**TECHNISCHE
UNIVERSITÄT
WIEN**

**VIENNA
UNIVERSITY OF
TECHNOLOGY**

DISSERTATION

Dynamic Models of Drug Users and Susceptibles: Optimal Mix of Use Reduction and Harm Reduction in Australia and the U.S.A.

ausgeführt zum Zwecke der Erlangung des akademischen
Grades einer Doktorin der technischen Wissenschaften unter
der Leitung von

Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Gernot Tragler
E105 Institut für Wirtschaftsmathematik

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Dipl.-Ing. Dagmar Wallner
Matrikelnummer: 0025614
Deinhardsteingasse 11/8
1160 Wien

Wien, am 12. Dezember 2008

Kurzfassung

Diese Dissertation analysiert Zwei-Zustands-Modelle zur optimalen Kontrolle von illegalem Drogenkonsum. Im Unterschied zu vielen bestehenden Modellen werden Personen, die für Drogenkonsum empfänglich sind, in einem separaten Zustand betrachtet. Die zweite Zustandsvariable gibt die Anzahl der Konsumenten an. Als Kontrollinstrument wird sogenannte "harm reduction" modelliert.

Zunächst wird das Grundmodell beschrieben, bei dem es sich um ein nichtlineares, autonomes Modell mit unendlichem Planungshorizont handelt. Der Einstieg in den Drogenkonsum ist eine Funktion der beiden Zustände, wobei die Wahrscheinlichkeit, dass das zufällige Aufeinandertreffen von Drogenkonsumenten und suszeptiblen Nicht-Konsumenten zu einem Neueinstieg führt, von der momentanen Anzahl an Konsumenten abhängt. Einerseits wirkt sich die Kontrolle positiv auf die im Zielfunktional akkumulierten sozialen Kosten aus, gleichzeitig bewirkt sie aber einen erhöhten Drogeneinstieg. Das Modell wird für die Kokainepidemie in den Vereinigten Staaten von Amerika und für "injecting drug users" in Australien betrachtet. Für die Kontrollvariable gibt es Kontrollbeschränkungen. Die komparative Analyse des deskriptiven Modells unter den Randkontrollen liefert erste wichtige Erkenntnisse. Im nächsten Schritt wird die Lösung des optimalen dynamischen Kontrollproblems ermittelt. Gleichgewichtslösungen mit Randkontrolle sind dabei von signifikanter Relevanz. "Harm reduction" soll in Australien nahezu immer umgesetzt werden, hingegen ist der maximal mögliche Einsatz von "harm reduction" in den U.S.A. nur in Phasen der Epidemie mit entweder sehr wenigen oder sehr vielen Drogenkonsumenten optimal. In den Bereichen dazwischen ist "pure use reduction" optimal.

Der zweite Teil der Dissertation präsentiert die optimalen Lösungen für einige Abwandlungen des Basismodelles. Zunächst wird angenommen, dass die Kokainepidemie in den USA weniger ansteckend ist. In einer weiteren Modifikation wird die Funktion für den Drogeneinstieg um sogenannte Innovatoren ergänzt. Die dritte Abwandlung besteht darin, in der Zielfunktion die für die Kontrolle anfallenden Kosten zu berücksichtigen. Sowohl für Australien als auch die USA werden mehrere verschiedene Szenarien betrachtet und miteinander verglichen. Abschließend wird die funktionale Form für den Drogeneinstieg verändert, allerdings erscheint die benutzte logistische Funktion nur für Australien sinnvoll.

Das Pontryagin'sche Maximumprinzip ist das geeignete Verfahren aus der optimalen Kontrolltheorie, um die vorliegenden Modelle zu lösen. Hierbei treten in fast allen Fällen Gleichgewichtspunkte mit Randkontrolle auf. In den Einzugsgebieten der Gleichgewichte existieren sogenannte Indifferenz-Kurven oder DNSS-Kurven (benannt nach Dechert, Nishimura, Sethi und Skiba). Entlang solcher Kurven ist ein Entscheidungsträger indifferent zwischen zwei Pfaden, die auf verschiedene Weise zu einem optimalen Gleichgewicht führen, beziehungsweise zwischen zwei Pfaden, die zu unterschiedlichen optimalen Gleichgewichten führen.

Abstract

This thesis investigates two-state optimal control models of illicit drug use. In contrast to most existing models, we explicitly consider people who are susceptible to drug use. This subpopulation of susceptibles is modeled as one state, with the current users of the drug making up the second state. The single control instrument is so-called "harm reduction".

The first part of the thesis introduces the base model, which is non-linear, autonomous, and of infinite time horizon. Initiation into drug use is driven by random mixing of users and susceptible non-users. The probability of adoption depends on the current size of the pool of users. Furthermore, harm reduction has the beneficial effect of reducing social costs, while at the same time it increases initiation. The model is parameterized both for the U.S. cocaine epidemic and for injecting drug use in Australia. For the control variable, there are boundary constraints. A static comparative analysis between control set equal to zero and control at its upper bound is conducted to derive first important insights. Next, the optimal dynamic control model is assessed. Boundary control solutions turn out to be of significant relevance. Harm reduction should be applied almost always in Australia, whereas for the U.S. cocaine epidemic it is optimal only at the very early stages of the epidemic and at high levels of use. At intermediate levels of cocaine use, the policy suggestion for the U.S. is not to apply harm reduction.

The second part of the thesis discusses various modifications of the model. First, the case of a less intense infectivity of the U.S. cocaine epidemic is analyzed. Second, innovators are introduced to the initiation function. Third, a linear cost term that penalizes control spending is introduced in the objective function. For both the U.S. and Australia, various scenarios are analyzed and compared with each other. The last modification makes initiation a logistic function of prevalence of injecting drug use in Australia.

The optimal control models are solved by applying Pontryagin's Maximum Principle. The optimal steady states are solutions with boundary control. In several scenarios, indifference curves and DNSS curves (named after Dechert-Nishimura-Sethi-Skiba) show up in the phase portrait. These curves consist of points in the state space at which a decision maker is indifferent between two different paths that lead to the same optimal steady state (indifference curve) or at which a decision maker can choose between two optimal steady states (DNSS curve), respectively.

Contents

Chapter 1

Introduction

Consumption of illicit unhealthy substances poses serious challenges for societies and decision makers all over the world. Policy makers are interested in ameliorating problems that stem from drug use. Consequently, there is a wide range of control interventions. Among the most traditional there are prevention, treatment, and various forms of enforcement. In the past and present, analysis of the optimal dynamic application of these interventions has been subject of abundant work in the field of optimal dynamic control of drug use (see e.g., Behrens et al., 2000; Tragler et al., 2001; Zeiler, 2007). The vast majority of prior studies on optimal control of drug epidemics emphasizes on numbers of users. Some of the models recognize that drug users progress through different levels of drug use. Amongst the models of drug epidemics that provide pioneering results we find the famous two-state LH-model (cf. Behrens et al., 1999; Knoll and Zuba, 2004). The L state models "light users" who consume a certain drug, but are not yet dependent, whereas the state variable H denotes "heavy users" who are dependent of the drug. In Zeiler (2007), the situation of a decision maker facing two qualitatively similar, interacting drug epidemics is modeled.

In marketing, the group of susceptible non-users plays an important role. Bass (1969) introduced his classic model of product diffusion via consumer contact. It posits that the rate of adoption by susceptible non-users is linearly increasing in the number of current product users. The role of susceptibles as potential users of a certain good is well known and recognized in marketing. Drug models and marketing models have many parallels, because the spread of drugs is a diffusion process. Nevertheless, only very few studies in the field of drug control optimization pay attention to those susceptible to drug use (an exception being provided by Almeder et al., 2001, 2004).

Concerning the question how to fight a drug problem in a society, a heated debate in drug policy has been going on for the last years. It concerns the relative merits of "use reduction" versus "harm reduction". Basically, the term use reduction refers to the control of drug use per se, whereas harm reduction focuses on the reduction of harms that arise due to drug consumption, e.g., drug-related HIV/AIDS transmission, transmission of Hepatitis C, health costs borne by drug users, or crime and violence. Complicating matters in the debate, when different people use the term "harm reduction" they often mean different things. In the past, there has been another vigorous and equally contentious debate between those placing primacy on "demand reduction" and the proponents of "supply control". Efficiency analyses (see e.g., Winkler et al., 2004) and optimal control models (Behrens et al., 1999, 2000; Tragler et al., 2001) suggested a peaceful solution to the bitterly opponent sides. The optimal control solutions revealed that each kind of control mechanism had a decisive role to play but the relative effectiveness varied over the course of the drug epidemic. The excited discussion between harm reduction and use reduction proponents makes it interesting to explore the possibility of a similar resolution for this case.

The model assessed in this thesis is revolutionary in two respects. Inspired by the discussion of harm reduction versus use reduction the control instrument in our model is harm reduction. With respect to the states, the model does no longer focus on numbers of users only, but also considers people who are susceptible to drug consumption. The users of a particular drug are lumped together into one state. It is the non-using remainder of the population where the second state emerges from. The non-users are split up into people who are not using that drug, but who are vulnerable to do so, and a large remainder of non-using non-susceptible persons. Following the concept of a drug epidemic with infectivity that may vary over the course of the epidemic, the probability that the contact between users and susceptible non-users leads to initiation depends on the number of current users.

The goal of this thesis is the analysis of the optimal dynamic application of harm reduction interventions and use reduction measures over the course of a drug epidemic, applying the tools of optimal control theory to a simple, dynamic, two-state, one-control model of drug use. The classic interventions of the use reduction framework (treatment, price-raising law enforcement, and prevention) are not modeled explicitly, but nevertheless they are an implicit part of the model, because parameterizations stem from data affected by the classic drug control policies. Thus setting the harm reduction control equal to zero represents the traditional policy of pure use reduction with application of the classic controls.

This thesis is structured as follows:

Chapter 2 presents the definition of harm reduction underlying the model. Furthermore, the mathematical formulation of the model is presented, the functional forms are specified, and the parameterization of the model for the U.S. cocaine epidemic and Australia's population of injecting drug users (mostly abbreviated by IDU in what follows) is presented.

Before analyzing an optimal control model, it is often useful to examine the behavior of the underlying uncontrolled (purely descriptive) model first. Chapter 3 presents a more advanced static analysis. It contrasts results from the system without control and with the maximum possible control. Results suggest that full harm reduction may yield great benefits. Depending on the drug, the country, and the stage of the epidemic, however, harm reduction also bears risks. In particular, the simple static analysis suggests that harm reduction may have always been advantageous for Australia's IDU problem, while for the cocaine epidemic in the U.S. it can be highly disadvantageous.

The optimal dynamic control version of the model under the base case parameterizations is presented in Chapter 4. In the base case, the optimal long-run steady states always have boundary control values. Trajectories leading to those steady states will typically involve several switches between boundary and interior control. The results obtained provide interesting interpretations for policy makers. Even for U.S. cocaine, where harm reduction is so bitterly assailed and demonized, harm reduction turns out to play a potentially helpful role.

Chapter 5 is devoted to the important issues of sensitivity and bifurcation analyses.

Some variations of the model are presented in Chapter 6. First, the case of increased infectivity of the U.S. cocaine epidemic is analyzed. Second, innovators are introduced to the initiation function. Third, a linear cost term is added to the objective function in order to penalize control spending. Various scenarios are analyzed and compared with each other for both the U.S. and Australia. The last variation modifies the initiation function and uses a logistic approach to model the feedback of current use.

The Appendix presents some technicalities of the model. Most results cannot be computed analytically. Therefore numerical methods have to be employed to obtain the solutions of the system. Information on the software used is given in Appendix A.4.

Chapter 2

The Model

2.1   Literature Review on Harm Reduction

MacCoun (1998) briefly introduces the issue of harm reduction with the following lines: "During the 1980s, a grassroots movement called harm reduction emerged in Amsterdam, Rotterdam, and Liverpool [. . .]. The movement gradually spread to many other European cities, eventually influencing the policies of several nations [. . .]. Harm reduction is not yet a well-developed approach. Rather, it is a set of programs that share certain public health goals and assumptions. Central among them is the belief that it is possible to modify the behavior of drug users, and the conditions in which they use, in order to reduce many of the most serious risks that drugs pose to public health and safety. Examples of specific harm reduction interventions for drug use include needle and syringe exchange, low-threshold methadone maintenance, "safe use" educational campaigns, and the use of treatment as an alternative to incarceration for convicted drug offenders."

Australia is a country that has a long history of harm reduction interventions and thus represents a concrete example. Monograph Nr. 6 (Ritter & Cameron, 2005) by Australia's Drug Policy Modelling Program (DPMP) sheds light on their understanding of the term harm reduction: "Harm reduction was defined as policies and interventions that focus on reducing the harms associated with drug use, not the amount of drugs used. The following interventions were reviewed: needle syringe programs; supervised injecting facilities; non-injecting routes of administration; outreach; HIV education and information and HIV testing and counselling; brief interventions (aimed at harm reduction); overdose prevention interventions and legal and regulatory frameworks." Furthermore, the authors explain: "The definition chosen

9

for this review was programs or policies that aimed to reduce harm but not reduce drug use per se; and do not have as a primary mechanism of action, the reduction of drug use. By way of example, needle syringe programs (NSP) are included because the goal of NSP is to reduce the harm associated with injecting. Methadone maintenance has been excluded from the review because the primary mechanism by which it achieves reductions in harm is through reductions in use. With outreach, or brief interventions, the focus is on interventions that reduce the risk of harm associated with injecting (overdose, blood borne viruses [BBV]) rather than interventions for reduction of drug use itself."

From the fact that MacCoun includes methadone maintenance treatment, whereas Ritter & Cameron exclude such interventions, we infer that depending on the particular background and specific attitudes different people may have different perceptions of harm reduction.

An example where harm reduction's definition and the outlook on future harm reduction interventions are targeted to specific needs of a certain region and its prevalent problems is Azim et al. (2005). They report on the effectiveness of harm reduction programs for IDUs in Dhaka city and describe "The services provided include needle/syringe exchange in the field, DIC based clinical services for the management of abscesses and sexually transmitted infections (STIs), drug detoxification camps, condom distribution, education and awareness on the harmful effects of drugs, safe injections, HIV/AIDS, STIs, other blood borne infections." This sounds similar to the Australian understanding, but the results of the study suggest riskier sexual behavior in IDUs who are involved in commercial sex, which seems less of a problem in Australia. Further, there are neighborhood differences. Thus, the article concludes that "the harm reduction programme has to develop a more comprehensive response" and that "consideration of structural interventions may be necessary if an HIV epidemic is to be averted."

The third interesting example out of the literature review is the harm reduction approach by the Chicago Recovery Alliance (CRA). Information on this program was retrieved online only. Their Harm Reduction Protocol (Scavuzzo (editor), 1996) sheds light on very important points and states the vitally important fact that harm reduction is different from classic drug control approaches in the U.S. Indeed, harm reduction rather stands for a change of the attitude towards drug use taking into account the most individual needs of drug users. Their definition is: "Harm Reduction is anything that reduces the risk of injury whether or not the individual is able to abstain from the risky behavior. Inherently, it is a staged form of behavioral change, which is consistent with all the prevalent models of sexual

and drug use behavior change and all the models of behavior change in these areas that have been shown to have some benefit."

Pointing out to the ongoing contentious discussion between proponents of use reduction and harm reduction advocates, it is outlined, that the first priority of harm reduction is a decrease in the negative consequences of drug use. But by contrast, North America's drug policy has traditionally focused on the reduction of the prevalence of use. Harm reduction establishes a hierarchy of goals. Amongst those, the more immediate and realistic ones are chosen to be achieved as first steps toward risk-free use or, if appropriate, abstinence. Needle exchange is mentioned as a form of harm reduction directed to the enormous risks injecting drug users face, primarily stemming from the sharing of needles. A crucial point is that needle exchange programs recognize that many IDUs are unable or unwilling to stop injecting, and that the goal of interventions must be reducing the risk of an HIV infection. The strategy is based on knowledge and means approach to change in behavior. The individuals are provided with information about the changes that are needed. Moreover, they are also provided with the means to make these changes. In the case of IDU, those means are sterile needles, syringes, other instruments for administering drugs, and condoms.

Furthermore, the issue of respect and trust is addressed. The Harm Reduction Protocol states that the essence of harm reduction is providing education in a respectful, non-judgemental and clearly understandable way. In a simple definition, harm reduction can be understood as "any positive change", which is eponymous for the entire CRA initiative. Positive change is relative to the individual, which is subject to a myriad of circumstances, and therefore needs to be determined by the individual. The CRA's Harm Reduction Advocates work with the individual to determine what changes are desired and to figure out goals and changes that are indeed realistic. As may be expected, the Advocates state that this task is usually an ambitious one. The Advocates report that individuals often arrive with arguments like "I want to stop shooting", "I plan quitting" or "I am interested in a methadone program", when in reality they are seeking clean needles or other sterile equipment. Their experience is that getting beyond this stage is critical. There are no requirements for participation in harm reduction and related activities. Individuals are informed that they can choose the avenues for change, when and if they want to. This is a form of self-empowerment allowing users to make the choice(s) in how they want to design their personal harm reduction efforts. This approach often creates a trusting, respectful relationship that allows for a positive harm reduction interaction. Beyond this, once IDUs understand how the concept of harm reduction works, they

will usually try to find out what other services are available.

Harm reduction tactics designed by the CRA are clearly focused on users. A harm reduction approach that also takes into account harms felt by non-users is prevalent in Canada. The Canadian government launched Canada's first National Drug Strategy (CDS) to address substance abuse problems. The strategy has four pillars providing a balance of measures to reduce the demand for and supply of illicit substances. One of them is harm reduction (Health Canada, 2003). According to Riley (1993), the CDS defined harm as "sickness, death, social misery, crime, violence and economic cost to all levels of government." The description is very general, but crime and violence usually account for harms that are felt by the community of non-users, which makes it quite different from the definitions we retrieved before.

Those extracts from literature show that definitions (which types of interventions are classified as reducing harm) and focuses (users vs. non-users) vary and that IDU populations and more generally drug users can have very specific problems, needs or behaviors. In the next sections the understanding of harm reduction underlying the current work is explained and the role of harm reduction in the model is discussed.

2.2 Harm Reduction versus Use Reduction

For the current model a definition of harm reduction has to be specified. Mac-Coun (1998) gives a definition of harm reduction that is essentially equal to minimization of social costs. He describes this as "macro harm reduction", which has to be distinguished from "micro harm reduction" standing for minimization of how harmful drugs are. The distinction between the aggregate level and the per-unit-level is captured by the simple equation

$$\text{Total Harm} = \text{Total Use} \cdot \text{Average Harm Per Unit of Use.} \qquad (2.1)$$

In the terminology of MacCoun macro harm reduction means the reduction of total harm. Looking at equation (2.1) a reduction of total harm can be achieved by either reducing the factor *Total Use* or by reduction of the *Average Harm Per Unit of Use*. The first mechanism reflects the idea of pure use reduction, where total harm is reduced by cutting down total use. Thereby, harmfulness is viewed as something that cannot change. Moreover, opponents of harm reduction measures worry that such interventions might induce higher levels of use. Thus, they shy away from such programs. Critics of use reduction usually bring in the counter-argument that it is not possible

to completely suppress drug use in a free society. Furthermore, they criticize that efforts to suppress use usually displace it into even more harmful forms. The classic example in that context is that prohibition of syringe possession leads injecting drug users to share and reuse their syringes, which increases the risk of infections and exacerbates the spread of blood borne diseases, notably HIV/AIDS and Hepatitis C. Harm reduction proponents suggest, e.g., that use can be made safer by providing supervised injection facilities (SIFs) where overdoses can be quickly detected and treated. They don't believe that harm reduction might induce more young people to try drugs, increase drug use in general, or literally "send the wrong message".

Minimization of total harm is indeed the objective in our model. When speaking about harm reduction in the sense of the control variable $v(t)$ of our model the reduction of the average harmfulness of a drug per unit of use is addressed.

