can be found as demo at http://www.eos.tuwien.ac.at/OR/OCMat/index.html

[2]The name of the file will be used as reference to the corresponding generated files and therefore has to be chosen carefully.

Here we can find the control constraints, in this case, the first (and only) control `u1` (corresponding to v) has to be greater or equal to `lb` and smaller or equal to `ub`. The size of these and the other parameters is specified next by:

```
r=0.04; p=1; k=1; delta=0.05; a=0.02;
b=0.01; al=1.75; beta=1; mu=0.0976;
```

Having written a file containing all these elements one can start initializing the model and create the necessary files by invoking

```
>>initocmat('price')
```

The toolbox then extracts the contained information of the initfile and uses it to symbolically express the optimal control, the costate equations, the Legendre-Clebsch condition, the Lagrange multipliers, etc. These information get stored in the `ocmat/data` folder contained in a structure. By the invocation of

```
>>files4model('price')
```

the toolbox automatically creates the files necessary for later numerical calculation and moves them to the `ocmat/model/` -folder when invoking

```
>>moveocmatfiles('price')
```

B.2 Numerical Calculations with OCMat

By

```
>>m=ocmodel('price')
```

one can get the data previously stored containing information about the model and initiate an `ocmodel`-object. Input argument is the name of the model as specified earlier, and an `ocmodel`-object is returned, which is inheriting from the `optdyn`-class. An object oriented programming approach was used here in order to allow an efficient treatment of the problem attributes and functions, and will considered more closely in App. B.3.

```
>>ep=calcep(m,[1;10;5;5])
```

The function calcep serves to calculate steady states of the canonical system. Input arguments are the ocmodel-object (containing all the required necessary information) and an initial point, needed for the equation solver provided by MATLAB. Since for many numerical algorithms it is not guaranteed that MATLAB can always find a solution nor that it returns all possible outcomes - however, if the initial point is not too distant to the equilibrium the algorithm converges and a solution is found. The output argument is a dynprimitive (see B.3). The returned object contains the coordinates, an attribute describing whether a constraint is active or not and various other attributes inherited from the class octrajectory, as well as the Jacobian matrix evaluated at this point and a period attribute[3], which is zero. If desired, one can now take a look at the eigenvectors and eigenvalues of this point, which can be calculated by

```
>>[evec,eval]=eig(ep{1})
```

in order to find out more about the stability properties of the equilibrium. However, finding out whether an equilibrium point or limit cycle has saddle properties can simply be done by using the function

```
>>b=issaddle(ep{1})
```

returning 1 or 0 depending on whether the point/limit cycle has saddle properties or is totally unstable, respectively. One can then use the steady state to initialize the necessary objects and structures needed to start the calculation of a stable path by solving the corresponding boundary value problem.

Before that, however, one should make sure, that the options such as the precision of the calculations or the used boundary value solver are set as needed by the user. The OCMAT toolbox uses a structure containing options for the various calculations, which are returned by

```
>>opt=defaultocoptions
```

One can display all of the options by using

```
>>showocoptions(opt)
```

However, this input argument is not necessary if one only wants to display the default options. The option structures consists of different fields, each containing options for different problems, solvers, etc. used in the toolbox.

[3]The purpose of the class dynprimitive is to contain points, to which the optimally controlled trajectories converge, meaning not only steady states, but also limit cycles. The period then reflects the period of the limit cycle.

The option categories are OCMAT, ODE, BVP, OCCONT, and MATCONT, containing options that can be used for the MATCONT toolbox. By entering one of these categories as second input argument, the previously described function displays the options occurring there. One can change specific options by the use of the function

```
>>opt=setocoptions(opt,'BVP','AbsTol',1e-08,'RelTol',1e-08)
```

which would increase the precision of the BVP solver. Similarly it is possible to extract specific options by

```
>>bvpsolver=getocoptions(opt,'OC','BVPSolver')
```

which would return the name of the MATLAB function containing the BVP solver which will be used later on.

```
>>initstruc=initocmat('extremal',m,'initpt',[1:2],[10;0],ep{1},...
>>'ContinuationType','f','IntegrationTime',1000)
```