Focusing on the question whose harms should get counted, there is again considerable discrepancy between harm reduction proponents and use reduction advocates. Many harm reduction advocates focus on harms felt and suffered by users exclusively, with special attention paid to the risk of overdoses and blood borne diseases like HIV/AIDS. Use reduction proponents rather tend to consider the harms felt by the non-using community, notably drug-related crime and violence.

It is not surprising that the U.S. differ from Australia and some European countries regarding their harm and use reduction objectives. Social costs and public concern about illicit drugs in the U.S. are dominated by crime and violence. The considerable share of two thirds of social costs stems from a single drug: cocaine (including crack). Cocaine is usually not injected in the U.S. In most other developed countries, violence is much less of a concern and the drug that dominates use is heroin, which is usually injected (UNODC, 2005). Injecting drug users face a high risk of infections when reusing or sharing their devices for administration, in the worst case they are infected with blood borne diseases. Coutinho (1998) reports on HIV and Hepatitis C among injecting drug users and concludes that "the low incidence (and prevalence) of HIV among injecting drug users in Australia may be ascribed to that country's public health approach, with wide implementation of preventive measures including needle and syringe programmes." Furthermore, heroin has a particularly high risk of overdose because of its low "safety ratio". Gable (2004) computed the "safety ratio" for several substances by comparing its reported acute lethal dose with the usual recreational dose. One of the key findings of the study is that "intravenous heroin appeared to have the greatest direct physiological toxicity". This implies that for the

U.S., where the most problematic drug is not administered by injection, harm reduction might probably play a less important role.

2.3   Risk Compensation

Considering equation (2.1) and assuming that harm reduction has no effect on levels of use, it is trivial to conclude that reducing harmfulness would reduce total harm. The crucial point is, that there is evidence for the possibility that people tend to participate more frequently in a risky activity if that activity is made safer. MacCoun (1998) discusses psychological and empirical evidence for the possibility of so-called "risk compensation" in various domains. He cites examples from literature that document that car drivers have responded to seat belts and other safety improvements in cars by driving more recklessly and faster, and that low-tar tobacco reduces the harmfulness per unit of tobacco, but yet numerous studies demonstrated that smokers compensate by smoking more cigarettes, inhaling more deeply, or blocking the filter vents.

Measures like improvements in car safety or low-tar tobacco aim at reduction of adverse consequences of certain behaviors. Nevertheless, increased participation may offset the per-unit reduction and lead to increases of harmfulness at the macro level. Another example for an idea that intends a beneficial effect, but in reality does not lead to a reduction of harms, are so-called difficult-to-reuse syringes. Sharing of syringes by injection drug users is a principal means by which HIV is spread. Some have suggested that distributing syringes that are difficult to reuse (DTR) would slow the spread of HIV. Caulkins et al. (1998) developed a simple mathematical model which describes how changes in the numbers of DTR syringes or regular syringes consumed over the course of a fixed number of injections affect the proportion of injections that are potentially infectious and, thus, the transmission of HIV. The authors find that increasing the number of either type of syringe will reduce the number of potentially infectious injections, but that, per syringe added, the reduction is always greater if a regular rather than a DTR syringe is added. Similarly, introduction of a certain number of DTR syringes and simultaneous reduction of regular syringes by the same number will increase, not decrease, the proportion of infectious injections. Given the fact that DTR syringes are more expensive than regular syringes, the authors conclude that there is little justification for substituting DTR syringes by regular syringes.

The term "moral hazard" describes a phenomenon similar to risk compensation. Moral hazard describes the possibility that an individual insulated

from certain risks may behave differently from the way he or she would behave if fully exposed to that risk. When a person (or institution) does not bear the full consequences of his/her actions, and therefore develops a tendency to act less carefully than it would otherwise, and thereby a third party is left to bear the responsibility for the consequences of the behavior, moral hazard occurs. As an example, one can imagine a person with insurance against automobile theft. The negative consequences of automobile theft are (partially) borne by the insurance company, consequently the person may care less about locking his/her car.

The essence of the discussion of risk compensation, moral hazard, and DTR syringes is that measures that aim at benefits need not necessarily be beneficial. From the principle that making an activity safer generally increases participation in that activity we can conclude ambiguous effects of harm reduction on total harm. Imagine harm reduction is done, meaning harm per unit of consumption is cut down. Via a risk compensation effect, total use increases due to pushed initiation and/or because current users increase their units of consumption. If the increase in use was high enough, total harm could go up even if the harmfulness per unit went down. However, MacCoun's conclusion on the topic's treatise is that there is "little evidence that behavioral responses produce net increases in harm [...]. Instead, most studies find that when programs reduce the probability of harm given unsafe conduct, any increases in the probability of that conduct are slight, reducing, but not eliminating the gains in safety [...]. As a result, in our terms, micro harm reduction produces macro harm reduction."

Risk compensation is considered in the current model by assuming that application of harm reduction measures reduces the risk of drug consumption and thus induces more people to try drugs. Consistent with MacCoun's concluding finding, the direct adverse effect of harm reduction is always less than the benefit resulting from reduced harm per unit of use. What remains to be explored is whether and how non-linear feedback effects in the dynamics of drug users affect total harm.

2.4   State Dynamics

In Wallner (2005), the so-called "SA model of Harm Reduction" inspired by Caulkins (2005) was stated in a general form. The analysis there is limited to a simple one-state model, and only basic features are explored. In this thesis, a more sophisticated analysis is conducted. The two states tracked over time are the number of users of a particular drug and people who are

not consuming that drug, but who are susceptible to start doing so. They are captured by the state variables $A(t)$ and $S(t)$, respectively.

The flows in the system are schematically described in the flow diagram in Figure 2.1.
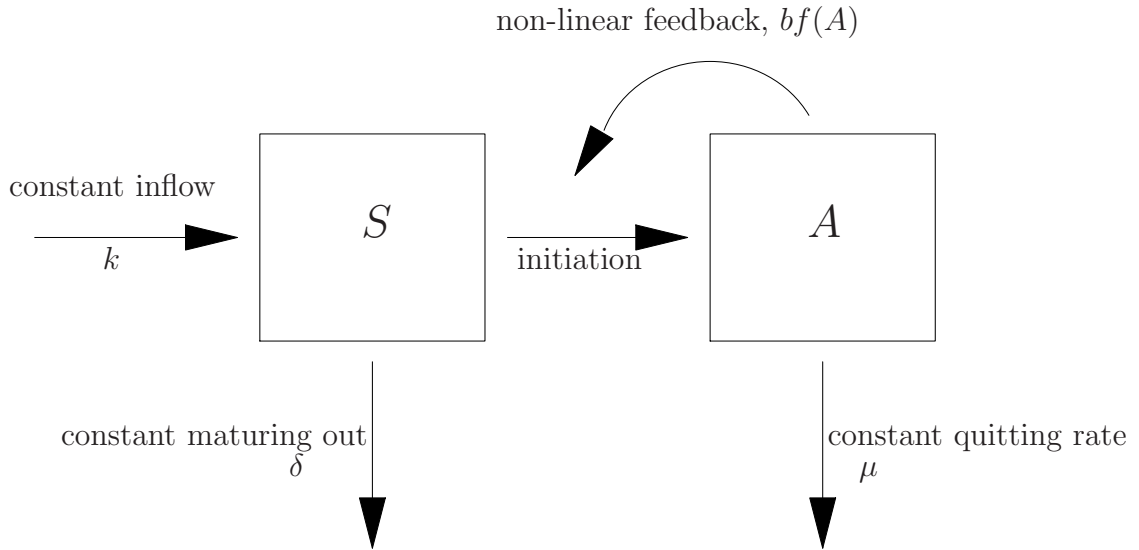
Figure 2.1: Flow diagram of the modeled system

2.4.1 Drug Users

The analysis in this thesis is twofold in the sense that the model is run with parameterizations stemming from two different countries. In particular, the drug use states $A(t)$ we deal with are cocaine users in the U.S. and injecting drug users in Australia. Within the use state, we do not take into account frequency of use, quantity consumed or degree of addiction.

Users can quit consumption. Reasons for such an exit may be death, ceasing use due to successful participation in a treatment program or desistance motivated by economic or other reasons. The model does not distinguish the particular reasons why people quit, they are all lumped together. This single outflow from the pool of users is modeled via a constant per capita rate, μ. Note that the fraction $\frac{1}{\mu}$ gives the average length of a career of drug use.

2.4.2 The Susceptible Individuals

The pool of susceptibles, $S(t)$, is fed by a constant inflow, k. This simple approach can be corroborated by the fact that most people who start using drugs are teenagers or young adults. Hence, people that reach the age when susceptibility to drug use starts can be seen as the inflow to the group of susceptibles. To be concrete, this inflow could be made up by all young people turning age 12.

After some ten or twenty years, people reach ages when they will no longer be vulnerable to drug use. This "maturing out" is governed by the constant per capita outflow rate δ. The outflow rate is roughly equal to one over the usual dwell time in the pool of people who are likely to consider starting drug consumption. Above we alluded to a duration of susceptibility between ten and twenty years, thus we expect δ to be in the range between 0.05 and 0.1.

2.4.3 Initiation

The link between the two states is initiation into drug use. It is a flow from the pool of susceptibles $S(t)$ to drug use $A(t)$ modeled by the so-called initiation function. The general initiation function reads

$$\mathrm{I}(A(t), S(t), v(t)) = \Big(\tau + b\,f\big(A(t)\big)\Big)\,S(t)\,g\big(v(t)\big). \tag{2.2}$$

Interaction between non-users $S(t)$ and users $A(t)$ leads to new infections over the course of the drug epidemic. A certain proportion of susceptibles $S(t)$ is "recruited" per unit time. That proportion is influenced by the current size of the community of users, which is modeled by the function $f\big(A(t)\big)$. The dependence on $A(t)$ is grounded on the perception of "drug epidemics". Of course, there is no pathogen that spreads drug use similar to the way that a pathogen spread the flu, malaria or HIV. However, Noymer (2001) finds that drug use is contagious in the same way fashions, laughter and even rumors can be. Indeed, almost everyone who engages in illicit drug consumption is introduced by a friend, sibling or acquaintance who is already using. The story of drug dealers seducing unsuspecting, naive persons is no more than a myth. What happens with drug use is no general spread in the sense of diffusion, but a process grounded on social interaction between current users and current non-users. One concrete approach for such an epidemic concept is the functional form $f\big(A(t)\big) = A(t)^\alpha$ chosen for the base case model. A so-called logistic approach is briefly discussed later on in this section, and a concrete example is investigated in the last section of Chapter 6.

Consistent with the idea of risk compensation discussed in section 2.3, the function $g(v(t))$ models effects of harm reduction $v(t)$ on initiation. Section 2.5.2 is devoted to the detailed discussion of this important function $g(v(t))$.

With respect to the reasons why susceptible non-users $S(t)$ start to consume drugs, the initiation function distinguishes two groups of people. The vast majority of new users enter the drug use state due to the contagious spread done by current users. In the jargon of product diffusion models, the term $b\,f(A(t))$ gives the share of susceptibles that start a career of drug use as so-called imitators. The parameter b is a proportionality constant that links the units of the number of users modified by the power function to the probability that the contact between a user and a susceptible non-user leads to an "infection". The smaller share of new users, $\tau S(t)$, is called innovators, because they are not recruited via such social interaction effects. They start to use drugs due to intrinsic interest. The parameter τ is the coefficient of innovation.

A basic assumption on initiation is that it is increasing in the number of users. The most unpretentious functions that fit that need are power functions $f(A(t)) = A(t)^\alpha$ with $\alpha > 0$. This leads to the formulation of the initiation function

$$\mathrm{I}(A(t), S(t), v(t)) = \left(\tau + b\,A(t)^\alpha\right) S(t)\, g(v(t)), \qquad (2.3)$$

that is underlying the lion's share of the investigations in this thesis. Depending on the exponent α, the initiation function will be concave ($\alpha < 1$) or convex ($\alpha > 1$) in the number of users. This suits our needs in drug epidemics with respect to the following two concepts of feedback effects from current users to initiation.

An individual who is offered an illicit drug for the first time might be less likely to accept the offer if he or she does not know much about the drug and/or if only a small, maybe even highly atypical population uses the drug. The person may be more likely to accept the offer knowing that some of his/her friends are already using it, and even more likely to engage in consumption of a drug that almost all of his/her peers use. Hence, we might expect convexity with respect to $A(t)$ at least for some illicit drugs.

Another concept of a drug epidemic observes the so-called Musto effect. Most of the severe adverse consequences of drug use do not manifest when there are only some users. They turn out later in the epidemic, when some people have used the drug for an extended period and when use is relatively widespread (Musto, 1987). People who have progressed to dependent use and people who manifest destructive consequences of drug use act as a brake on
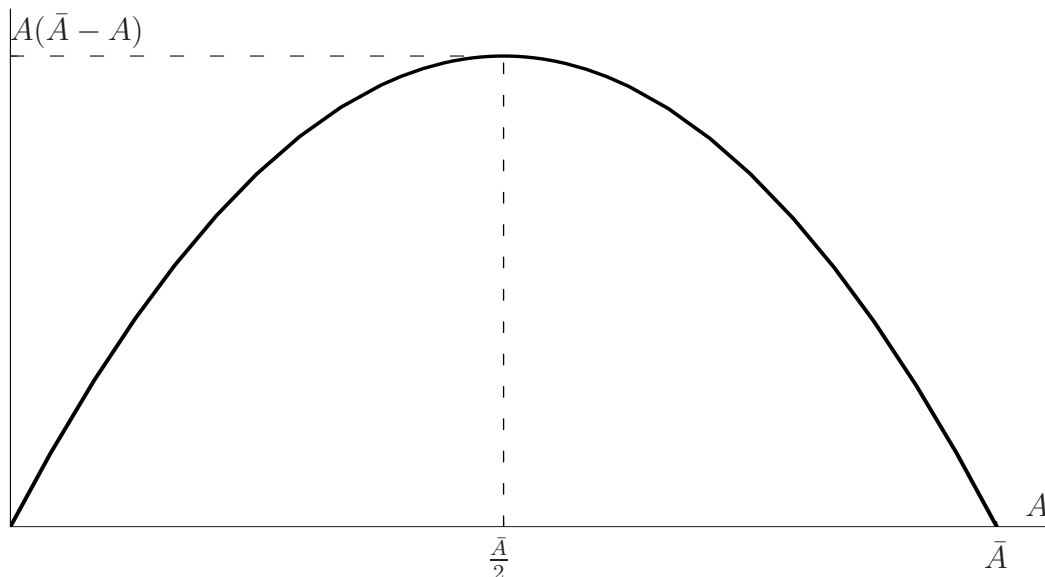
Figure 2.2: Logistic function $f(A) = A\left(\bar{A} - A\right)$

initiation. The virulence of the epidemic then develops less than proportional to the number of users. The idea of saturation takes the same line, but works as an entirely separate argument. When use is relatively widespread, non-users are offered the drug in multiple occasions, and people who did not accept earlier offers will most probably also reject future offers. This means that at the later stages of the epidemic, an increase in users $A(t)$ leads to a less than proportionate increase in initiation. Thus, an initiation function concave in $A(t)$ may be appropriate to model the spread of certain drugs.

Please note that initiation functions with power functions $A(t)^{\alpha}$ and functions increasing in $A(t)$ are not the single possibility to model initiation. One could assume that manifestation of negative consequences of drug use does not only slow down contagion, but that initiation declines in the number of users when use is high. Then, a logistic approach with $f\left(A(t)\right) = A(t)\left(\bar{A} - A(t)\right)$ is appropriate. With this function, a logistic initiation function could read

$$\mathrm{I}_{\log}(A(t), S(t), v(t)) = \left(\tau + b\,A(t)\,(\bar{A} - A(t))\right) S(t)\, g\left(v(t)\right). \qquad (2.4)$$

$\bar{A}$ denotes the carrying capacity of the model. The number of users cannot grow beyond that critical value, because then initiation is negative. As sketched in Figure 2.2, the number of infections is increasing in $A(t)$ as long as $A(t) < \frac{\bar{A}}{2}$. But for use growing beyond $A(t) = \frac{\bar{A}}{2}$, the contagious effect decreases.

Aiming at construction of a realistic model the feasible domains of the initiation functions have to be determined. Initiation is the flow from the pool of susceptibles to the drug use state, negative numbers of initiates are not sensible. Negative numbers of drug users $A(t)$ and/or susceptibles $S(t)$ do not occur, hence we only have to concentrate on the first quadrant of the (A,S)-plane, i.e. we have to check if $\mathrm{I}(A(t), S(t), v(t)) \geq 0$ for $A(t) \geq 0$, $S(t) \geq 0$ and the actual parameters. Areas where initiation is negative have to be excluded from the feasible domain. For the functional form $\mathrm{I}(A(t), S(t), v(t)) = \left(\tau + b\, A(t)^{\alpha} \right) S(t) g\left(v(t) \right)$ with parameters $\tau \geq 0$ and $b > 0$, non-negativity in the first quadrant is assured. For the logistic initiation function introduced in equation (2.4) the feasible domain is restricted to $A(t) \leq \bar{A}$.

This thesis elaborates on the spread of a "bad" fashion, of a substance that triggers harms and social costs to society. In marketing, many product diffusion models have been explored. These models focus on the adoption of a consumer good, whose consumption is usually beneficial to the buyer and/or decision maker, by people who are not yet buying that product. In the original model by Bass (1969), the number of new adoptions is given by the stock of current users $A(t)$, multiplied with the non-using remainder of the population, $1 - A(t)$, and a proportionality constant b. It gives the probability that the random mixing between users and non-users leads to an adoption. Hence, the number of new adoptions is $\left(\tau + b\, A(t) \right) \left(1 - A(t) \right)$. Notable among the extensions of the Bass model are the non-uniform influence models by Easingwood et al. (1983). These models allow for a nonlinear effect from current buyers on the non-buying population, which modifies the adoption term to $\left(\tau + b\, A(t)^{\delta} \right) \left(1 - A(t) \right)$. The current model tracks two separate states, the number of drug users, $A(t)$, and those who are susceptible to start consuming that drug, $S(t)$. To the best of my knowledge, the present model is the first time that an adoption decision is described as the result of random interaction between users $A(t)$ and susceptible non-users $S(t)$ with nonlinear feedback from the pool of current users.

Note that the states $A(t)$ and $S(t)$ are given in numbers of millions throughout the whole thesis.

To conclude this section, we summarize the above discussed flows and write down the system dynamics in general form, in which the time argument of the states and the control variable is omitted:

$$
\begin{aligned}
\dot{A} &= \mathrm{I}(A, S, v) - \mu\, A & (2.5)\\
\dot{S} &= k - \delta\, S - \mathrm{I}(A, S, v). & (2.6)
\end{aligned}
$$

2.5 The Control Variable

The control variable harm reduction is denoted as $v(t)$. It represents the reduction in the harmfulness of drug use at time t in terms of a number between 0 and 1. Beyond the natural upper bound of $v(t) \leq 1$ we can expect that there will exist limits to what extent the harmfulness of a drug can be reduced.

In this thesis the control variable $v(t)$ focuses on harms felt by the users, whereas we assume that harms felt by the non-users are not affected or wiped out by those harm reduction measures targeted on the users. Consequently, there exists an upper boundary control condition demanding $v(t) \leq v_{\max} < 1$, whereas the lower boundary control condition is given by the non-negativity condition $v(t) \geq 0$.

The notion of harm reducion $v(t)$ in the current model and the parameterization of $v_{\max}$ are based on Cost of Illness (COI) studies (see Caulkins et al., in submission). The costs listed there are used as a proxy for social costs that stem from drug abuse and are borne by society in general, i.e. those harms and social costs entailed by illicit drug consumption are diluted over users and non-users. We assume that the health-related costs listed in COI studies are borne by the users, whereas the remaining components of COI studies (e.g. also crime costs) are borne by third parties.

As mentioned above, the harm reduction control variable $v(t)$ in the current model focuses only on the harms felt by the users, and we assume that those costs can be wiped out. Hence, $v_{\max}$ denotes the share of costs that is felt by the users. As derived in Caulkins et al. (in submission), the proportion of COI study costs that is health-related gives this upper bound $v_{\max}$ for the harm reduction control variable $v(t)$. Please note that given this definition of harm reduction, the specific value of $v_{\max}$ will vary depending on the particular drug and the country.

Reducing harms felt by drug users has a beneficial effect on the objective function in the model. Nevertheless, the model also considers fears that harm reduction might trigger behavioral responses. The opponent effects in the objective function and the dynamics are explained in the following two subsections.

2.5.1 Objective Function: The Benefit Attributable to Harm Reduction

The objective is to minimize total discounted drug-related social costs over the planning horizon. Social costs are the product of

- the baseline social costs when there is no harm reduction, which is normalized to $\kappa = 1$ without loss of generality,

- the number of drug users, $A(t)$, and

- $1 - v(t)$, the proportion of total social costs that is not averted via harm reduction policies.

Additionally one might want to consider

- costs for the control $v(t)$, captured by a cost function $c(v(t))$.

Most of the optimal control models with classic controls consider control costs in the objective function. Budget spending for law enforcement, treatment, or prevention is given in U.S. dollars or some other currency, but always synchronized to the occurring social costs terms. Harm reduction is modeled as a percentage, not as a monetary budget. Thus, we cannot simply add the term $v(t)$ in order to penalize spending on harm reduction. An appropriate cost function $c(v(t))$ has to be considered. There is evidence that $c(v(t)) = 0$ might be most appropriate. In countries like Australia or the Netherlands, where harm reduction is the centerpiece of national drug control strategies, harm reduction programs actually receive very modest levels of funding (Moore, 2005; Rigter, 2006). More generally, one should think of harm reduction as a kind of policy that integrates a new attitude to existing drug control strategies, and not consider it as a program with a budget. A simple example is a jurisdiction that pursues a harm reduction strategy and tells police not to arrest people for possessing a syringe. Such measures are easy to implement and essentially do not cost much money. This is the motivation why in the base case parameterization, the cost function $c(v(t))$ is omitted. Later on in this thesis, when variations of the model are assessed, a linear function $c \cdot v(t)$ is investigated.

Putting the above items together, we formulate the objective functional in generalized form

$$\min_{0 \leq v \leq v_{\max}} \left\{ J = \int_0^\infty e^{-rt} \Big(A(t) \big(1 - v(t) \big) + c\big(v(t) \big) \Big) \, \mathrm{d}t \right\}. \qquad (2.7)$$

In the best case of application of full harm reduction, where all harms felt by the users are wiped out, the share $v_{\max}$ of total harm is reduced, but there is still the remainder $(1 - v_{\max})$ that is felt by the non-using community.

Please note that the objective functional's value can be interpreted equally as aggregated harms or aggregated costs. The background of COI studies for the parameterization makes the two terms "harms" and "social costs" essentially equal.

2.5.2 The Negative Effect of Harm Reduction: Initiation Increases

As discussed in sections 2.2 and 2.3 harm reduction can have adverse effects on total use. Risk compensation due to harm reduction is a controversial issue. There are claims and fears that harm reduction might lead to adverse behavioral reactions, but no one knows for sure if this downside really exists. To the best of my knowledge, no one has ever tried before to find a function to model such effects. The present model does not consider the possibility of an increase in the quantity consumed by current users, it limits the adverse effect of harm reduction to an increase in initiation. That increase is modeled by the function $g\big(v(t)\big)$ which multiplies the baseline (i.e. when there is no harm reduction) initiation term.

The properties of the function $g\big(v(t)\big)$ are:

- When harm reduction is not applied, i.e. for $v(t) = 0$, there is no increase beyond the baseline initiation:

$$g(0) = 1. \tag{2.8}$$

- For any positive amount of harm reduction that is done, there is an increase in initiation, if only slightly. More mathematically stated,

$$g\big(v(t)\big) > 1 \quad \text{for} \quad v(t) > 0. \tag{2.9}$$

- The negative effect in initiation is higher, when more harm reduction is done. This means, that the function $g(v(t))$ is increasing in its argument $v(t)$:

$$g'(v(t)) > 0 \quad \text{for} \quad v(t) > 0. \tag{2.10}$$

- The direct percentage change in initiation into drug use is always less than the percentage change in harms felt by the user. In mathematical terms this can be expressed as

$$g\big(v(t)\big) - 1 \leq v(t) \quad \text{for} \quad 0 \leq v(t) \leq v_{\max}. \tag{2.11}$$

The functional form of $g\big(v(t)\big)$ follows a simple approach. We first assume that changes in the non-monetary costs of drug use (e.g. health risks) induce the same changes in the consumption decision of a user as changes in the monetary costs (money spent for purchase of drugs) of drug use do. This is particularly convenient because there is a growing empirical literature on drug price elasticities that deals with how responsive drug use is to changes in drug price (e.g. Grossman, 2004; Dave, 2004, 2008).

Several cost terms occurred in the preceding text: social costs, monetary and non-monetary costs of drug use. It is important not to mix them up. In particular, the personal, non-monetary costs of drug use are distinct from the social costs of drug use in two ways.

First, some of the costs included in social costs are externalities from a user's point of view. The most notable share among those is costs that are imposed on third parties, stemming from drug-related crime and violence. These costs can make up a considerable share of the overall social costs (e.g. in the case of the U.S. cocaine epidemic). Of course, a decision maker should be interested in implementation of strategies to reduce them. An example for such interventions is to push drug markets away from street corners, where dealing takes place under rather violent conditions, into more covert and less destructive forms. But such policies do not activate the risk of a behavioral response that increases participation in drug use. And the costs are not only borne by the users, but also by third parties. Thus, as important such policies may be, they are orthogonal to the current model. As outlined in section 2.5.1, the model takes costs listed in COI studies as a proxy for social costs. It assumes that the health-related costs documented in those studies are borne by the users (i.e. those costs are the personal, non-monetary costs of use), and that the other components of the costs listed there are seen as externality from the point of view of the users.

Second, even a fraction of the health-related costs may be an externality to the users. For example, when a user overdoses, one can distinguish costs that are borne by the user (e.g. reduced income when work time is lost because of morbidity and/or increased mortality) and other costs that are not (e.g. emergency care provided at no charge or medical treatment covered by state administered health services like Medicaid in the U.S.). Thus, we

assume that drug users do not fully factor the health-related costs into their consumption decision.

In the function $g\big(v(t)\big)$ the parameters c_m and c_s denote the monetary and social costs of drug consumption, respectively. The parameter ω denotes the proportion of health-related non-monetary costs of drug consumption the users factor into their consumption decisions. In the absence of empirical evidence, we set this value equal to $\omega = 0.5$. Then, the cost recognized by the user when no harm reduction is done is given by $c_m + \omega c_s v_{\max}$. When harm reduction is done to some extent $v(t) \in (0, v_{\max}]$, the costs felt by the user are reduced to $c_m + \omega c_s\big(v_{\max} - v(t)\big)$. A constant elasticity of demand model suggests the functional form

$$g\big(v(t)\big) = \left(\frac{c_m + \omega c_s\big(v_{\max} - v(t)\big)}{c_m + \omega c_s v_{\max}} \right)^{\eta}, \tag{2.12}$$

with η being the elasticity of participation in drug use with respect to cost of drug consumption.

It is noteworthy that we only need to synchronize the units of price c_m and other costs c_s with each other, but not with the objective function coefficient κ normalized to 1. This is because in the function $g\big(v(t)\big)$, the units in the numerator and the denominator cancel out.

There is a certain subtlety about the elasticity η when we want to utilize the above mentioned price elasticities. Grossman (2004), Dave (2004, 2008) and others empirically estimated the elasticity of participation with respect to price (which is captured by c_m in the model). We have to find a link between elasticity with respect to cost and with respect to price. Therefore we interpret the function g as a function of c_m and let the constant c_0 denote the baseline value of costs felt by the user. This yields

$$g(c_m) = \left(\frac{c_m + \omega c_s(v_{\max} - v)}{c_0} \right)^{\eta}.$$

Per definition the empirically measured price elasticity γ is

$$
\begin{aligned}
\gamma &= \frac{\mathrm{d}g(c_m)}{g(c_m)} \frac{c_m}{\mathrm{d}c_m} \\
&= \eta \left(\frac{c_m + \omega c_s(v_{\max} - v)}{c_0} \right)^{\eta-1} \frac{1}{c_0} c_m \left(\frac{c_m + \omega c_s(v_{\max} - v)}{c_0} \right)^{-\eta} \\
&= \frac{\eta c_m}{c_m + \omega c_s(v_{\max} - v)}.
\end{aligned}
$$

This formula helps to derive an expression for the elasticity with respect to cost, η, given the elasticity with respect to price, γ, namely

$$\eta = \frac{c_m + \omega c_s(v_{\max} - v)}{c_m}\gamma.$$

2.6 Parameter Values

The derivation of the base case parameter values summarized in Table 2.1 is presented in detail in Caulkins et al. (in submission). For the sake of conciseness of this thesis, the detailed re-exposition of the parameter derivation is avoided.

Remember that the states tracked in the current model are drug users, A, and a pool of susceptible non-users, S. The Australian parameterization with respect to states is based on a Multiple State Markov Chain Model (Caulkins et al., 2007). The drug of major interest in the U.S. is cocaine, and our A state is derived from the LH Markov Chain Model in Caulkins et al. (2004). Based on those sources, the reader should associate the state A with injecting drug users (mostly heroin) for the results derived for Australia. For the U.S., the state A shall be thought of as cocaine users, independent on the frequency of use or whether the user is dependent on the drug or not. The pool of susceptible non-users S is best imagined as adolescents and young adults that are vulnerable to try that drug.

Please note that the data underlying the parameterization of the dynamics of users A and susceptible persons S is given in years.

When looking at the results presented in the following sections, please keep in mind that the parameter estimates are by no means precise since the data may not be perfectly accurate due to the illicit nature of the drug industry.

We assume $\tau \geq 0$ with the special case of $\tau = 0$ in the base case computations, $b > 0$ for the initiation proportionality constant, and $\alpha > 0$ for the initiation function exponent. For the parameters of the function $g(v(t))$, social costs c_s and monetary costs c_m are assumed to be non-negative, the factor ω can vary in the interval $[0, 1]$, the elasticity η of drug participation with respect to cost is negative. The parameter k is assumed to be non-negative, whereas for the rates μ and δ we assume that they are positive. Positivity is also demanded for the discount rate r.

In the objective functional J in equation (2.7), $r > 0$ is the time discount rate. It is assumed to be constant over the whole planning horizon. If the

Parameter	Symbol	Australian IDU	U.S. Cocaine
Inflow into S-state	k	0.0526	1.3417
Maturing out from S-state	δ	0.0952	0.0605
Coefficient of innovation	τ	0	0
Proportionality constant	b	0.5112	0.0090
Exponent in initiation function	α	0.8622	1.5604
Exit rate from A-state	μ	0.1136	0.1661
Social costs of use	c_s	\$ 39,255/yr	\$ 223.56/g
Monetary costs of use	c_m	\$ 13,537/yr	\$ 106.54/g
Health costs factor in $g(v)$	ω	0.5	0.5
Upper bound for control v	$v_{\max}$	0.53	0.17408
Price elasticity of participation	γ	-0.21	-0.45
Cost elasticity of participation	η	-0.371	-0.532
Annual discount rate	r	0.04	0.04

Table 2.1: Parameter values for the base case model for IDU in Australia and cocaine consumption in the U.S.

decision maker is interested in what happens in the future, r will be small, i.e. close to zero, and the decision maker is said to be "farsighted". A "myopic" planner, who does not care much about the future development, is characterized by large values of r. We model a farsighted decision maker, discounting at 4% per year, i.e. $r = 0.04$.

Regarding the planning horizon, one could argue that for drug problems, policy programs will change when a new election period begins, and that thus a finite time horizon T would be more appropriate than the infinite time horizon we use in the objective functional (2.7). Nevertheless, we follow the spirit of Arrow & Kurz (1970, p. xvii), who justify consideration of an infinite time horizon in the following way: "The infinite horizon is an idealization of the fundamental point that the consequences of investment are very long-lived; any short horizon requires some methods of evaluating end-of-period capital stocks, and the only proper evaluation is their value in use in the subsequent future."

Examining the values estimated for the proportion of social costs that can be averted in the best case, in the model's notation concisely captured by the variable $v_{\max}$, we find that the upper bound for the control $v(t)$ is much larger in Australia (53%) than in the U.S. (17%). This perfectly reflects the fact that much of the social costs attributable to heroin consumption in Australia are health-related and stem from problems for which effective harm reduction tactics exist (e.g. preventing overdose and the spread of blood
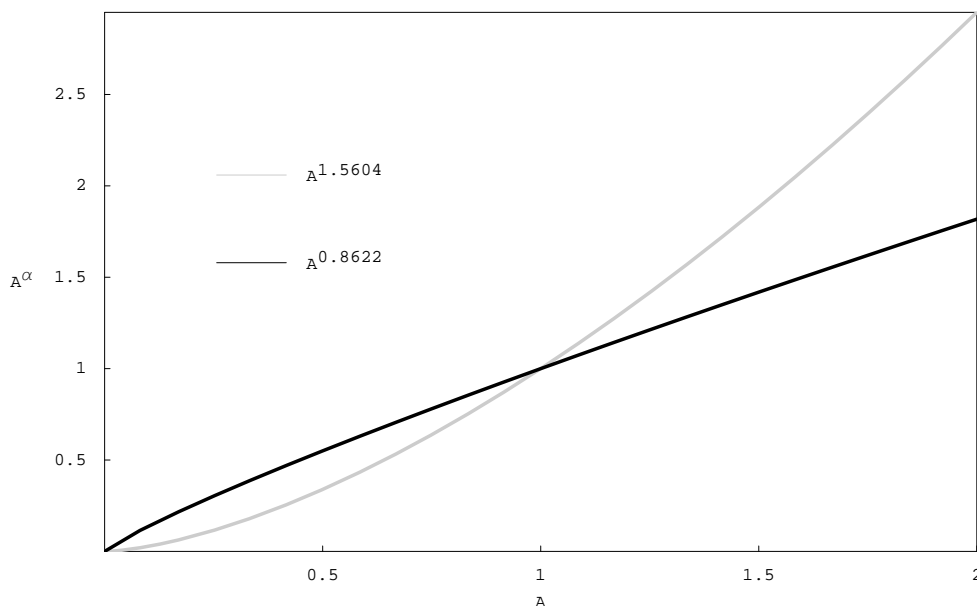
Figure 2.3: Functions $f(A) = A^\alpha$ for the U.S. parameter $\alpha = 1.5604$ in gray and for the Australian parameter $\alpha = 0.8622$ in black

borne infections), whereas the lion's share of social costs stemming from cocaine abuse in the U.S. is not health-related, but associated with crime and violence.

The discussion of the epidemic concept that underlies the initiation mechanism in section 2.4.3 presents the idea of concave and convex dependence of initiation on the number of current users. In the base case initiation function $I(A(t), S(t), v(t))$ from equation (2.3) the distinction is governed by the exponent α. The parameterizations presented above indeed yield a convex function with $\alpha = 1.5604 > 1$ for the U.S. cocaine epidemic and a concave function with $\alpha = 0.8622 < 1$ for Australia's IDU population. Figure 2.3 shows the corresponding curves $f(A(t)) = A(t)^\alpha$ for the U.S. cocaine epidemic in gray and for Australian IDUs in black.

Structurally the parameter α is the most interesting and important parameter of the initiation function. For the U.S. cocaine epidemic we find $\alpha > 1$, thus initiation is convex with respect to the number of users. Convexity implies the possibility that multiple stable equilibria emerge, with their basins of attraction separated by a curve of tipping points. In contrast, for Australia's injection drug use, where the parameter estimate is $\alpha < 1$, we expect to encounter only one stable steady state. The fact of existence of tip-
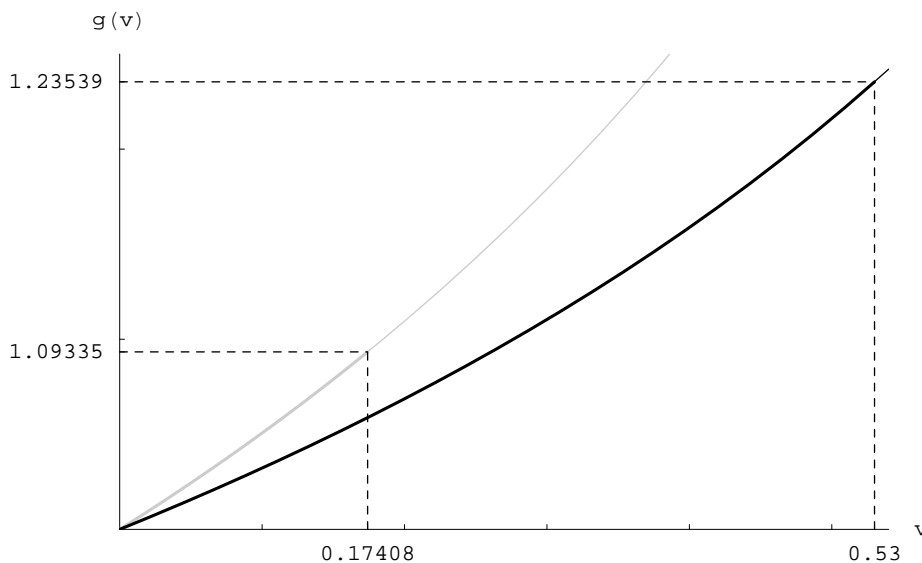
Figure 2.4: Functions $g(v)$ for the U.S. cocaine epidemic (gray) and Australian IDU (black)

ping points in the model for the U.S. and the absence of such a phenomenon for the Australian parameterization suggests that in the U.S., harm reduction has greater potential to trigger catastrophic consequences in trajectories of use.

It is easy to see that the function $g\big(v(t)\big)$ presented in equation (2.12) fulfills that $g(0) = 1$, which is the property demanded by equation (2.8). Figure 2.4 shows the function $g\big(v(t)\big)$ parameterized for Australian IDU and the U.S. cocaine epidemic. The plot shows the entire curves, but due to the control constraint $0 \leq v(t) \leq v_{\max}$ the relevant part is limited to this interval. To emphasize this limitation, the graph is plotted bold on the relevant interval. The black curve is for Australian IDU, where the upper bound for control interventions is quite high, $v_{\max} = 0.53$. The maximum reduction of 53% of social cost that wipes out all the harms felt by the users yields an adverse effect of an increase of less than 24% in initiation. In the U.S. (gray curve), the maximum reduction of social costs is more modest, about 17%. The negative behavioral response is less than 9% increased initiation. Figure 2.4 shows that the fully parameterized function $g\big(v(t)\big)$ for either case indeed has the properties (2.9) - (2.11).

Note that with the parameters presented in Table 2.1 for both U.S. cocaine use and Australian IDU the function $g\big(v(t)\big)$ is convex, though only modestly. Convexity is a property that meets the following considerations

about how a rational drug control planner would align harm reduction interventions. Among the various programs that rank among harm reduction policies a smart policy maker would figure out which program generated the least impact on initiation per unit reduction in harm and implement that program first. Then the policy maker would determine the next program, the one with the second lowest impact on initiation, and implement it, and so on. Thus, the different programs are implemented from the least problematic ones to those that have more adverse effects on use. Consequently, one can presume that $g\big(v(t)\big)$ is convex.

Additionally, the only modest convexity points to the fact that a linear approximation of the function $g\big(v(t)\big)$ may be suitable for some applications. We deal with nonlinear dynamic systems and do not focus on models that are linear with respect to control. Such models have so-called bang-bang solutions. This means that for a certain time, the optimal strategy is to choose control at either the lower or the upper bound, but then a point $t = \theta$ in time is reached, where it is optimal to switch to the other extreme of the control domain and stay there until the end time T is reached. Modest convexity of $g\big(v(t)\big)$ opens the possibility that the optimal control results involve boundary solutions with respect to control, meaning that for certain intervals of time t, $v^*(t) = 0$ or $v^*(t) = v_{\max}$ is the optimal policy.