With this function one collects the necessary information for starting the calculation of the BVP. When invoking the function as given one wants to find a solution of the `ocmodel m` starting at an initial point with the state values `[10;0]` truncating the infinite time horizon by 1000 time units. Among other information, it also contains the initial solution `ep1`, required by

```
>>[sol vio]=occont(m,initstruc,opt)
```

This function provides an algorithm for the calculation of the optimal solution with the help of a boundary value solver. This algorithm is the core of the toolbox as it allows efficient calculations. In the actual version it was implemented by Dieter Grass using the default MATLAB BVP solver; see (Grass et al., 2008, Chap. 7). The boundary value problem is such that a solution has to start at a certain point (e.g., where the state variables have certain values, or the objective function has a certain value), has to end at a certain point (an equilibrium or limit cycle) and that it has to be a solution of the corresponding canonical system. However, to start the numerical calculations it is necessary to provide an approximate guess for the solution. To find such an approximate solution, `occont` takes an initial solution which starts and ends in the steady state (provided by the input argument `initstruc`) and then extends it into the direction of the initially specified starting point step by step, being able to solve initially "smaller" boundary value problems by using the previously calculated solution as approximate guess for the next solution, until the desired point is reached. This method and more sophisticated versions are called continuation algorithms and constitute another core part of the toolbox. A big advantage of this function is that it can also handle constraints; when one is violated, calculations will be aborted in order to allow the user to continue further calculations as desired by the user. Then the variable `vio`, which would have been empty if the desired starting point is reached without any violations, returns an `ocasymptotic` object (which handle candidates for optimal solutions) with more detailed information about the actual violation type occurred at this trajectory. By

```
>>initstruc1=initocmat('extremal',m,'initpt',[1:2],[10;0],solv,...
>>'ContinuationType','f','IntegrationTime',1000)
```

a multi-point boundary value is formulated connecting, e.g., arcs with active and inactive control constraints.(i.e. the optimal solution where the constraint is active has to start at a certain point and end where the constraint becomes inactive, which is the same point where the part of the solution with no active constraint starts, which has to end in the steady state). This problem can be solved by the same invocation of `occont`.

```
>>[sol1 vio1]=occont(m,initstruc1,opt)
```

The function `occont` returns an `ocasymptotic` object containting the solution, some information about it such as where constraints are active, which solver was used, etc. and the steady state, where the solution ends.

One can store the found results, which are at this point also locally stored[4] in a `mat`-file as a structure and assign them to the object which handles the corresponding model by invoking

```
>>m=store(m)
```

When a control constraint becomes violated it is also interesting to see at which other points of the stable manifold of a certain steady state such an event takes place. One can then use the previously found trajectory with the control violation an initial solution for solving the boundary problem where one seeks to find trajectories ending in the steady state and starting at a point where the control starts taking its boundary value. Such a boundary value problem can be initiated by

[4]Mainly to ensure that not all results are lost when it becomes necessary to interrupt or cancel the calculations.

```
>>initStruct=initoccont('boundary',m,'initpoint',[1],[0],solv,...
>>'ContinuationType','f','IntegrationTime',1000)
```

and be solved by

```
>>solb=occont(m,initStruct,opt)
```

and save the results again by

```
>>m=store(m)
```

Still it can be interesting to find optimal solutions as an initial value problem (IVP), e.g., when one wants to see the path leading to a certain point without specifying any further conditions, with the function

```
>>ocTrj=odesolve(m,-10,octrajectory(sol.dynVar(1:4,1),2),...
>>  opt)
```

It is not necessary here that the solution of the problem has to end in a steady state, the input argument used here is the first point of the previously found solution (the starting point of the trajectory). The function calculates the trajectory over ten time units back in time and it can be seen how the optimal solution develops before it reaches the previously found initial point. The function returns an `octrajectory` object.