Chapter 3

Base Case Models with Static Control

For the ease of exposition and if there is no ambiguity, the time argument t is mostly omitted in what follows. Furthermore, the denominations equilibrium, steady state, fixed point or critical state/point are used synonymously.

The ultimate aim of our investigations is to solve optimal control problems based on a two-dimensional system

$$
\begin{aligned}
\dot{A} &= \mathrm{I}(A, S, v) - \mu A, & A(0) &= A_0, \\
\dot{S} &= k - \delta S - \mathrm{I}(A, S, v), & S(0) &= S_0.
\end{aligned}
\tag{3.1}
$$

Before analyzing an optimal control model, the uncontrolled (purely descriptive) model is often assessed as a benchmark case. Here the uncontrolled model is assessed, i.e. the model with $v(t) = 0$ for all t, and additionally the system is run with control v set to $v(t) = v_{\max}$ for all t. We assume that implementation of harm reduction is a one-time, irrevocable decision. The focus is not yet put on choosing the best control value $v(t)$ over time for a particular set of initial conditions $(A(0), S(0))$, but on comparing the performance of pure use reduction (denoted as $v \equiv 0$) vs. full harm reduction (labeled $v \equiv v_{\max}$) for various initial conditions. The simple comparative study conducted for the two-dimensional system using the base case initiation function $\mathrm{I}(A, S, v) = b\, A^\alpha\, S\, g(v)$, which only takes into account imitators, given by

$$
\begin{aligned}
\dot{A} &= b\, A^\alpha\, S\, g(v) - \mu A, & A(0) &= A_0, \\
\dot{S} &= k - \delta S - b\, A^\alpha\, S\, g(v), & S(0) &= S_0,
\end{aligned}
\tag{3.2}
$$

yields constitutional results and is an important and fruitful point of departure for the optimally controlled versions of the model, which are presented in Chapter 4 for the base case and in Chapter 6 for some interesting variations.

Please note that the states A and S are denoted in millions, only sometimes numbers like $A = 0.3$ are reported as 300,000 users.

3.1 General Remarks

Some analysis can be done easily before the parameters values specific for a particular drug and country are substituted into the model.

To determine the steady states $(\hat{A}, \hat{S})$ of the generalized system (3.1), the equations for $\dot{A}$ and $\dot{S}$ are set equal to zero and this system is solved simultaneously. Except for special cases of α that do not occur in the present parameterizations, the solution cannot be computed analytically. Still, we can directly derive some insights. Summing up the equations $\dot{A} = 0$ and $\dot{S} = 0$, the terms including the initiation function $I(A, S, v)$ cancel out, which yields that in steady state there holds $k - \delta \, \hat{S} - \mu \, \hat{A} = 0$. Solving for $\hat{S}$ yields that the steady states of the system satisfy

$$\hat{S} = \frac{k - \mu \hat{A}}{\delta}. \tag{3.3}$$

This is a downward sloping line between $(A, S) = (0, \frac{k}{\delta})$ and $(A, S) = (\frac{k}{\mu}, 0)$ and is not dependent on the harm reduction control variable v. So, no matter if control is applied or not, the steady states of the system are located according to this linear relation. Consequently, the highest possible value of susceptibles in steady state is $\hat{S}_{\max} = \frac{k}{\delta}$, whereas the highest possible steady state number $\hat{A}$ for the users is located at $\hat{A}_{\max} = \frac{k}{\mu}$.

In the base case model with dynamics (3.2), all trajectories starting in the positive quadrant never leave it. The A-axis acts as a repeller, because

$$\lim_{S \to 0} \dot{S} = \lim_{S \to 0} \left(k - \delta S - b A^\alpha S g(v) \right) = k > 0.$$

The S-axis (i.e. $A = 0$) is a natural delimiter of the system, because there the dynamics of users comes to rest for any $\alpha > 0$:

$$\dot{A}|_{A=0} = \left(b A^\alpha S g(v) - \mu A \right)|_{A=0} = 0.$$

3.2 Australia: Isoclines, Steady States and Phase Portraits

We start with the static investigation of the model for the Australian parameters. The uncontrolled model is analyzed first, i.e. $v \equiv 0$. Please note that when control is set equal to zero for all time, the function $g(v)$ takes the value $g(v) = g(0) = 1$ as demanded in equation (2.8).

In order to locate the equilibria of the two-dimensional system (3.2), the equations $\dot{A}|_{v\equiv0} = 0$ and $\dot{S}|_{v\equiv0} = 0$ have to be solved simultaneously, which is done numerically. The locus where the dynamics of a state comes to rest is called isocline. For example, all points in the (A, S)-plane, where $\dot{A}|_{v\equiv0} = 0$ holds, are conditions for which the number of users does not change over time when harm reduction is never applied. In a steady state $(\hat{A}_{v\equiv0}, \hat{S}_{v\equiv0})$ of the uncontrolled system, neither A nor S changes over time, thus the intersections of the isoclines $\dot{A}|_{v\equiv0} = 0$ and $\dot{S}|_{v\equiv0} = 0$ provide the system's steady states. In the case of the Australian parameterization two equilibria are computed which are located at

$$\begin{pmatrix} \hat{A}_{v\equiv0}^1 \\ \hat{S}_{v\equiv0}^1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.552521 \end{pmatrix}, \qquad \begin{pmatrix} \hat{A}_{v\equiv0}^2 \\ \hat{S}_{v\equiv0}^2 \end{pmatrix} = \begin{pmatrix} 0.304916 \\ 0.188672 \end{pmatrix}.$$

In Figure 3.1 the isocline $\dot{S}|_{v\equiv0} = 0$ given by $S = \frac{k}{\delta + bA^\alpha}$ is depicted as a gray dashed curve. The isocline $\dot{A}|_{v\equiv0} = 0$ has two branches. For $A = 0$ there holds $\dot{A}|_{v\equiv0} = 0$, independent from the actual number of susceptibles S. So the axis $A = 0$ is a branch of the isocline. For any value $A > 0$ a unique value for S can be determined, where the dynamics with respect to A comes to rest. This function $S = \frac{\mu A}{bA^\alpha}$ gives the second branch of the isocline $\dot{A}|_{v\equiv0} = 0$. In Figure 3.1 it is shown as a solid gray curve. The steady state $(\hat{A}_{v\equiv0}^2, \hat{S}_{v\equiv0}^2)$ located at the intersection of the isoclines in the interior of the first quadrant is depicted as a black dot.

In order to analyze the computed steady states with respect to their qualitative properties graphically, the phase diagrams in the neighborhood of the equilibria are constructed. The little arrows give the direction of evolution of the dynamics of users and susceptibles over time. It can easily be detected in Figure 3.1 that the steady state $(\hat{A}_{v\equiv0}^2, \hat{S}_{v\equiv0}^2)$ is a stable focus. The no-use steady state $(\hat{A}_{v\equiv0}^1, \hat{S}_{v\equiv0}^1)$ can only be achieved for initial conditions with $A(0) = 0$. This represents the case that the epidemic never starts, which is of no interest here. The steady state is a saddle point.

The stability properties can also be shown analytically by investigation of the Eigenvalues of the Jacobian Matrix evaluated at the critical point or
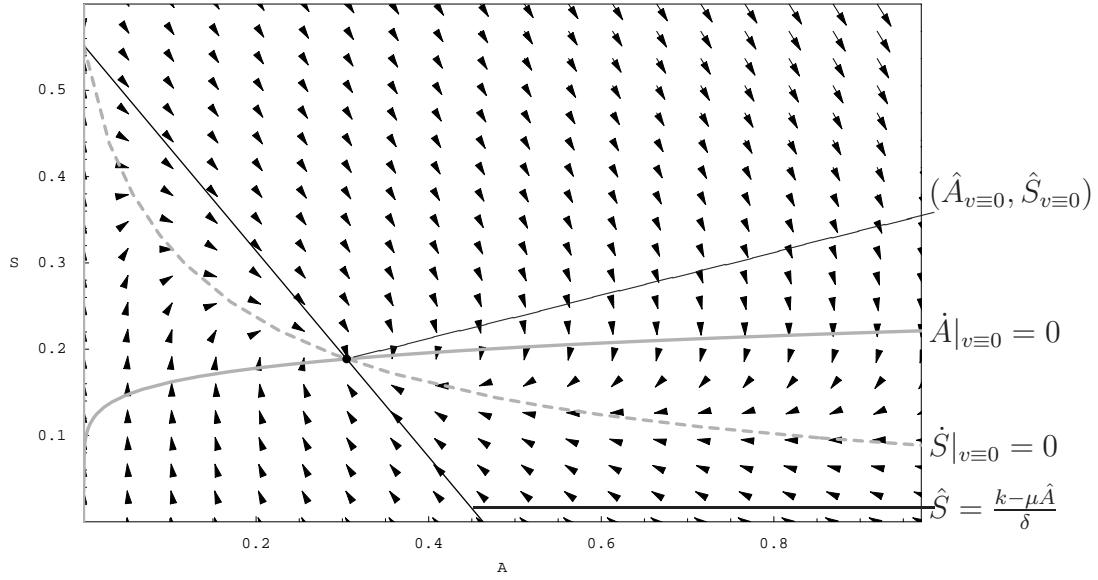
Figure 3.1: Phase portrait for the Australian base case parameter set with static control $v \equiv 0$

by calculating the trace and the determinant of the Jacobian Matrix, respectively. The theoretical background on the determination of the stability properties of steady states of two-dimensional systems is presented in Appendix A.1. The mathematically more proper, non-graphic stability analysis for the above presented equilibria is conducted in section 3.3.

For the model with static control at the upper bound, $v \equiv v_{\max}$, the phase portrait is qualitatively equivalent to the one described above. The intersections of the isoclines $\dot{S}|_{v\equiv v_{\max}} = 0$ and $\dot{A}|_{v\equiv v_{\max}} = 0$ are located at

$$\begin{pmatrix} \hat{A}^1_{v\equiv v_{\max}} \\ \hat{S}^1_{v\equiv v_{\max}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0.552521 \end{pmatrix}, \qquad \begin{pmatrix} \hat{A}^2_{v\equiv v_{\max}} \\ \hat{S}^2_{v\equiv v_{\max}} \end{pmatrix} = \begin{pmatrix} 0.333455 \\ 0.154617 \end{pmatrix},$$

where the first one is a saddle point which is only achieved when one moves along the S-axis. The second one is a stable focus.

Figure 3.2 depicts the isocline $\dot{S}|_{v\equiv v_{\max}} = 0$ for the system (3.2) with static control $v \equiv v_{\max}$ as a black dashed curve. The black solid curve is the interior (i.e. $A > 0$) branch of the isocline $\dot{A}|_{v\equiv v_{\max}} = 0$. At their intersection, the steady state $(\hat{A}^2_{v\equiv v_{\max}}, \hat{S}^2_{v\equiv v_{\max}})$ is depicted as a black dot. Furthermore, the arrows trace the vector field of the system dynamics under static control $v \equiv v_{\max}$. The isoclines of the uncontrolled system are shown in gray. The

$(\hat{A}^2_{v\equiv 0}, \hat{S}^2_{v\equiv 0})$ $(\hat{A}^2_{v\equiv v_{\max}}, \hat{S}^2_{v\equiv v_{\max}})$

$\dot{A}|_{v\equiv 0} = 0$

$\dot{A}|_{v\equiv v_{\max}} = 0$

$\dot{S}|_{v\equiv 0} = 0$

$\dot{S}|_{v\equiv v_{\max}} = 0$
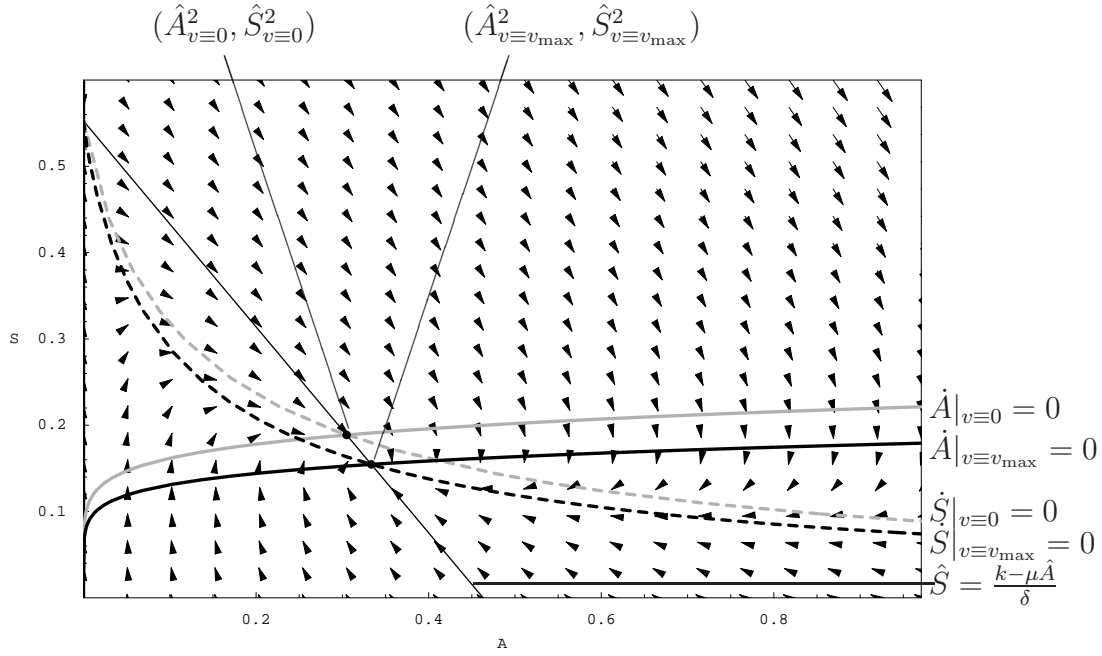
$\hat{S} = \frac{k-\mu\hat{A}}{\delta}$

Figure 3.2: Phase portrait for the Australian base case parameter set with static control $v \equiv v_{\max}$

line $\hat{S} = \frac{k-\mu\hat{A}}{\delta}$ given by equation (3.3) is depicted in black. In section 3.1 this line was identified as the characteristic line along which the steady states of the system are located. It is clearly visible that the computed equilibria indeed lie on the downward sloping line. Intuitively, one assumes that when full harm reduction is done and if it triggers an increase in initiation, the pool of users is larger throughout the epidemic and hence is also larger in equilibrium, whereas there are less susceptibles in steady state, because more individuals out of the S-state are recruited to drug use. Figure 3.2 provides the graphical confirmation of this intuition, because the full harm reduction steady state $(\hat{A}^2_{v\equiv v_{\max}}, \hat{S}^2_{v\equiv v_{\max}})$ is located to the right and below the steady state $(\hat{A}^2_{v\equiv 0}, \hat{S}^2_{v\equiv 0})$ of the system without harm reduction.

3.3 Local Stability Behavior of the System Around Steady States, Australia

In order to investigate the local stability behavior of the system, we derive the Jacobian Matrix of the two-dimensional base case system of state dynamics.

For general v it is of the form

$$J = \begin{pmatrix} b\,\alpha\,A^{\alpha-1}\,S\,g(v) - \mu & b\,A^\alpha\,g(v) \\ -b\,\alpha\,A^{\alpha-1}\,S\,g(v) & -\delta - b\,A^\alpha\,g(v) \end{pmatrix}.$$

Evaluated at the equilibrium $(\hat{A}^2_{v\equiv0}, \hat{S}^2_{v\equiv0})$ we get

$$J = \begin{pmatrix} -0.0156541 & 0.183591 \\ -0.0979459 & -0.278791 \end{pmatrix},$$

which has the conjugate complex Eigenvalues

$$e_{1,2} = -0.147222 \pm 0.0259179\,\mathrm{i}.$$

According to Appendix A.1 this indicates that the fixed point $(\hat{A}^2_{v\equiv0}, \hat{S}^2_{v\equiv0})$ is a stable focus. Equivalently, we find $\Delta = 0.0223462$ for the determinant of the Jacobian Matrix and $\tau = -0.294445$ for its trace, which yields $\tau^2 - 4\Delta = -0.00268696 < 0$. This also identifies the steady state as a stable focus.

For the no-use steady state $(\hat{A}^1_{v\equiv0}, \hat{S}^1_{v\equiv0})$ the analysis is less straightforward. In the Australian parameter set, the exponent α is $\alpha = 0.8622 < 1$. The Jacobian Matrix involves the term $A^{\alpha-1}$, which has a pole at $A = 0$. We cannot calculate the Jacobian Matrix and determine its Eigenvalues, but we can look at the expressions for its trace τ and the determinant Δ:

$$\begin{aligned} \tau &= b\,\alpha\,A^{\alpha-1}\,S - \mu - \delta - b\,A^\alpha \\ \Delta &= -b\,\alpha\,\delta\,A^{\alpha-1}\,S + b\,\mu\,A^\alpha + \delta\,\mu. \end{aligned}$$

The parameters μ, δ and the value of the product $\delta\mu$ are fixed and independent of A. For S we substitute $\hat{S} = \frac{k}{\delta} + dA$, with $d \in \mathbf{R}$. For $0 < \alpha < 1$, the terms including A^α tend to zero for $A \to 0$ (those including d converge independently from this constant), while the terms with $A^{\alpha-1}$ diverge. This leads to

$$\begin{aligned} \lim_{A\to0} \tau &= \lim_{A\to0} b\,\alpha\,\frac{k}{\delta}\,A^{\alpha-1} = +\infty \\ \lim_{A\to0} \Delta &= \lim_{A\to0} -b\,\alpha\,k\,A^{\alpha-1} = -\infty. \end{aligned}$$

The determinant is negative, hence the fixed point without users is a saddle point.

Substituting $v \equiv v_{\max}$ into the Jacobian Matrix, at $(\hat{A}^2_{v\equiv v_{\max}}, \hat{S}^2_{v\equiv v_{\max}})$ we evaluate

$$J = \begin{pmatrix} -0.0156541 & 0.244995 \\ -0.0979459 & -0.340195 \end{pmatrix}.$$

The Eigenvalues are real and negative

$$e_1 = -0.226251, \qquad e_2 = -0.129598.$$

For the trace and the determinant of the Jacobian Matrix we get

$$\tau = -0.355849, \ \Delta = 0.0293217, \ \text{with} \ \tau^2 - 4\Delta = 0.00934172 > 0,$$

consequently the fixed point $(\hat{A}^2_{v\equiv v_{\max}}, \hat{S}^2_{v\equiv v_{\max}})$ satisfies the properties of a stable node.

Analogously to the above investigation for the case $v \equiv 0$ it can be shown that the no-use equilibrium $(\hat{A}^1_{v\equiv v_{\max}}, \hat{S}^1_{v\equiv v_{\max}})$ is a saddle point.

3.4 U.S. Cocaine: Isoclines, Steady States and Phase Portraits

The exponent $\alpha > 1$ in the initiation function raises the possibility of multiple stable equilibria. Such multiplicity often generates "tipping points" in models. From the mathematical point of view this is caused by the occurrence of saddle points. The importance of this will be explained below.

Concerning drug use, there is the following interpretation of tipping points. When the number of persons engaging in the drug market (be it using or selling) is of modest size, the drug does not spread for different reasons. First, if little is known about a drug or its consumption is associated with small, atypical populations, people are less likely to accept offers to consume. Second, if the market is still small in the sense that there are only a few sellers and a few drug consumers, they may have a hard time to locate each other. This keeps the market from spreading. A third argument is that in a "thin" market, a modest level of enforcement considerably increases the risk of getting caught for each of the few participants in the market (cf. Kleiman, 1993). The other extreme is widespread use of the drug. When use is so common that non-using persons receive a multitude of offers, they are more likely to try the drug that most of their friends consume. Furthermore, if there are many users, the corresponding demand is accommodated by a considerable number of sellers. The tipping point is the hairline case between those extreme stages. It identifies the critical market size which determines the borderline between drug use being limited to a modest number of users and widespread drug consumption.

If a drug epidemic is near such a tipping point, small changes may have a potential to tip trajectories of prevalence away from convergence to modest

levels to convergence to high-use levels. Effects of very modest changes can be large in terms of numbers of drug users, long-lasting in terms of time and may impose high social costs. Thus, a decision maker should act with caution if he knows the drug epidemic he or she is confronted with might have such tipping points.

Multiple stable equilibria separated by tipping points indeed occur in the model of the U.S. cocaine epidemic.

Solving the system ($\dot{S}|_{v\equiv0} = 0$, $\dot{A}|_{v\equiv0} = 0$) for the base case dynamics (3.2) and the U.S. parameterization gives three steady states located at

$$\begin{pmatrix} \hat{A}^1_{v\equiv0} \\ \hat{S}^1_{v\equiv0} \end{pmatrix} = \begin{pmatrix} 0 \\ 22.1769 \end{pmatrix}, \ \begin{pmatrix} \hat{A}^2_{v\equiv0} \\ \hat{S}^2_{v\equiv0} \end{pmatrix} = \begin{pmatrix} 0.8867 \\ 19.7426 \end{pmatrix}, \ \begin{pmatrix} \hat{A}^3_{v\equiv0} \\ \hat{S}^3_{v\equiv0} \end{pmatrix} = \begin{pmatrix} 5.4888 \\ 7.1076 \end{pmatrix}.$$

For the Australian parameter set, the no-use steady state could only be achieved if the epidemic never started. In the U.S. case, the parameter $\alpha = 1.5604 > 1$ induces a structural change. The no-use steady state plays a more important role, because there are initial conditions with $A(0) > 0$ for which the system will converge to the no-use steady state ($\hat{A}^1_{v\equiv0}, \hat{S}^1_{v\equiv0}$). In Figure 3.3 the two branches of the isocline $\dot{A}|_{v\equiv0} = 0$ are depicted in gray (solid). The dashed gray curve represents the locus where $\dot{S}|_{v\equiv0} = 0$. The three fixed points are located at their intersections, and are depicted as black dots in Figure 3.3.

In order to analyze the qualitative properties of the steady states graphically, the phase portrait of the system was constructed in a neighborhood of the fixed points. The little arrows in Figure 3.3 represent the vector field. Aiming for better visualization of the dynamics some trajectories are depicted as black curves. If we encounter the system in a state in the region under the dashed isocline $\dot{S}|_{v\equiv0} = 0$, the number of susceptibles S increases. Located above this dashed curve, S decreases. For initial conditions below and to the left of the inner branch of the isocline $\dot{A}|_{v\equiv0} = 0$ which is the gray solid curve, the number of users decreases, whereas above this gray curve use spreads.

The no-use steady state ($\hat{A}^1_{v\equiv0}, \hat{S}^1_{v\equiv0}$) and the fixed point ($\hat{A}^3_{v\equiv0}, \hat{S}^3_{v\equiv0}$) with a high number of users are stable. The no-use equilibrium is a node, and the high-use steady state is a focus. This can be directly conducted by interpretation of the phase portrait and can also be shown by investigation of the Eigenvalues of the Jacobian Matrix of the system at the steady states. The following section 3.5 is devoted to this analysis.

The intermediate-use steady state ($\hat{A}^2_{v\equiv0}, \hat{S}^2_{v\equiv0}$) is a saddle point. Only paths starting on its stable manifold end up in the saddle point. Such initial
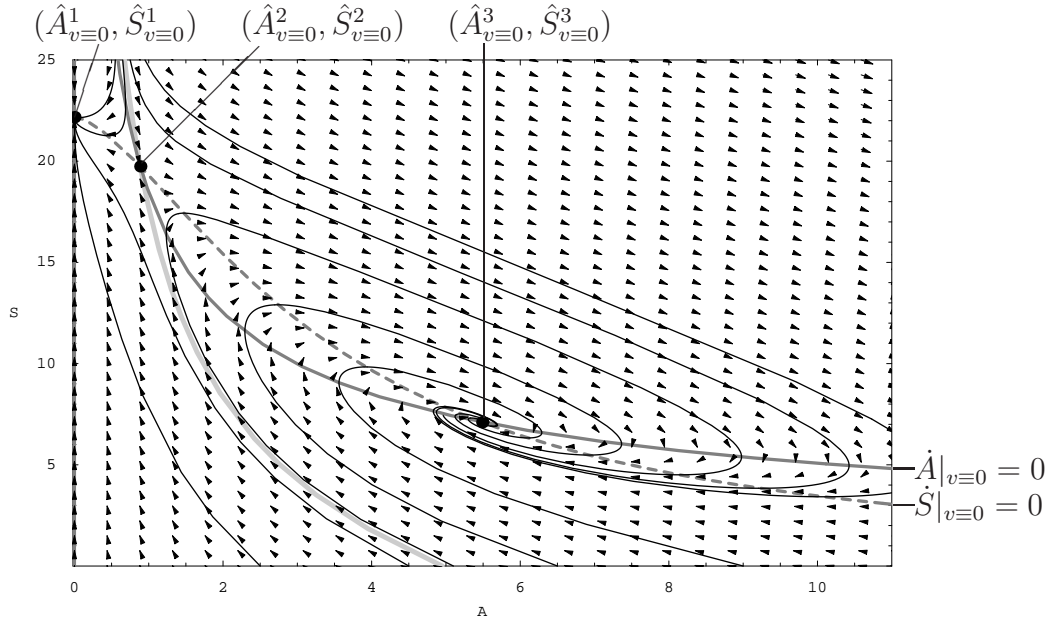
Figure 3.3: Phase portrait for the U.S. base case parameter set with static control $v \equiv 0$

conditions are a hairline case in the (A, S)-plane. The light gray curve in Figure 3.3 depicts the stable manifold of the intermediate-use steady state $(\hat{A}^2_{v\equiv0}, \hat{S}^2_{v\equiv0})$. It divides the first quadrant of the (A, S)-plane into two non-overlapping regions. Paths that emanate on the left of the curve converge to the no-use steady state. For any initial value to the right of the curve, the system dynamics approach the high-use steady state, possibly overshooting the steady state value of users considerably. The stable manifold of the saddle point separates the basins of attraction of the stable steady states. This property justifies calling the stable manifold a "separatrix".

In general, in purely descriptive analyses of dynamical systems, saddle points are quite unimportant because the probability to reach the saddle point is equal to zero. Notwithstanding, the saddle points or more concretely speaking their stable manifolds and the regions between and around them play a crucial role in the comparison of the two static control settings $v \equiv 0$ and $v \equiv v_{\max}$ in sections 3.6 and 3.7.

In the investigation of steady states of the statically controlled system (3.2), the case $v \equiv v_{\max}$ for the U.S. parameterization is still missing. The

steady states of the system $(\dot{A}|_{v\equiv v_{\max}} = 0, \dot{S}|_{v\equiv v_{\max}} = 0)$ are located at

$$\begin{pmatrix} \hat{A}^1_{v\equiv v_{\max}} \\ \hat{S}^1_{v\equiv v_{\max}} \end{pmatrix} = \begin{pmatrix} 0 \\ 22.177 \end{pmatrix}, \begin{pmatrix} \hat{A}^2_{v\equiv v_{\max}} \\ \hat{S}^2_{v\equiv v_{\max}} \end{pmatrix} = \begin{pmatrix} 0.727 \\ 20.181 \end{pmatrix}, \begin{pmatrix} \hat{A}^3_{v\equiv v_{\max}} \\ \hat{S}^3_{v\equiv v_{\max}} \end{pmatrix} = \begin{pmatrix} 5.777 \\ 6.317 \end{pmatrix}.$$

In analogy to the previously investigated case, from the left to the right we encounter a stable node, a saddle point and a stable focus. The phase portrait is qualitatively equivalent to the one for $v \equiv 0$ shown in Figure 3.3, hence the exposition is omitted.

3.5 Local Stability Behavior of the System Around Steady States, United States

The Jacobian Matrix at the high-use equilibrium $(\hat{A}^3_{v\equiv 0}, \hat{S}^3_{v\equiv 0})$ in the pure use reduction scenario is

$$J = \begin{pmatrix} 0.0930824 & 0.128269 \\ -0.259182 & -0.188769 \end{pmatrix}.$$

The Eigenvalues are conjugate complex with negative real parts

$$e_{1,2} = -0.0478434 \pm 0.115694\,\mathrm{i}.$$

For the determinant and the trace of the Jacobian Matrix evaluated at the equilibrium we find $\Delta = 0.015674$, $\tau = -0.0956868$ and the relation $\tau^2 - 4\Delta = -0.0535402 < 0$. Thus, the steady state $(\hat{A}^3_{v\equiv 0}, \hat{S}^3_{v\equiv 0})$ is classified as a stable focus.

Next, the intermediate-use steady state $(\hat{A}^2_{v\equiv 0}, \hat{S}^2_{v\equiv 0})$ is investigated. The Jacobian Matrix evaluated there is

$$J = \begin{pmatrix} 0.0930824 & 0.00745974 \\ -0.259182 & -0.0679597 \end{pmatrix}$$

with determinant $\Delta = -0.00439243$ and trace $\tau = 0.0251227$. The negative determinant immediately indicates that the fixed point is a saddle point. Indeed, we evaluate Eigenvalues

$$e_1 = 0.0800166, \qquad e_2 = -0.0548939,$$

which are both real and have opposite sign.

For the no-use steady state $(\hat{A}^1_{v\equiv 0}, \hat{S}^1_{v\equiv 0})$, in the Jacobian Matrix

$$J = \begin{pmatrix} b\,\alpha\,A^{\alpha-1}\,S\,g(v) - \mu & b\,A^{\alpha}\,g(v) \\ -b\,\alpha\,A^{\alpha-1}\,S\,g(v) & -\delta - b\,A^{\alpha}\,g(v) \end{pmatrix}$$

the terms with A^{α} and $A^{\alpha-1}$ both tend to zero for $A \to 0$, because in the U.S. case there holds $\alpha > 1$. So, the Jacobian Matrix is reduced to

$$J = \begin{pmatrix} -\mu & 0 \\ 0 & -\delta \end{pmatrix} = \begin{pmatrix} -0.1661 & 0 \\ 0 & -0.0605 \end{pmatrix},$$

which has the Eigenvalues

$$e_1 = -\mu = -0.1661, \qquad e_2 = -\delta = -0.0605.$$

The Eigenvalues are real and have negative signs, consequently the no-use steady state $(\hat{A}^1_{v\equiv 0}, \hat{S}^1_{v\equiv 0})$ is a stable node.

Finally, we investigate the local stability behavior of the steady states when $v \equiv v_{\max}$. The Jacobian Matrix evaluated at the high-use steady state $(\hat{A}^3_{v\equiv v_{\max}}, \hat{S}^3_{v\equiv v_{\max}})$ is

$$J = \begin{pmatrix} 0.0930824 & 0.151889 \\ 0.259182 & -0.212389 \end{pmatrix}.$$

Its determinant is $\Delta = 0.0195972$, the trace is computed as $\tau = -0.119306$. The positive determinant indicates that the critical point is either a focus or a node, the negative τ classifies the fixed point to be stable. For the distinction between focus and node, one evaluates $\tau^2 - 4\Delta = -0.064155 < 0$. Consequently, the high-use steady state is a focus. The Eigenvalues of J are indeed conjugate complex with negative real part

$$e_{1,2} = -0.0596532 \pm 0.126644\,\mathrm{i}.$$

The Jacobian Matrix evaluated at the fixed point with intermediate use, $(\hat{A}^2_{v\equiv v_{\max}}, \hat{S}^2_{v\equiv v_{\max}})$, is

$$J = \begin{pmatrix} 0.0930824 & 0.00598441 \\ -0.259182 & -0.0664844 \end{pmatrix}$$

with $\Delta = -0.00463748$, $\tau = 0.026598$, and Eigenvalues

$$e_1 = 0.0826845, \quad e_2 = -0.0560864.$$

Consequently, the fixed point $(\hat{A}^2_{v\equiv v_{\max}}, \hat{S}^2_{v\equiv v_{\max}})$ is a saddle point.

Analogously to the case with $v \equiv 0$, the Jacobian Matrix evaluated at the no-use steady state $(\hat{A}^1_{v \equiv v_{\max}}, \hat{S}^1_{v \equiv v_{\max}})$ is reduced to

$$J = \begin{pmatrix} -\mu & 0 \\ 0 & -\delta \end{pmatrix} = \begin{pmatrix} -0.1661 & 0 \\ 0 & -0.0605 \end{pmatrix}.$$

Hence, one concludes directly that it has the local stability properties of a stable node.

3.6 Comparative Analysis of the System with Static Control

This section contrasts the evolution of the states and the resulting value of the objective functional under the pure use reduction regime $v \equiv 0$ and under the full harm reduction regime $v \equiv v_{\max}$, i.e. under the static settings as explained at the beginning of Chapter 3.

3.6.1 Tipping Point Curves for the U.S. Cocaine Epidemic

We first conduct some investigations for the U.S. cocaine epidemic. Figure 3.4 depicts the inner branch of the isocline $\dot{A}|_{v \equiv 0} = 0$ in the system without harm reduction in gray and dashed, the gray solid curve represents the locus where $\dot{S}|_{v \equiv 0} = 0$. In black we find the isoclines of the system with full harm reduction. The thin black downward sloping line is the line $\hat{S} = \frac{k - \mu \hat{A}}{\delta}$ from equation (3.3), along which any equilibrium of the system is located. When harm reduction is pursued with full force the inner branch of the isocline $\dot{A} = 0$ is pulled closer to the origin (switch from the gray to the black solid curve), and the isocline $\dot{S} = 0$ moves closer down to the A-axis (switch from the gray dashed to the black dashed curve). This means that application of full harm reduction pulls the high-use steady state to the lower right, increasing the equilibrium number of users and inducing less susceptibles in steady state. Switching from $v \equiv 0$ to $v \equiv v_{\max}$, the intermediate-use steady state is shifted to the upper left along the line $\hat{S} = \frac{k - \mu \hat{A}}{\delta}$. The high-use and intermediate-use steady state $(\hat{A}^3_{v \equiv 0}, \hat{S}^3_{v \equiv 0})$ and $(\hat{A}^2_{v \equiv 0}, \hat{S}^2_{v \equiv 0})$, respectively, are depicted as gray dots in Figure 3.4. The analogous fixed points of the system under static control $v \equiv v_{\max}$ are shown as black dots. The no-use steady state is not affected by a switch in static control. Hence, at $(A, S) = (0, \frac{k}{\delta})$
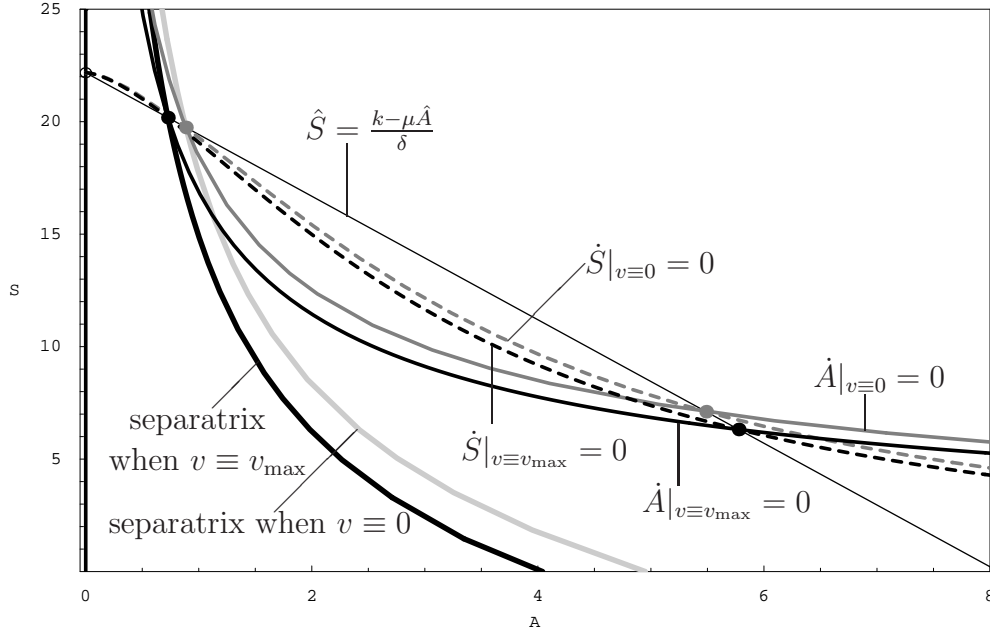
Figure 3.4: Comparison of isoclines, steady states and separatrices in the U.S. base case parameter set with static control $v \equiv v_{\max}$ and $v \equiv 0$

we encounter a circle representing both $(\hat{A}^1_{v\equiv 0}, \hat{S}^1_{v\equiv 0})$ and $(\hat{A}^1_{v\equiv v_{\max}}, \hat{S}^1_{v\equiv v_{\max}})$. In the previous section 3.5 we identified the intermediate-use steady states as saddle points. Now we look at their stable manifolds and the effect of a shift of policy emphasis.

The stable manifold of the intermediate-use steady state $(\hat{A}^2_{v\equiv 0}, \hat{S}^2_{v\equiv 0})$ is the light gray bold curve in Figure 3.4. The stable manifold of the saddle point $(\hat{A}^2_{v\equiv v_{\max}}, \hat{S}^2_{v\equiv v_{\max}})$ is the black bold curve. As mentioned in section 3.5, the probability that the initial condition is located exactly on the stable manifold is zero. Nevertheless, those curves play an important role because they act as frontiers between the basins of attraction of the stable equilibria in the system. We assume that the initial condition is located somewhere between the black and the gray manifold. If the policy maker in his one-time irrevocable decision chooses the no harm reduction regime $v \equiv 0$, the initial condition lies on the left hand side of the gray manifold (separatrix in the system with $v \equiv 0$). Hence, the epidemic converges to the no-use steady state $(\hat{A}^1_{v\equiv 0}, \hat{S}^1_{v\equiv 0})$. The drug epidemic will die out, which is most desirable. If the decision maker decides for the full harm reduction strategy, the initial condition is located to the right of the system's separatrix (black bold line). Consequently, the corresponding trajectory will lead to the high-use steady state $(\hat{A}^3_{v\equiv v_{\max}}, \hat{S}^3_{v\equiv v_{\max}})$ with almost six million drug users in equilibrium.

In order to exemplify the extreme difference in evolution of the epidemic graphically, the initial condition $(A(0), S(0)) = (0.85, 19.843)$ located in the region between the separatrices is chosen. Figure 3.5 outlines the appalling contrast between convergence to a steady state with zero users when $v \equiv 0$ is chosen, and approaching the catastrophic high-use steady state under $v \equiv v_{\max}$. The stable manifold of the saddle point in the uncontrolled system is again shown in gray, the stable manifold of the intermediate-use steady state in the system with full harm reduction in black. The short gray curve emanating from the above mentioned initial condition is the path for $v \equiv 0$. It approaches the no-use steady state with zero users $\hat{A}^1_{v\equiv 0} = 0$ and $\hat{S}^1_{v\equiv 0} = 22.177$ million susceptible individuals. This fixed point is shown as a gray dot in Figure 3.5. The black curve shows how the system evolves under the static control $v \equiv v_{\max}$. The curve eventually curls into the high-use steady state with $\hat{A}^3_{v\equiv v_{\max}} = 5.777$ million users and $\hat{S}^3_{v\equiv v_{\max}} = 6.317$ million susceptibles, which is depicted as a black dot. The gray dot close to the black high-use steady state is the high-use equilibrium $(\hat{A}^3_{v\equiv 0}, \hat{S}^3_{v\equiv 0})$.