Having found such solutions, one can take a look at various aspects of these solution, such at the optimal control values at the points of this solution by

```
>>ct=control(m,sol1)
```

or at the value of the Hamiltonian by

```
>>ha=hamiltonian(m,sol1)
```

As it can be rather hard to interpret the results simply by looking at numbers, the OCMat toolbox also provides functions to easily plot the calculated results. A phase portrait showing the previously calculated `ocasymptotic` object can be plotted by

```
>>plotphase(m,sol1,'state',2)
```

where by setting the plot option 'state' to 2 means that the second state variable will be plotted on the abscissa. Similarly it is possible to plot a timepath showing the value of the optimal control of the trajectory by

```
>>plottimepath(m,sol1,'control')
```

or to extract the calculated information from the used `ocmodel` object and plot a found solution and the curve describing where the control constraint becomes inactive by

```
>>plotphaseocresult(m,'state',2,[1],'only',{'ExtremalSolution',...
>> 'BoundaryCurve'})
```

B.3 Some Remarks on the Design and Implementation of OCMat

When implementing a toolbox, with the purpose that optimal control problems can be efficiently solved numerically, not only problems of mathematical nature arise since a large amount of information has to be handled. It is particularly important to ensure that users are able to easily access what is on the one hand relevant from a mathematician's point of view, but on the other hand also what is relevant for economic interpretation. Due to the large number of ways in which one might extend such a tool in further development one must also not neglect while designing and implementing the toolbox, that the design and the source code has to be readable and reusable for other people. It is intended that the toolbox can be used by people without a large experience in optimal control theory. Therefore, the implementation must be such that as much as possible has to be handled automatically without losing information desired by more sophisticated users. A certain traceability about what is actually done by the toolbox is also very important to ensure that the calculated results can be used for whatever purpose they are intented.

The OCMat toolbox was developed in and for the Matlab environment. With the intention that the toolbox can be used for different fields of optimal control theory, e.g., one-stage and multi-stage models, an object orientated approach, as far as supported by Matlab, was used to design the classes[5]. Since optimal control problems do have some common attributes and functions, the concept of inheritance is particularly important for these classes. Therefore, many functions can be efficiently (re)used for different types of problems with common attributes. This is also intended to enhance the expandability of the toolbox. A simple example of a common attribute of an object of a one-stage-model-class (`ocmodel`) and a multi-stage-model-class (`msmodel`) are the model parameters (inherited from the class `optdyn`). A

[5]Note that the toolbox gets continuously improved and therefore some of the attributes might be subject to change in future versions of OCMat.

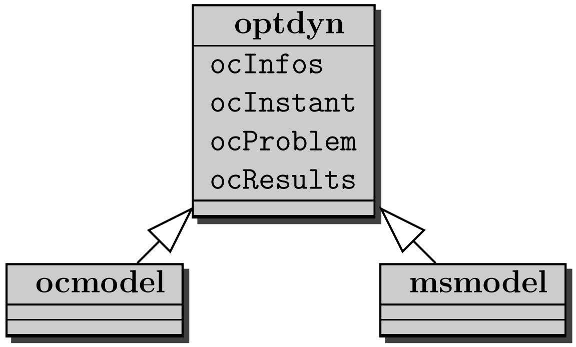

Figure B.1: Class diagram of the classes dealing with the used models, and their attributes. The corresponding functions are omitted in this diagram.

user of the toolbox might want to access the parameters and change them, no matter what kind of model one deals with. Yet, it is necessary to distinguish between the two mentioned classes, since there are some differences arising through the different stages in the representation of the model and the calculations of the results. Fig. B.1 shows a class diagram describing the attributes and the relation between the classes handling the model information.

The nature of the different numerical results also motivates using an object-oriented approach there. A class diagram for classes that are intended to handle results can be seen in Fig. B.2. At this stage of implementation there are trajectories (class `octrajectory`), i.e. candidates for solution paths, their attributes contain the most important information about the coordinates, the constraints active at these points, whether a constraint is violated at some point, at which time these points are reached, and some additional informations about and returned by the solver used to calculate this trajectory.

The next relevant class is `dynprimitive`, the purpose of which is to handle steady states and limit cycles. As such it inherits the properties of the class `octrajectory`, defining both elements as trajectories that start and end at the same points. However, in contrast to ordinary trajectories, when considering steady states and limit cycles, two additional properties become relevant: for a cycle, its period is needed (for an equilibrium it is set to zero) and for both elements the linearization of its stable manifold, found by the calculation of the Jacobian matrix of a steady state or the monodromy matrix of a limit cycle.

Using the concepts of inheritance and aggregation, an object of the class `ocasymptotic` contains a candidate for an optimal solution, i.e. a trajectory

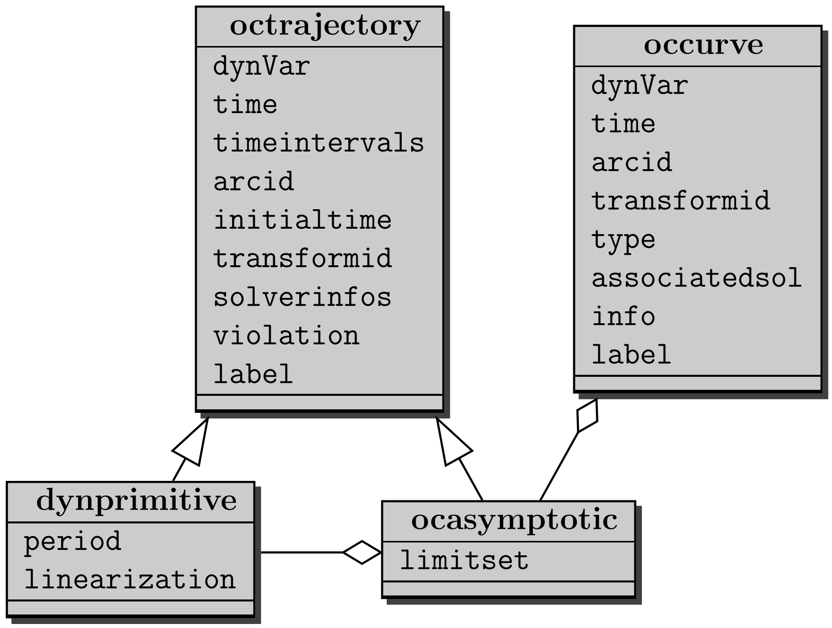

Figure B.2: Class diagram of the classes for handling different types of results, and their attributes. The corresponding functions are again omitted due to the required space.

lying on the stable manifold of a steady states, therefore it inherits its attributes from the class `octrajectory` and has the steady state to which the optimal solution converges (which is an object of the class `dynprimitive`) as an attribute.

Points, which are not lying on the same trajectory, but are related because they share some common characteristic, such as DNSS points or points where constraints become active or inactive are dealt with in the class `occurve`. An object of this class must, among other properties, contain the coordinates of the points and also an exemplary solution, belonging to the class of `ocasymptotic`, starting at this curve in order to ensure that one can easily continue calculations[6] with the previous results.

There are many ways in which one could still extend the toolbox from an informatics point of view (e.g., adding a graphical user interface, the current version is basically command-line orientated), but certainly even more from a mathematician's perspective – on the one hand one could for instance include optimal control problems with a linear control or differential games on the other hand currently the toolbox relies heavily on the solution algorithms

[6]The reason for this is that the BVP solver requires an initial solution to be able to find a solution fulfilling the boundary conditions.

provided by Matlab which are not always completely suited to adequately deal with the problem and one could try to implement more fitting algorithms.