The difference in the development of the numbers of users and susceptibles under the distinct policies $v \equiv 0$ and $v \equiv v_{\max}$ could hardly be more striking. When harm reduction is never applied, the number of users decreases and the pool of susceptibles grows towards the final steady state number. Very different to this, when harm reduction is always pursued to the full extent, the pool of susceptibles is exploited first. At the same time, use grows and overshoots the equilibrium value considerably. Only later on, use starts to ebb down a bit, but finally the transient curls into a high-use steady state. High numbers of users might induce high costs. Intuitively, one expects that the objective functional's value for the black trajectory will be larger than the amount of social costs and harm imposed on society along the short gray segment.

3.6.2 First Insights Into Effects of Full Harm Reduction

The following analyses shed light on the question for which initial conditions the full harm reduction option $v \equiv v_{\max}$ is a good alternative to the pure use reduction scenario $v \equiv 0$. Both systems are run forward numerically with $t = 200$. For the resulting trajectories $(A_{v\equiv 0}(t), S_{v\equiv 0}(t))$ and $(A_{v\equiv v_{\max}}(t), S_{v\equiv v_{\max}}(t))$ the net present values of aggregated use U and ag-
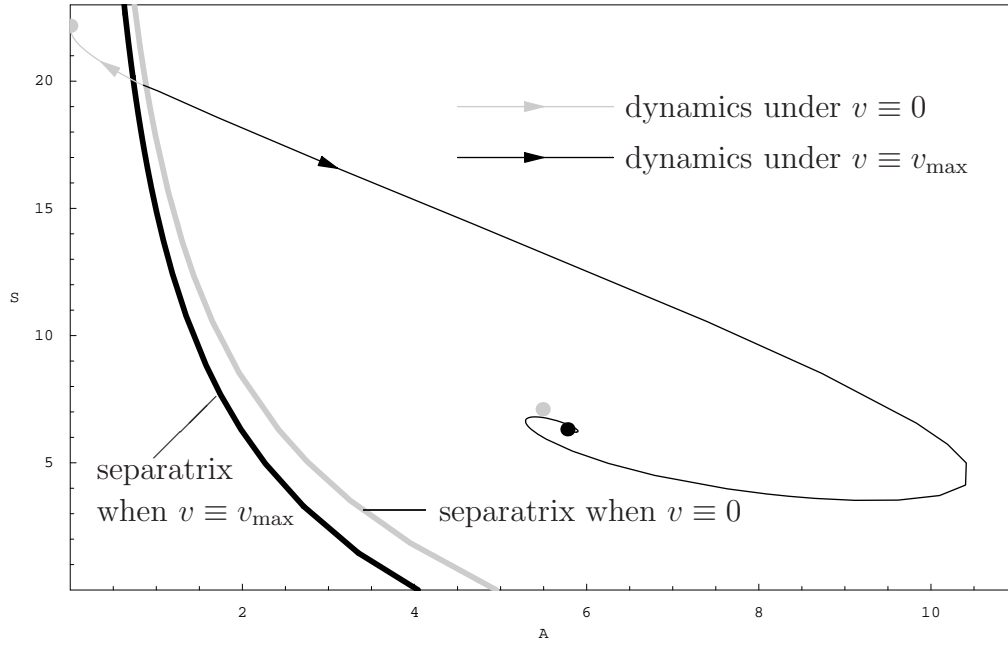
Figure 3.5: Initial condition $(A(0), S(0)) = (0.85, 19.843)$ between the separatrices in the model of the U.S. cocaine epidemic: When static control $v \equiv 0$ is chosen, the initial condition is located to the left of the system's separatrix (gray). The short gray segment shows convergence to the no-use steady state $(\hat{A}^1_{v\equiv 0}, \hat{S}^1_{v\equiv 0})$. Given the choice of full harm reduction $v \equiv v_{\max}$, the initial condition lies to the right of the system's separatrix (black). The black trajectory shows convergence to the high-use steady state $(\hat{A}^3_{v\equiv 0}, \hat{S}^3_{v\equiv 0})$.

gregated harm J

$$U|_{v\equiv 0} = \int_0^{200} e^{-rt} A_{v\equiv 0}(t) \, \mathrm{d}t,$$

$$J|_{v\equiv 0} = \int_0^{200} e^{-rt} A_{v\equiv 0}(t) \overbrace{(1-v)}^{=1} \, \mathrm{d}t,$$

$$U|_{v\equiv v_{\max}} = \int_0^{200} e^{-rt} A_{v\equiv v_{\max}}(t) \, \mathrm{d}t,$$

$$J|_{v\equiv v_{\max}} = \int_0^{200} e^{-rt} A_{v\equiv v_{\max}}(t)(1 - v_{\max}) \, \mathrm{d}t$$

are evaluated numerically. Finally, the relations

$$R_U := \frac{U|_{v\equiv v_{\max}}}{U|_{v\equiv v_0}} \quad \text{and} \quad R_J := \frac{J|_{v\equiv v_{\max}}}{J|_{v\equiv v_0}}$$

are considered.

The initial values $A(0)$ for the numbers of users vary between $A = 0$ and the high-use equilibrium number of users when harm reduction is not pursued. For Australian IDU, this is $\hat{A}^2_{v\equiv0} = 0.304916$ million users, while we have $\hat{A}^3_{v\equiv0} = 5.4888$ million cocaine users in the U.S. In order to be able to present the results for Australia and the U.S. on the same graph, the horizontal axis in Figure 3.6 gives the initial number of users $A(0)$ as proportion of the above mentioned steady states. The initial numbers of susceptibles are set to $S(0) = \frac{k-\mu A(0)}{\delta}$ (cf. equation (3.3)). Typically, the number of users tends to increase in the early stages of an epidemic, and there are many susceptibles, thus the chosen initial conditions indeed model important points in time at which harm reduction interventions could begin. The resulting ratios R_U and R_J are shown on the vertical axis of Figure 3.6.

Figure 3.6 reveals that for injecting drug users in Australia, full force harm reduction $v \equiv v_{\max}$ increases aggregate use compared to the static no harm reduction policy $v \equiv 0$ (black dashed curve, $R_U > 1$), but it reduces aggregate harm in the society (gray dashed curve, $R_J < 1$). The reduction is considerable, total harm is cut down by 44-49% for most of the initial conditions. This goes along with an increase by 8 to a maximum of 20% in drug use. For cocaine use in the U.S., aggregated harm (gray curve) is reduced although total use (black curve) increases, when initial conditions involve $A(0)$ smaller than about 10.8% and larger than about 23% of the steady state value $\hat{A}^3_{v\equiv0} = 5.4888$. For the intermediate region, the choice of static full force harm reduction $v \equiv v_{\max}$ does not only increase aggregated drug use, but also aggregated harm. The increases are dramatic for some initial conditions. The maximum increase in total harm and total use was computed for the initial condition $(A(0) = 0.808, S(0) = 19.9585)$ and amounts to $R_J = 5.54$, which means a more than fivefold increase in aggregated use. Aggregated harm multiplies with the factor 4.57. This initial condition is quite close to the initial condition $(A(0) = 0.85, S(0) = 19.843)$ chosen for the exposition in Figure 3.5. Hence, the expectation that in Figure 3.5 the black trajectory is worse than the gray one is now proven. Please note that the terrifying high increases in both use and harm occur only for initial conditions that are located between or close around the separatrices. The results are a clear sign that for initial conditions around there, the U.S. drug policy makers might better shy away from a harm reduction policy. Nevertheless, for a large region of other initial values of users $A(0)$ on the line, application of full harm reduction $v \equiv v_{\max}$ indeed reduces aggregated harm, hence such interventions cannot be demonized in general.
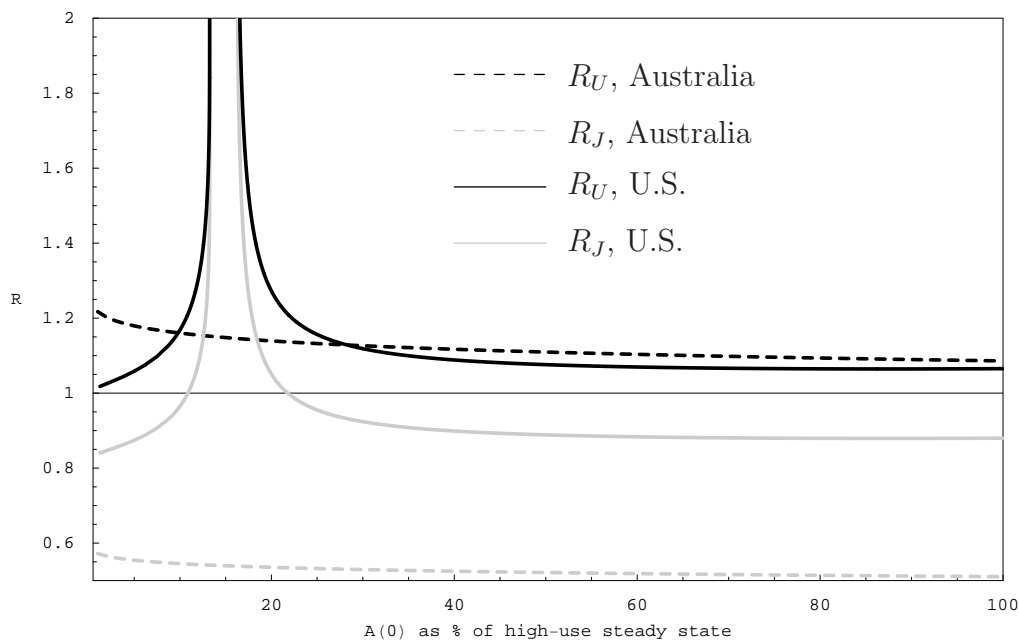
Figure 3.6: Effects of application of full harm reduction $v \equiv v_{\max}$ compared to a static use reduction policy $v \equiv 0$ on total aggregated use U and harm J, measured in terms of the ratios R_U and R_J for the U.S. cocaine epidemic (solid) and IDU in Australia (dashed).

3.6.3 Comprehensive Investigation of Effects of Full Harm Reduction

Next, a more comprehensive answer to the question "For which initial conditions does full harm reduction lead to higher value of aggregated harm than does the pure use reduction regime?" is given.

Figure 3.7 gives this answer graphically for the U.S. cocaine epidemic's base case parameterization. The region colored in gray depicts initial conditions $(A(0), S(0))$ where $R_J > 1$, which means that $J|_{v \equiv v_{\max}}$ evaluates higher total harm than does $J|_{v \equiv 0}$. Figure 3.7 also redraws the separatrices of the systems and the trajectories picked to show the possibility of tipping the epidemic to a high-use equilibrium. The gray region is somewhat broader than the sliver between the tipping point curves, but it still includes only a modest share of the state space.

Figure 3.8 displays the results from the analogous investigation for Australian IDU. For any initial condition shown in the Figure, the ratio of total aggregated harm evaluates to $R_J < 1$. The maximum of the computed ratios
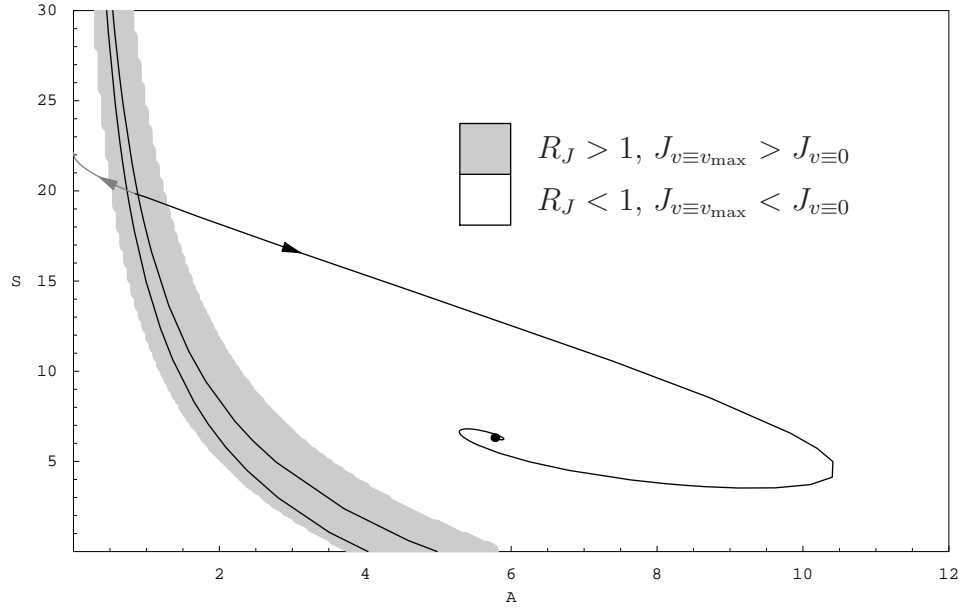
Figure 3.7: Distinction of $R_J > 1$ (gray) and $R_J < 1$ (white) for several initial conditions of the U.S. cocaine epidemic on the (A, S)-plane.

is 0.5947, the minimum is 0.4870. The ratio is higher, the closer the underlying initial condition is located to the S-axis. The different gray shadings in Figure 3.8 give different ranges for the ratios. Nevertheless, $J_{v \equiv v_{\max}} < J_{v \equiv 0}$ holds everywhere. When applying a static drug control policy to IDU in Australia, the full harm reduction regime $v \equiv v_{\max}$ is always preferred to the control regime $v \equiv 0$ without harm reduction.

The results for static control of Australia's IDU population are intuitively clear. The health-related share of social cost stemming from drug abuse is high for that drug epidemic. When use is widespread, the harms felt by users are correspondingly high. A full harm reduction intervention is appealing because it considerably reduces the amount of harm (the harms felt by the users account for 53% of the objective function). No matter where on the (A, S)-plane the initial conditions are located, on the way to a steady state, the drug epidemic will always progress through stages of high use, namely when the epidemic is close to the steady state or if it overshoots the equilibrium number of users. The static full harm reduction regime has the benefit of this huge reduction in harms, whereas the static pure use reduction regime does not. In the short run, it seems irrational to apply the initiation-increasing policy of harm reduction at low levels of use, but the static control choice $v \equiv v_{\max}$ shows its merits in the long run,

although the steady state level of users is higher and the transient under the full harm reduction policy overshoots the levels of use occurring under the pure use reduction regime. As an example, Figure 3.8 shows two trajectories emanating from $(A(0), S(0)) = (0.02, 0.95)$. The solid trajectory shows the path for static control $v \equiv 0$, the dashed one shows the evolution of the system when applying $v \equiv v_{\max}$. The arrows indicate the convergence towards the steady states of the statically controlled system, which are shown as black dots.
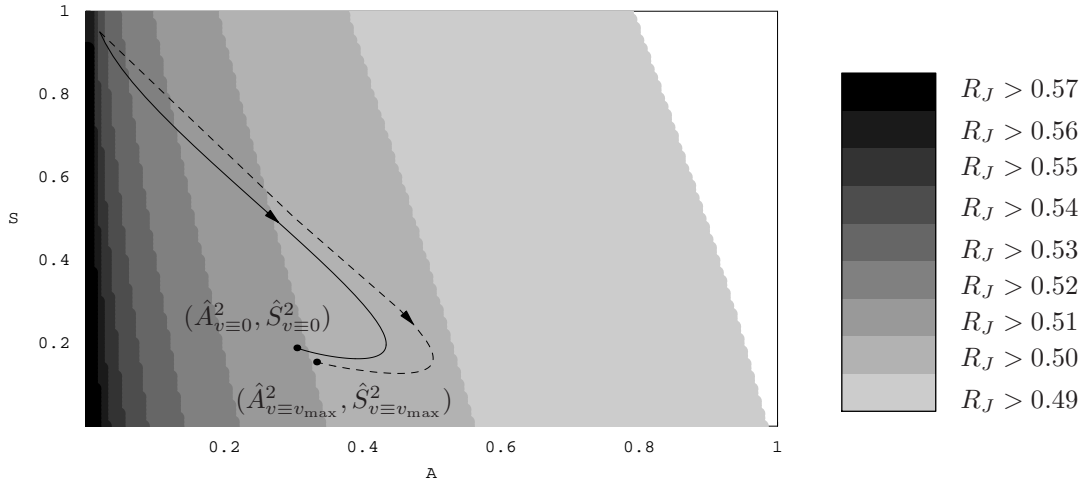

Figure 3.8: $R_J < 1$ for several initial conditions in the (A, S)-plane for IDUs in Australia. The dashed trajectory outlines the overshoot in number of users when $v \equiv v_{\max}$ is applied compared to $v \equiv 0$ (solid trajectory).

3.6.4 Conclusions from the Comparative Analysis in the Base Case

Recapitulating the results from the different steps of the investigation of the static options $v \equiv v_{\max}$ and $v \equiv 0$, the results derived for the U.S. cocaine epidemic are less in favor of harm reduction than the results for Australian IDU. The first finding of the comparative study was that applying the full harm reduction policy $v \equiv v_{\max}$ for an initial condition located between the

separatrices existing in the case of the U.S. cocaine epidemic's parameter set tips the epidemic towards a high-use steady state.

At this step, we did not yet know if the aggregated harm along the trajectories leading to the high-use steady state was indeed higher than when the system evolves under a pure use reduction policy. A comparison of total aggregated use and harms for initial conditions located on a single line revealed that for the U.S. cocaine epidemic there exists a region for which application of full harm reduction $v \equiv v_{\max}$ increases total use and total harm, dramatically so for certain initial conditions.

Inspection of total harm $J_{v \equiv v_{\max}}$ and $J_{v \equiv 0}$ for a large set of initial conditions led to the gray region in Figure 3.7. Starting static control in this gray region, the total aggregated harm under a static harm reduction policy $v \equiv v_{\max}$ exceeds the total aggregated harm when a pure use reduction regime $v \equiv 0$ is pursued. The gray region is broader than the domain between the stable manifolds. This reveals that even if harm reduction does not tip the epidemic, it may affect the course of the epidemic negatively, such that more harm is accumulated compared to the pure use reduction policy.

Looking at the diverging trajectories shown e.g. on Figure 3.7, the strong contrast in steady state numbers of users $\hat{A} = 0$ and $\hat{A} = 5.77$ million, and the gray region shown in Figure 3.7, the immediate conclusion is that even the modest amount ($v_{\max} = 17.4\%$) of harm reduction available in the U.S. could have terrible and long-lasting effects on the cocaine epidemic.

However, having a look on the results from a less nervous (fearful) point of view, the "dangerous" initial conditions where full harm reduction tips the epidemic are located in a quite narrow sliver of the (A, S)-plane. The gray region where $J_{v \equiv v_{\max}} > J_{v \equiv 0}$ is a bit broader, but still a rather narrow region in the entire state space. Furthermore, for any initial condition located to the left of the separatrix of the system with $v \equiv 0$ (gray bold curve in Figure 3.4), the population of drug users $A(t)$ declines under the pure use reduction regime. Confronted with a drug epidemic at such a stage (rather modest number of users, declining number of users), a policy maker might most probably not think about new drug control policies like harm reduction. This argument shows that from about two third of the initial conditions that were encountered as problematic, there is actually no great danger. Only the sliver on the right outside the region between the separatrices remains. In this region, no matter if harm reduction is implemented to the full possible extent $v \equiv v_{\max}$ or if the decision maker sticks to the pure use reduction regime $v \equiv 0$, the system will converge to a high-use steady state. Static full harm reduction leads to higher total social cost along the epidemic's

transients, making it the worse option. Note that this is only the case for $(A(0), S(0))$ close to the separatrix. For initial conditions further to the right, the static policy $v \equiv v_{\max}$ is preferred. Although this softens many of the concerns against harm reduction, ill-timed harm reduction clearly has negative effects.