Appendix C

Miscellaneous

C.1 Concavity of the Function Describing the Influence of the Control on Initiation Due to the Legendre-Clebsch Condition

In the marketing model it was assumed that the function describing the influence of the control on initiation $g(v)$ is concave with $g_{vv} \leq 0$. The reason for this will be shown here, but let us first consider the Lagrangian and its derivation with respect to the control again.

$$\begin{aligned}
\mathcal{L} &= \pi A(1-v) + \lambda_1(k + \delta S - f(A)Sg(v)) + \lambda_2(f(A)Sg(v) - \mu A) + \\
&\quad \nu_1 v + \nu_2(v_{\max} - v) && \text{(C.1)} \\
\mathcal{L}_v &= -\pi A + (\lambda_2 - \lambda_1)f(A)Sg_v + \nu_1 - \nu_2 = 0 && \text{(C.2)} \\
\mathcal{L}_{vv} &= (\lambda_2 - \lambda_1)f(A)Sg_{vv} \leq 0 && \text{(C.3)}
\end{aligned}$$

If no control constraints are active, the Legendre-Clebsch condition is fulfilled if $\mathcal{L}_{vv} \leq 0$. Since the number of susceptibles S and the initiation function $f(A)$ are assumed to be always equal or greater than zero that means either

(a) $\lambda_1 \geq \lambda_2$ and $g_{vv} \geq 0$ or

(b) $\lambda_1 \leq \lambda_2$ and $g_{vv} \leq 0$.

Due to Pontryagin's Maximum Principle the condition $\mathcal{L}_v = 0$ has to be fulfilled. Assuming now that no control constraint is active ($\nu_1 = \nu_2 = 0$) we find by transforming (C.2)

$$g_v = \frac{A}{(\lambda_2 - \lambda_1) f(A) S} \tag{C.4}$$

Due to the assumption that $g_v \geq 0$, we can see that (C.4) can only be fulfilled if $\lambda_2 \geq \lambda_1$, consequentially meaning that, if no control constraint is active, g_{vv} has to be non-positive.

Because of the transversality condition, the considered candidates for an optimal solution considered are found on the stable manifold of steady states. This also implies that a solution can only be optimal if the Legendre-Clebsch condition is fulfilled in the steady state (if no control constraint is active).

Looking at the costate equation concerning λ_1

$$\dot{\lambda}_1 = (r + \delta)\lambda_1 + (\lambda_1 - \lambda_2) f(A) g(v), \tag{C.5}$$

we find that in a steady state, i.e. $\dot{\lambda}_1 = 0$,

$$\hat{\lambda}_2 = \hat{\lambda}_1 \frac{r + \delta + f(\hat{A}) g(\hat{v})}{f(\hat{A}) g(\hat{v})}. \tag{C.6}$$

Due to the assumption that the parameters r and δ, as well as the functions $f(A)$ and $g(v)$ are positive (or zero), $\hat{\lambda}_2 > \hat{\lambda}_1$ since the numerator of the fraction in (C.6) is always greater than the denominator.

Bibliography

Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science*, *15*(5), 215–227.

Bass, F. M. (2004). Comments on "A new product growth for model consumer durables". *Management Science*, *50*(12), 1833–1840.

Behrens, D., Caulkins, J., Tragler, G., & Feichtinger, G. (2000). Optimal control of drug epidemics: Prevent and treat - but not at the same time? *Management Science*, *46*(3), 333–347.

Beyn, W. J., Pampel, T., & Semmler, W. (2001). Dynamic optimization and Skiba sets in economic examples. *Optimal Control Applications and Methods*, *22*(5–6), 251–280.

Caulkins, J. P. (2001). The dynamic character of drug problems. *Bulletin on Narcotics*, *53*(1), 11–23.

Caulkins, J. P. (2004). *Models Pertaining to How Drug Policy Should Vary Over the Course of a Drug Epidemic*. University of Lund, Lund, Sweden: Paper Presented at the Arne Ryde Symposium on the Economics of Substance Abuse.

Caulkins, J. P. (2005). SA model of harm reduction. Memo 80.

Caulkins, J. P. (2007). Two-state, two-stage marketing model. Memo 93.

Caulkins, J. P. & Reuter, P. (1997). Setting goals for drug policy: Harm reduction or use reduction? *Addiction*, *92*(9), 1143.

Dechert, W. D. & Nishimura, K. (1983). A complete characterization of optimal growth paths in an aggregated model with a non-concave production function. *Journal of Economic Theory*, *31*, 332–354.

Dhooge, A., Govaerts, W., & Kuznetsov, Y. A. (2003). Matcont: A MATLAB package for numerical bifurcation analysis of ODEs. *ACM Transactions on Mathematical Software*, *29*(2), 141–164.

Dhooge, A., Govaerts, W., Kuznetsov, Y. A., Mestrom, W., Riet, A. M., & Sautois, B. (2006). *MATCONT and CL_ MATCONT: Continuation toolboxes in MATLAB*. available at http://www.matcont.ugent.be/manual.pdf (last accessed 13-Feb-2009).

Etro, F. (2007). *Competition, Innovation, and Antitrust - A Theory of Market Leaders and Its Policy Implications*. Berlin: Springer.

Everingham, S. S. & Rydell, C. P. (1994). *Modeling the Demand for Cocaine*. Santa Monica CA: RAND.

Feichtinger, G. (1992). Hopf bifurcation in an advertising diffusion model. *Journal of Economic Behavior and Organization*, *17*(3), 401–411.

Feichtinger, G., Ghezzi, L. L., & Piccardi, C. (1995). Chaotic behavior in an advertising diffusion model. *International Journal of Bifurcation and Chaos*, *5*(1), 255–263.

Feichtinger, G. & Hartl, R. F. (1986). *Optimale Kontrolle ökonomischer Prozesse*. Berlin: deGruyter.

Glaeser, E. L., Sacerdote, B. I., & Scheinkman, J. A. (2002). The social multiplier. *HIER Discussion Paper*, *1968*.

Gould, J. (1970). *Microeconomic Foundations of Employment and Inflation Theory*, chapter Diffusion process and optimal advertising policy, (pp. 338–368). London: Macmillan.

Grass, D., Caulkins, J. P., Feichtinger, G., Tragler, G., & Behrens, D. A. (2008). *Optimal Control of Nonlinear Processes: With Applications in Drugs, Corruption and Terror*. Heidelberg: Springer.

Hartl, R. F. & Kort, P. M. (2005). *Optimal Control and Dynamic Games*, chapter Advertising directed towards existing and new customers, (pp. 3–18). Heidelberg: Springer.

Himmelstein, J. L. (1983). From killer weed to drop-out drug: The changing ideology of marihuana. *Contemporary Crisis*, *7*, 13–38.

Hoyer, W. D. & MacInnis, D. J. (2004). *Consumer Behavior* (3rd ed.). Boston: Houghton Miffin Company.

Kleiman, M. A. R. (1993). Enforcement swamping: A positive-feedback mechanism in rates of illicit activity. *Mathematical and Computer Modeling*, *17*(2), 65–75.

Kotler, P. (1999). *Marketing - Märkte schaffen, erobern und beherrschen*. München: Econ Verlag.

Kotler, P. & Armstrong, G. (2008). *Principles of Marketing*. Upper Saddle River, NJ: Prentice Hall.

Lenton, S. & Single, E. (1998). The definition of harm reduction. *Drug and Alcohol Review*, *17*(2), 213–220.

Léonard, D. & Long, N. V. (1992). *Optimal Control Theory and Static Optimization in Economics*. Cambridge: Cambridge University Press.

Lilien, G. L. & Kotler, P. (1983). *Marketing Decision Making - A Model-Building Approach*. New York: Harper & Row.

MacCoun, R. (1998). Toward a psychology of harm reduction. *American Psychologist*, *53*(11), 1199–1208.

MacCoun, R., Reuter, P., & Schelling, T. (1996). Assessing alternative drug control regimes. *Journal of Policy Analysis and Management*, *15*(3), 330–352.

Mahajan, V., Muller, E., & Bass, F. M. (1990). New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, *54*(1), 1–26.

Mahajan, V. & Peterson, R. A. (1978). Innovation diffusion in a dynamic potential adopter population. *Management Science*, *24*(15), 1589–1597.

Makris, M. (2001). Necessary conditions for infinite-horizon discounted two-stage optimal control problems. *Journal of Economic Dynamics & Control*, *25*(12), 1935–1950.