The most important conclusion from the static analysis is that harm reduction seems to have great potentials to ameliorate social costs and harms associated with injecting drug use in Australia. For the base case scenario of the U.S. cocaine epidemic, it seems to have great potential at the early stages of the epidemic - when drug use is rather rare and there is no danger that initiation will explode - or when the number of drug users is high in the later stages of the epidemic. At these levels of high use even the modest reduction of 17% beneficially cuts down social cost and harms in society. Nevertheless, it shall be emphasized that if control intervention is timed incorrectly, it may tip the epidemic and/or induce increased aggregated harm.

3.7   Parameter Variations

In drug-related modeling, parameterizations may be tenuous. Indeed, the present parameterizations are not precise. Therefore, analyses of the sensitivity of results with respect to parameter variations are indispensable.

The first scenario investigated is that the infectivity of the U.S. cocaine epidemic dropped. In terms of the parameters this means that the original value $b = 0.009$ is replaced by the lower $b = 0.0075$, whereas the other parameter values from the U.S. base case remain unchanged. On the one hand the choice of a decreased infectivity of cocaine in the U.S. is based on the fact that this modification changes results dramatically compared to the base case. On the other hand, there is indeed a possibility that the virulence of the epidemic has declined in the later stages of the cocaine epidemic (Tragler et al., 2001; Caulkins et al., 2004; Johnson et al., 1996).

The factor $\omega = 0.5$ in the base case parameterization is rather arbitrary, thus we allow for the further modification to $\omega = 1$. That change means that users will now fully factor their non-monetary costs of drug use into their consumption decision. In the base case, only a proportion of 50% was recognized by them, the rest was regarded to be an externality from the point of view of the users. With this change, the effects modeled by the function $g(v)$ are more pronounced.

Analogously to the base case investigation conducted in the preceding

section 3.6, we first analyze the ratios R_J and R_U for initial conditions located on the line $S = \frac{k - \mu A}{\delta}$ (cf. equation (3.3)) between the no-use steady state and the high-use steady state of the system without harm reduction. Then we investigate the stable manifolds of the saddle point steady states and compute the ratio R_J for a larger set of initial conditions. The respective results from the base case parameterization of the U.S. cocaine epidemic are shown in panels a) and b) of Figure 3.9 as a reference.

Panel c) of Figure 3.9 reveals that for initial conditions on the line $S(0) = \frac{k - \mu A(0)}{\delta}$, full harm reduction always increases total use (gray curve). Furthermore, it increases total harm (black curve) for any initial number of users greater than about 15% of the uncontrolled high-use steady state, which has now about $\hat{A}^3_{v \equiv 0} = 4.14$ million users. Harm reduction is a good choice at the early stages of the epidemic, but only then.

Panel d) in Figure 3.9 shows the stable manifolds of the intermediate-use saddle points of the system. The tipping point region between the separatrices is now considerably broader than in the base case. Additionally, the gray region where full harm reduction $v \equiv v_{\max}$ counterproductively affects the system in terms of $R_J > 1$ is substantially larger than in the base case. Note that the high-use steady state of the uncontrolled model is now located at $\hat{A}^3_{v \equiv 0} = 4.1366$, $\hat{S}^3_{v \equiv 0} = 9.9940$. In panel d) it is shown as a little dark gray dot. Use is widespread there, the epidemic has reached an endemic level. Intuitively, we would expect that harm reduction's merits in cutting down harms and social cost via the objective function are appealing in such circumstances. Counterintuitively, the steady state happens to fall in the gray region, where full harm reduction is worse than pure use reduction. We conclude that even if the epidemic has reached levels around a high-use steady state of the system under the pure use reduction, harm reduction is not necessarily beneficial. Use must grow a whole lot more, until the white region of save application of $v \equiv v_{\max}$ is reached.

From an economic point of view - seen through the eyes of a user - it is plausible that in the stylized example harm reduction is a poor choice for a broader range of initial conditions. The larger the proportion ω, the more do drug users factor health-related costs into their decision to consume. Thus, in the fraction involved in the function $g(v)$, the denominator will be large, when ω is large. Full harm reduction decreases the fraction under the elasticity radical in $g(v)$ when ω is larger. The large ω says that potential users fully recognize any economic reduction of health-related cost of use, and full harm reduction gives a large economic benefit, which considerably fuels up initiation.
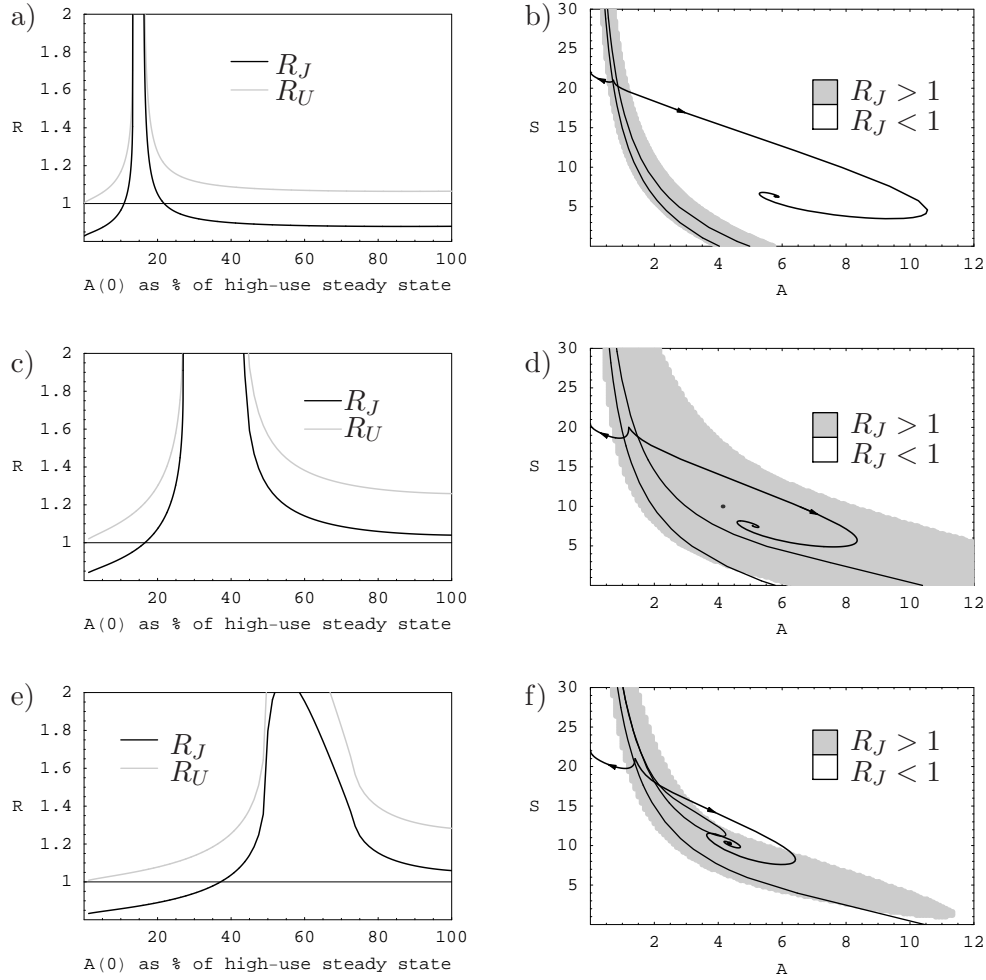
Figure 3.9: Overview of the results for the static system under parameter changes. Panels a) and b) show the results for the base case with $b = 0.009$, $\omega = 0.5$. Panels c) and d) present the results for parameter values $b = 0.0075$, $\omega = 1$. Panels e) and f) refer to the case of $b = 0.0065$, $\omega = 0.5$.

The next step of the analysis decreases infectivity a bit more to $b = 0.0065$. The other parameters remain unchanged, i.e. we use $\omega = 0.5$ again. Panel e) in Figure 3.9 shows that for initial conditions along the line $S(0) = \frac{k - \mu A(0)}{\delta}$ total use increases for any initial condition, but at the early stages of the epidemic, full harm reduction $v \equiv v_{\max}$ can bring benefits. The peaks in R_J and R_U are less pronounced than in the base case.

Panel f) shows what happens to the stable manifolds. The separatrix in the system with $v \equiv v_{\max}$ is still a curve of the same shape as in the examples before, though shifted further to the right. In the system with $v \equiv 0$, the separatrix forms a pocket around the high-use steady state at $\hat{A}^3_{v\equiv 0} = 3.3521, \hat{S}^3_{v\equiv 0} = 12.9739$. For initial conditions that happen to fall inside the pocket, static control $v \equiv 0$ will lead to this high-use steady state. Under $v \equiv 0$, initial conditions located outside the pocket exhibit transients that will eventually approach a no-use steady state. Consequently, for most initial conditions, eradication of the epidemic is possible with the pure use reduction strategy. Choosing an initial condition outside the pocket and on the right of the left separatrix, the static no-harm reduction policy $v \equiv 0$ leads to convergence to zero use, whereas static full harm reduction $v \equiv v_{\max}$ lets the system converge to a high-use steady state. The region where harm reduction tips the epidemic to a high-use equilibrium is now extended to a considerable part of the (A, S)-plane. Moreover, we evaluated the ratio R_J for a large set of initial conditions. The results are shown by the usual shading in white and gray in panel f). The large tipping point region is now split up in a gray part and a white part. The gray part, where full harm reduction actually increases J, is smaller, the larger part of the tipping point region is shaded white. This means that even when harm reduction tips the epidemic to a high-use steady state, its application yields less social cost and harms than the no harm reduction regime under which zero use can be approached. When the epidemic has grown beyond a certain number of users and susceptibles, it is the better option to accommodate the existing drug use problems by the new harm reduction mechanism and cut down harm and social cost to the maximum possible extent.

3.8 Conclusions from the Model with Static Control

The numbers of users and susceptibles in the steady states determined in the preceding sections are realistic. Remaining in the first quadrant is plausible for a drug epidemic, and the model nicely captures this property. By

construction of the base case initiation function $I(A, S, v) = bA^\alpha Sg(v)$, initiation cannot be negative. The stability properties of the steady states are "stable" or "saddle", instable fixed points did not occur. Altogether, the model's properties are satisfactory and we do not need to adapt functional forms and re-estimate parameters to expunge unwanted features. All this clearly speaks for our model.

A major comment with respect to no-use steady state in the uncontrolled model is that it is viewed as unrealistic by many people. They interpret this steady state in the sense that the cocaine epidemic dies out by itself, and that the decision maker does not have to undertake policy interventions to ameliorate problems that stem from illicit drug consumption. With respect to the present model, one has to take into account that the control $v(t) = 0$ means that the policy maker decides not to apply harm reduction as a drug control tool at time t. Still, the system is in the pure use reduction regime. This means that even if not modeled explicitly in our current approach, in a system with $v(t) = 0$ something is done to fight the drug problem in the society, namely classic drug control policy (law enforcement, prevention, treatment) is applied.

The term "uncontrolled model" is used some times in the text. It is important to emphasize that this does not allude to drug legalization. The word "uncontrolled" merely states that the control variable $v(t)$ of the present model is set equal to zero.

We conclude the static analysis with the observation that whether static full harm reduction $v \equiv v_{\max}$ is beneficial or not depends on a lot of factors. The particular drug, the specific country and the actual parameter values influence the results on potential merits and dangers of harm reduction interventions. Given the parameters shown in Table 2.1 for Australia, static full harm reduction $v \equiv v_{\max}$ is to be preferred in comparison to a static pure use reduction regime where $v \equiv 0$. Policy recommendations for the U.S. cocaine epidemic are a more sensitive endeavor, even more because the parameter variations in the previous section do not lead to unitary results.

Static analysis is important and revealed interesting features of the model. We have seen that results are sensitive to parameter changes, which points to the need of sensitivity analysis and bifurcation investigations. Such analyses are conducted in Chapter 5. We found that at certain stages of a drug epidemic, full harm reduction has great advantages, whereas for other initial conditions it can tip the epidemic to non-desired levels of use and/or negatively affect transients in terms of increased aggregated harm. The distinction of various stages of the epidemic and the insight that for distinct

stages of the epidemic appropriate control interventions may be different, are crucial in a system, in which social costs are proportional to the number of users, and in which the number of users over time is affected by the control interventions. This leads directly to the importance of investigation of the optimal dynamic control version of the model. It is analyzed in detail in Chapter 4.

Chapter 4

Base Case Models with Optimal Control

4.1 The Optimal Control Model

In the previous Chapter, control was determined at time $t = 0$ in form of a so-called one-shot policy. This means that at the beginning of the planning horizon, the decision maker chooses among the options $v \equiv 0$ (pure use reduction) and $v \equiv v_{\max}$ (full harm reduction). The chosen policy is then applied once and for all time. In an optimally controlled system, the control intervention is allowed to vary dynamically over time, i.e. obeying the non-negativity condition

$$v(t) \geq 0 \quad \forall\, t, \tag{4.1}$$

and the upper boundary control condition

$$v_{\max} - v(t) \geq 0 \quad \forall\, t, \tag{4.2}$$

the control $v(t)$ can take any value between $v(t) = 0$ and $v(t) = v_{\max}$. Of special interest are time spans where interior control values $0 < v_i(t) < v_{\max}$ occur. This is where the harm reduction policy is introduced and implemented, where only certain harm reduction programs are done, or where harm reduction is reversed. This may for example happen when certain measures are discontinued because they trigger adverse behavioral responses in the current community of drug users and susceptibles. In what follows, v_i refers to interior values of control $v(t)$.

Please remember that the upper boundary value $v_{\max}$ varies for specific drugs and countries and can depend on exact definitions of the term harm

and whose harms get counted. The bounds of the feasible control region play an important role in the optimal control solutions of the current model.

The optimization problem is first solved for the base case model, where initiation into drug use is exclusively driven by current users who recruit new users from the pool of susceptibles and control costs are assumed to be negligible. The base case parameterization for the U.S. and Australia can be found in Table 2.1 in section 2.6.

The optimal control version of the SA Harm Reduction Model in the base case formulation is

$$\min_{0 \le v(t) \le v_{\max}} \left\{ J = \int_0^\infty e^{-rt} \Big(A(t) \big(1 - v(t)\big) \Big) \mathrm{d}t \right\} \tag{4.3}$$

subject to the system dynamics for A and S with initial condition (A_0, S_0)

$$\dot{A}(t) = b\, A(t)^\alpha\, S(t)\, g(v(t)) - \mu\, A(t), \qquad A(0) = A_0, \tag{4.4}$$
$$\dot{S}(t) = k - \delta\, S(t) - b\, A(t)^\alpha\, S(t)\, g(v(t)), \;\; S(0) = S_0, \tag{4.5}$$

and to the boundary conditions (4.1) and (4.2) for the control variable v.

It is a nonlinear, autonomous, continuous-time optimization problem with infinite time horizon, has two states and one control. The decision maker's goal is to minimize the discounted stream of social costs (interpreted as harms borne by the society) caused by illicit drug use.

Optimal control problems can be solved with Pontryagin's Maximum Principle (see, e.g. Feichtinger & Hartl, 1986; Leonard & Long, 1992; or more recently formulated in Grass et al., 2008).

To enhance readability, we will mostly omit the time argument t in what follows.

Note that the problem formulation is given as a minimization problem in equation (4.3). Analogously to the Maximum Principle the Minimum Principle could be stated, but here for personal convenience we stick to the Maximization notation. Consequently, we simply switch to maximization of the negative objective cost functional:

$$\max_{0 \le v \le v_{\max}} -J.$$

Next, the costate variables λ_A (costate of A) and λ_S (costate of S) are introduced. The current-value Hamiltonian H is given by

$$H(A, S, v, \lambda_A, \lambda_S) = -\lambda_0 \big(A(1 - v)\big) + \lambda_A\, \dot{A} + \lambda_S\, \dot{S}. \tag{4.6}$$

We deal with an infinite planning horizon, thus the additional constant multiplier $\lambda_0 \geq 0$ has to be associated with the integrand of the objective function. For $\lambda_0 \neq 0$, $\lambda_0 = 1$ can be set without loss of generality. Among others, Leonard and Long (1992) point out that $\lambda_0 = 0$ is the case only in pathological examples. For the sake of completeness, we show in the Appendix that $\lambda_0 = 1$ may be used for the base case model. This result can be easily transferred to the slightly modified model formulations in Chapter 6. In what follows, we will hence omit the factor $\lambda_0 = 1$.

We proceed stating the necessary optimality conditions for solutions of the infinite time horizon problem with boundary control constraints. In order to do so, we first define the Lagrangian function

$$L(A, S, v, \lambda_A, \lambda_S, \pi_1, \pi_2) = H(A, S, v, \lambda_A, \lambda_S) + \pi_1 \, v + \pi_2 \, (v_{\max} - v), \quad (4.7)$$

where π_1 and π_2 denote the Lagrange Multipliers.

The necessary optimality conditions that have to be satisfied by an optimal solution of the optimal control problem with boundary conditions are:

$$H(A^*, S^*, v^*, \lambda_A, \lambda_S) = \max_{0 \leq v \leq v_{\max}} H(A^*, S^*, v, \lambda_A, \lambda_S) \quad (4.8)$$

$$L_v = 0 \quad (4.9)$$

$$\dot{\lambda}_A = r\lambda_A - L_A \quad (4.10)$$

$$\dot{\lambda}_S = r\lambda_S - L_S \quad (4.11)$$

$$\pi_1 \geq 0, \ \pi_1 v^* = 0 \quad (4.12)$$

$$\pi_2 \geq 0, \ \pi_2 \, (v_{\max} - v^*) = 0. \quad (4.13)$$

These conditions are necessary conditions, thus they only help for determination of candidates for the optimal solutions. But we need to ensure that we indeed obtain maxima. Unfortunately, the Mangasarian sufficiency conditions for optimality are not fulfilled. They are stated in Appendix A.3 and applied to the current optimal control problem. Still, due to the fact that the steady states of the dynamics are confined to the line $\hat{S} = \frac{k - \mu \hat{A}}{\delta}$ (see equation (3.3)) in the (A, S)-plane, we can be quite sure that the solutions derived here are indeed optimal.

Please note that in equations (4.10) and (4.11) L_A and L_S denote the derivative of the Lagranian function L with respect to the states A and S, respectively.

The limiting transversality conditions for the costates λ_A and λ_S are

$$\lim_{t \to \infty} e^{-rt} \lambda_A(t) \, A(t) \;=\; 0,$$
$$\lim_{t \to \infty} e^{-rt} \lambda_S(t) \, S(t) \;=\; 0.$$

They hold if the states approach a stable steady state, which will be the case in the optimal control solutions.

Equations (4.10) and (4.11) are the so-called costate equations. Together with the state dynamics (4.4) and (4.5), they build the so-called canonical system.

The complementary slackness conditions (4.12) and (4.13) state that for an interior solution $0 < v^* < v_{\max}$ the Lagrange Multipliers have to be zero. So, if we restrict ourselves to the inner region of the feasible domain, we have $\pi_1 = 0$ and $\pi_2 = 0$. The Lagrangian function L is reduced to the Hamiltonian function H. The necessary optimality condition is the Hamiltonian maximizing condition of the control, given by

$$v^* = \arg\max_v H. \tag{4.14}$$

Assuming that the optimal solution is located in the interior of the feasible control domain we substitute v^* derived from $H_v = 0$ into the four-dimensional canonical system of the model, set the equations equal to zero and solve the system simultaneously.

The derivative of the Hamiltonian H with respect to v is

$$H_v = A + b \, A^\alpha \, S \, (\lambda_A - \lambda_S) \, g'(v).$$

Investigating the concavity of the Hamiltonian H with respect to control v, it is easy to see that due to the fact that the function $g(v)$ is convex, H is concave if $\lambda_A - \lambda_S < 0$ holds.