Motta, M. (2004). *Competition Policy - Theory and Practice*. Cambridge: Cambridge University Press.

Muller, E. (1983). Trial/awareness advertising decisions - A control problem with phase diagrams with non-stationary boundaries. *Journal of Economic Dynamics and Control*, *6*(1), 333–350.

Ofir, C. & Winer, R. S. (2002). Pricing: Economic and behavioral models. In B. A. Weitz & R. Wensley (Eds.), *Handbook of Marketing*. SAGE.

ONDP (2004). *The Economic Costs of Drug Abuse in the United States 1992-2002*. Washington: Office of National Drug Control Policy.

ONDP (2006). *National Drug Control Strategy*. Washington: Office of National Drug Control Policy.

Pampel, T. (2000). *Numerical approximation of generalized connecting orbits*. PhD thesis, University of Bielefeld.

Reuter, P. (2001). The need for dynamic models of drug markets. *Bulletin on Narcotics*, *53*(1 and 2).

Rogers, E. (2003). *Diffusion of Innovation* (5th ed.). New York: The Free Press.

Rossi, C. (2001). A mover-stayer type model for epidemics of problematic drug use. *Bulletin on Narcotics*, *53*(1), 39–64.

Seidl, A. (2005). Analysis and various applications of epidemic models. Master's thesis, TU Wien.

Sethi, S. (1977). Nearest feasible paths in optimal control problems: Theory, examples, and counterexamples. *Journal of Optimization Theory and Applications*, *23*(4), 563–579.

Skiba, A. (1978). Optimal growth with a convex-concave production function. *Econometrica*, *46*, 527–539.

Steindl, A. (1984). *Numerische Behandlung von Randwertproblemen auf unendlichen Intervallen*. PhD thesis, TU Wien.

UNODC (2006). *2006 - World Drug Report*, Volume 1. Vienna: United Nations Office on Drugs and Crime.

Wallner, D. (2005). Optimal control of the U.S. cocaine epidemic: Traditional approach vs. harm reduction. Master's thesis, TU Wien.

Wallner, D. (2008). *Dynamic Models of Drug Users and Susceptibles: Optimal Mix of Use Reduction and Harm Reduction in Australia and the U.S.A.* PhD thesis, TU Wien.

Winkler, D., Caulkins, J. P., Behrens, D. A., & Tragler, G. (2004). Estimating the relative efficiency of various forms of prevention at different stages of a drug epidemic. *Socio-Economic Planning Sciences*, *38*, 43–56.

Acknowledgments

Here I want to gratefully acknowledge the people who helped to make this thesis possible:

First, I want to express my gratitude to my supervisor Gustav Feichtinger for his continuous support and advice. I am also greatly indebted to Dieter Grass for his help and patience. Furthermore, I want to thank in particular Gernot Tragler, Jon Caulkins as well as Dieter Merkl for providing valuable comments and suggestions. I also wish to thank Dagmar Wallner, and Gerti Kappel for discussions concerning the subject of this thesis.

My deepest thanks also to my family and friends for their continuous support.

This thesis was supported by the Austrian science fund (FWF) project 18527.

Lebenslauf

Name: Mag.rer.soc.oec. Andrea Seidl

Wohnort: Wienerstr. 230, 3400 Klosterneuburg

Geburtsdatum: 02. 04. 1982

Geburtsort: Wien

Staatsangehhörigkeit: Österreich

Ausbildung & Beruflicher Werdegang

09/92 - 06/00	BG Klosterneuburg
10/00 - 04/05	Studium der Wirtschaftsinformatik an der TU Wien und der Universität Wien Diplomarbeit: *Analysis and Various Applications of Epidemic Models*
04/05 -	Doktoratsstudium der Sozial- und Wirtschaftswissenschaften an der TU Wien Studienzweig Wirtschaftsinformatik
08/06 -	Mitarbeit am Institut für Wirtschaftsmathematik, Forschungsgruppe für Operations Research und Nichtlineare Systeme