Claiming $H_v = 0$, we can then solve for v. This yields

$$\frac{-A}{b \, A^\alpha \, S \, (\lambda_A - \lambda_S)} = g'(v) = \frac{-c_s \, \omega \, \eta}{c_m + c_s \, \omega \, v_{\max}} \left(\frac{c_m + c_s \, \omega \, (v_{\max} - v)}{c_m + c_s \, \omega \, v_{\max}} \right)^{-1+\eta}.$$

To avoid lengthy formulas, we define

$$\Psi := \frac{A}{b \, A^\alpha \, S \, (\lambda_A - \lambda_S)}, \quad \Phi := c_m + c_s \, \omega \, v_{\max}, \tag{4.15}$$

and get

$$v^* = \frac{\Phi}{-c_s\omega}\left(\left(\frac{\Psi\Phi}{c_s\omega\eta}\right)^{\frac{1}{-1+\eta}} - 1\right). \tag{4.16}$$

Next, we tackle the explicit formulation of the canonical system for the model. The four differential equations read

$$\begin{align}
\dot{A} &= b\,A^\alpha\,S\,g(v) - \mu\,A, \tag{4.17}\\
\dot{S} &= k - \delta\,S - b\,A^\alpha\,S\,g(v), \tag{4.18}\\
\dot{\lambda}_A &= 1 - v + (r+\mu)\lambda_A + b\alpha A^{\alpha-1}S(\lambda_S - \lambda_A)\,g(v), \tag{4.19}\\
\dot{\lambda}_S &= (r+\delta)\lambda_S + b\,A^\alpha\,(\lambda_S - \lambda_A)\,g(v). \tag{4.20}
\end{align}$$

In what follows, the steady states of the model are computed and analyzed with respect to optimality.

4.2 Searching for Steady States with Interior Control

We first search for steady states with interior control, substituting v^* from equation (4.16) into the canonical system (4.17) - (4.20). The resulting four equations are set equal to zero and the system is solved simultaneously for $\hat{A}$, $\hat{S}$, $\hat{\lambda}_A$ and $\hat{\lambda}_S$. Then we can substitute those results back into equation (4.16) for optimal control v^*.

4.2.1 Australia

Conducting analytical investigations, which are not presented here for the sake of conciseness, it can be shown that steady states with $0 < \hat{v}_i < v_{\max}$ do not exist.

4.2.2 United States

For the U.S. parameter set the following steady state values are computed:

$$\begin{pmatrix} \hat{A}_1 \\ \hat{S}_1 \\ \hat{\lambda}_{S,1} \\ \hat{\lambda}_{A,1} \end{pmatrix} = \begin{pmatrix} 0.059236 \\ 22.014230 \\ -0.909179 \\ -0.004025 \end{pmatrix}$$

with control $\hat{v}_1 = 1.047218$, which is larger than the upper bound, and

$$\begin{pmatrix} \hat{A}_2 \\ \hat{S}_2 \\ \hat{\lambda}_{A,2} \\ \hat{\lambda}_{S,2} \end{pmatrix} = \begin{pmatrix} 4.212853 \\ 10.610664 \\ -33.954817 \\ -13.453296 \end{pmatrix}$$

with control $\hat{v}_2 = -0.684428$, which lies to the left of the feasible control domain.

The steady states with v^* from equation (4.16) violate the control constraint $0 \le v \le v_{\max}$ in both the case of U.S. cocaine and Australian IDU. Hence we conclude that the optimal solution in both cases will involve boundary control steady states of the system. Those steady states are analyzed in the next sections.

4.3 Boundary Control Steady States and Lagrange Multipliers

For determination of boundary control steady states the four equations of the canonical system (4.17) - (4.20) are set equal to zero simultaneously for $\hat{v} = 0$ and $\hat{v} = v_{\max}$, respectively. The steady state values $\hat{A}$ and $\hat{S}$ in this section are the ones derived in sections 3.2 and 3.4 devoted to the statically controlled system. Dealing with the optimally controlled system the steady state values $\hat{A}$ and $\hat{S}$ are complemented by the steady state costate values $\hat{\lambda}_A$ and $\hat{\lambda}_S$.

The economic interpretation of the costate variables is that they give a shadow price of an increment in the associated state. Users and susceptibles are bad in the current model, because they cause costs (the users do so directly, whereas the susceptibles are possible future users). Consequently, we expect negative or at least non-positive costates λ_A and λ_S in the current model.

Calculating steady states with boundary control, the Lagrange Multiplier of the active constraint has to be investigated. The complementary slackness conditions (4.12) and (4.13) state that only steady states at which the associated Lagrange Multiplier is non-negative can be candidates for the optimal solution of the current problem.

The Lagrange Multipliers for the current optimal control problem are

$$\pi_1 = -A - b\,A^\alpha\,S\,(\lambda_A - \lambda_S)\,g'(0), \tag{4.21}$$
$$\pi_2 = A + b\,A^\alpha\,S\,(\lambda_A - \lambda_S)\,g'(v_{\max}). \tag{4.22}$$

4.3.1 Australia

We start with the parameterization for Australian IDU and upper boundary control. Substituting $v = v_{\max}$ into equations (4.17)-(4.20), we set them equal to zero simultaneously. The complementary slackness conditions (4.12) and (4.13) state that the Lagrange Multipliers have to be non-negative for an optimal solution. In steady states with upper boundary control $\hat{v} = v_{\max}$, the lower boundary constraint $v \geq 0$ is not active. So, for its Lagrange Multiplier, $\pi_1 = 0$ holds automatically. $\pi_2 \geq 0$ must be fulfilled to make a steady state a candidate for optimality.

We numerically determine the following fixed points $(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S)$ and their Lagrange Multipliers π_2:

$$\hat{E}^1_{\hat{v}=v_{\max}} = \begin{pmatrix} 0.333455 \\ 0.154617 \\ -3.95724 \\ -2.55002 \end{pmatrix} \quad \text{with} \quad \pi_2 = 0.304802,$$

$$\hat{E}^2_{\hat{v}=v_{\max}} = \begin{pmatrix} 0 \\ 0.552521 \\ 0.076176 \\ 0 \end{pmatrix} \quad \text{with} \quad \pi_2 = 0.$$

The steady state $\hat{E}^1_{\hat{v}=v_{\max}}$ is feasible. The discussion of $\hat{E}^2_{\hat{v}=v_{\max}}$ in the next subsection excludes this steady state.

For lower boundary control $\hat{v} = 0$, we get the following steady states $(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S)$ and Lagrange Multipliers π_1:

$$\hat{E}^1_{\hat{v}=0} = \begin{pmatrix} 0.304916 \\ 0.188672 \\ -8.92372 \\ -5.13915 \end{pmatrix} \quad \text{with} \quad \pi_1 = -0.265058,$$

$$\hat{E}^2_{\hat{v}=0} = \begin{pmatrix} 0 \\ 0.552521 \\ 0.237137 \\ 0 \end{pmatrix} \quad \text{with} \quad \pi_1 = 0,$$

where the first one has a negative Lagrange Multipliers π_1 and is thus not among the candidates for the optimal long-run solution. The second steady state $\hat{E}^2_{\hat{v}=0}$ is discussed in the subsequent section.

4.3.2 Considering the No-Use Steady States in Depth

The striking feature of the no-use steady states $\hat{E}^2_{\hat{v}=v_{\max}}$ and $\hat{E}^2_{\hat{v}=0}$ are the positive values of the costate $\hat{\lambda}_A$. In its economic interpretation, the costate variable gives the shadow price of the associated state. This means that it quantifies the internal value of an infinitesimal increment in the state. For the current problem's formulation, we clearly expect the shadow price of use to be negative.

To clear the discrepancy about the no-use steady states, we set the costate equations (4.19) and (4.20) simultaneously equal to zero and solve for $\hat{\lambda}_S$ and $\hat{\lambda}_A$. This yields

$$\hat{\lambda}_S = \frac{b\hat{A}^{1+\alpha}(v-1)g(v)}{\hat{A}(r+\delta)(r+\mu) - b\alpha(r+\delta)\hat{S}\hat{A}^\alpha g(v) + b(r+\mu)\hat{A}^{1+\alpha}g(v)},$$

$$\hat{\lambda}_A = \frac{\hat{A}(v-1)(r+\delta+b\hat{A}^\alpha g(v))}{\hat{A}(r+\delta)(r+\mu) - b\alpha(r+\delta)\hat{S}\hat{A}^\alpha g(v) + b(r+\mu)\hat{A}^{1+\alpha}g(v)}.$$

In order to investigate the costates when the steady state at $(\hat{A}, \hat{S}) = (0, \frac{k}{\delta})$ is approached, we set $S = \frac{k}{\delta} + d \cdot A$, where $d \in \mathbf{R}$, and then consider the expressions for the costates for $A \to 0$. With the help of d convergence to the no-use steady state can be analyzed coming from any direction from the first quadrant of the (A, S)-plane. Setting $D := r + \delta$ and $M := r + \mu$ we get

$$\lambda_S = \frac{b\delta A^{1+\alpha}(v-1)g(v)}{A\delta DM - bk\alpha DA^\alpha g(v) - b\delta A^{1+\alpha}g(v)(d\alpha D - M)},$$

$$\lambda_A = \frac{\delta A(v-1)(r+\delta+bA^\alpha g(v))}{A\delta DM - bk\alpha DA^\alpha g(v) - b\delta A^{1+\alpha}g(v)(d\alpha D - M)}.$$

For A tending to zero, denominator and numerator both converge to zero, which yields the indeterminate form "$\frac{0}{0}$". Unfortunately, L' Hôpital's rule does not help here. A trick often used is manipulation of the expression, such that common factors cancel out, giving a new fraction which is no longer of indeterminate form. Here, we can extract A^α and obtain

$$\lambda_S = \frac{A^\alpha b\delta A(v-1)g(v)}{A^\alpha[A^{1-\alpha}\delta DM - bk\alpha Dg(v) - b\delta Ag(v)(d\alpha D - M)]}.$$

The factor A^α is in the numerator and the denominator and cancels out. Remember that $\alpha < 1$ for Australia. Taking the limit for $A \to 0$, we finally get $\hat{\lambda}_S = 0$ because

$$\lim_{A \to 0} \lambda_S = \frac{b\delta(v-1)g(v)\overbrace{\lim_{A \to 0} A}^{=0}}{\delta DM \underbrace{\lim_{A \to 0} A^{1-\alpha}}_{=0} -bk\alpha Dg(v) + b\delta g(v)(d\alpha D - M)\underbrace{\lim_{A \to 0} A}_{=0}}$$

$$= 0.$$

Please note that the convergence of the third term in the denominator is independent of the constant d.

The equation for λ_A is modified in the same way

$$\lambda_A = \frac{A^\alpha \delta A^{1-\alpha}(v-1)(r + \delta + bA^\alpha g(v))}{A^\alpha[A^{1-\alpha}\delta DM - bk\alpha Dg(v) - b\delta Ag(v)(d\alpha D - M)]},$$

$$\lim_{A \to 0} \lambda_A = \frac{\overbrace{\lim_{A \to 0} A^{1-\alpha}}^{=0} \delta(v-1)(r + \delta + b\overbrace{\lim_{A \to 0} A^\alpha}^{=0} g(v))}{\delta DM \underbrace{\lim_{A \to 0} A^{1-\alpha}}_{=0} -bk\alpha Dg(v) + b\delta g(v)(d\alpha D - M)\underbrace{\lim_{A \to 0} A}_{=0}}$$

$$= 0,$$

which yields $\hat{\lambda}_A = 0$.

Substituting the resulting $\hat{A} = 0$, $\hat{S} = \frac{k}{\delta}$, $\hat{\lambda}_A = 0$, $\hat{\lambda}_S = 0$ into the canonical system, we get $\dot{\lambda}_A = 1 - \hat{v} = 1 - v_{\max} \neq 0$, and $\dot{\lambda}_A = 1 - \hat{v} = 1 - 0 = 1 \neq 0$ respectively, while in steady state there should hold $\dot{\lambda}_A = 0$. Hence, we arrive at a contradiction. Analysis of the numerical solution shows that any positive value of $\hat{\lambda}_A$ is the product of division by an approximation of $\hat{\lambda}_S = 0$.

Concluding, the no-use steady states $\hat{E}^2_{\hat{v}=v_{\max}}$ and $\hat{E}^2_{\hat{v}=0}$ for the population of IDUs in Australia do not exist because the steady state costate value $\hat{\lambda}_A$ cannot be determined.

4.3.3 United States

For the U.S. parameters, we start with the boundary control steady states without harm reduction, i.e. $\hat{v} = 0$. The three pairs of steady state values

$(\hat{A}, \hat{S})$ are already known from section 3.4. They are now amplified by the steady state costate values $\hat{\lambda}_A$, $\hat{\lambda}_S$ and the Lagrange Multipliers:

$$\hat{E}^1_{\hat{v}=0} = \begin{pmatrix} 5.488796 \\ 7.107619 \\ -10.84137 \\ -6.07868 \end{pmatrix} \quad \text{with} \quad \pi_1 = -3.43948,$$

$$\hat{E}^2_{\hat{v}=0} = \begin{pmatrix} 0.886661 \\ 19.742572 \\ 28.43041 \\ 1.96447 \end{pmatrix} \quad \text{with} \quad \pi_1 = -2.72627,$$

$$\hat{E}^3_{\hat{v}=0} = \begin{pmatrix} 0 \\ 22.176860 \\ -4.85201 \\ 0 \end{pmatrix} \quad \text{with} \quad \pi_2 = 0.$$

The Lagrange Multipliers π_1 at $\hat{E}^1_{\hat{v}=0}$ and $\hat{E}^2_{\hat{v}=0}$ are negative. Hence, those steady states are not among the candidates for optimality. The no-use steady state $\hat{E}^3_{\hat{v}=0}$ is investigated in more detail in section 4.3.4.

The three steady states $(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S)$ with upper boundary control $\hat{v} = v_{\max}$ are

$$\hat{E}^1_{\hat{v}=v_{\max}} = \begin{pmatrix} 5.776702 \\ 6.317187 \\ -8.02684 \\ -4.83059 \end{pmatrix} \quad \text{with} \quad \pi_2 = 4.0649,$$

$$\hat{E}^2_{\hat{v}=v_{\max}} = \begin{pmatrix} 0.727089 \\ 20.180670 \\ 21.4433 \\ 1.20511 \end{pmatrix} \quad \text{with} \quad \pi_2 = 2.09133,$$

$$\hat{E}^3_{\hat{v}=v_{\max}} = \begin{pmatrix} 0 \\ 22.176860 \\ -4.00738 \\ 0 \end{pmatrix} \quad \text{with} \quad \pi_2 = 0.$$

The Lagrange Multipliers π_2 are non-negative, thus those steady states are candidates for optimality and will be analyzed further. In particular, the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$ is considered in detail in the following section 4.3.4.

Nevertheless, we are heuristically able to exclude the steady state $\hat{E}^2_{\hat{v}=v_{\max}}$ from being a candidate for optimality. Keeping the economic interpretation

of the costate variable as shadow price of the stock of users in mind again, we expect that λ_A is negative, because users are "a bad" in the system. The absolute value $|\lambda_A|$ describes the system's internal validation of an infinitesimal increment in drug use. Susceptibles S will potentially flow on to drug use. Hence, their shadow price λ_S is expected to be negative as well. Due to the fact that only a certain proportion of them indeed initiate drug use, we expect that they are not validated as bad as drug users are. In more mathematical terms this means $\lambda_A \leq \lambda_S \leq 0$, or in absolute values $|\lambda_S| < |\lambda_A|$. In the intermediate-use steady state $\hat{E}^2_{\hat{v}=v_{\max}}$, we found $\hat{\lambda}_A = 21.4433$ and $\hat{\lambda}_S = 1.20511$. Positive shadow prices do not correspond to the economic interpretation explained above. Anyway, that steady state is dominated by trajectories leading to another steady state, which will be shown in section 4.5 where the optimal paths are finally computed.

4.3.4 Closer Inspection of the No-Use Steady States

For the parameterization for IDU in Australia, no-use steady states of the four-dimensional canonical system were encountered numerically, but did not exist. Although there is nothing suspect about the no-use steady states for the U.S. cocaine epidemic, we decided to investigate it analytically and in more depth. This will finally allow us to show that the Lagrange Multiplier π_2 is positive in the neighborhood of the steady state $\hat{E}^3_{\hat{v}=v_{\max}}$.

In the U.S. parameter setting, we have $\alpha > 1$. Analogously to section 4.3.2, we modify the costate equation for λ_S such that in

$$\lambda_S = \frac{Ab\delta A^\alpha (v-1)g(v)}{A[\delta DM - bk\alpha Dg(v)A^{\alpha-1} - b\delta A^\alpha g(v)(d\alpha D - M)]}.$$

the factor A in the numerator and the denominator cancels out. Taking the limit for $A \to 0$, we finally get

$$\lim_{A\to 0} \lambda_S \;=\; \frac{b\delta(v-1)g(v)\overbrace{\lim_{A\to 0} A^\alpha}^{=0}}{\delta DM - bk\alpha Dg(v)\underbrace{\lim_{A\to 0} A^{\alpha-1}}_{=0} - b\delta g(v)(d\alpha D - M)\underbrace{\lim_{A\to 0} A^\alpha}_{=0}} \tag{4.23}$$

$$=\; 0.$$

The equation for λ_A is re-arranged to

$$\lambda_A \;=\; \frac{A\delta(v-1)(r+\delta+bA^\alpha g(v))}{A[\delta DM - bk\alpha DA^{\alpha-1}g(v) + b\delta g(v)(d\alpha D - M)A^\alpha]},$$

where the factor A drops out after cancellation. The limit for $A \to 0$ is

$$
\lim_{A \to 0} \lambda_A = \frac{\delta(v-1)(r + \delta + b \overbrace{\lim_{A \to 0} A^\alpha}^{=0} g(v))}{\delta DM - bk\alpha Dg(v) \underbrace{\lim_{A \to 0} A^{\alpha-1}}_{=0} + b\delta g(v)(d\alpha D - M) \underbrace{\lim_{A \to 0} A^\alpha}_{=0} g(v)}
$$

$$
= \frac{\delta(v-1)(r+\delta)}{\delta(r+\delta)(r+\mu)}
$$

$$
= \frac{v-1}{r+\mu}.
$$

Hence, we conclude that $\hat\lambda_S = 0$ and $\hat\lambda_A = \frac{v-1}{r+\mu}$. Please note that the terms including the constant d converge to zero independently of that constant.

Looking at equation (4.23), in the denominator we find the positive constant $\delta DM = \delta(r+\delta)(r+\mu)$, whereas the other terms converge to zero for $A \to 0$. In the numerator we find positive factors, with the exception of the factor $v - 1$. Concluding from this, the costate λ_S converges to zero from the left when $A \searrow 0, S \to \frac{k}{\delta}$, i.e. $\lim_{A \searrow 0, S \to \frac{k}{\delta}} \lambda_S = 0^{(-)}$. Consequently, in the neighborhood of the steady state the costate is negative, as we expect for an optimal solution.

Substituting $\hat A = 0$, $\hat S = \frac{k}{\delta}$, $\hat\lambda_A = \frac{v-1}{r+\mu}$, $\hat\lambda_S = 0$ into the canonical system (4.17) - (4.20), we find that the dynamics indeed come to rest there.

With the parameters specified for the U.S. cocaine epidemic we get

$$
\hat\lambda_A|_{\hat A = 0, \hat v = 0} = \frac{v-1}{r+\mu} = \frac{-1}{0.04 + 0.1661} = -4.85201,
$$

$$
\hat\lambda_A|_{\hat A = 0, \hat v = v_{\max}} = \frac{v_{\max}-1}{r+\mu} = \frac{0.17408 - 1}{0.04 + 0.1661} = -4.00738.
$$

These results coincide with the numerically determined costates $\hat\lambda_A$ of the no-use steady states. The steady states with zero use exist for the four-dimensional canonical system with the parameterization for the U.S. cocaine epidemic.

What remains to be investigated are the Lagrange Multipliers in the neighborhood of the steady states.

Substituting the expressions obtained for $\hat\lambda_A$ and $\hat\lambda_S$ into the Lagrange Multiplier π_1 from equation (4.21), we get

$$
\pi_1 = -A + b\, A^\alpha\, S\, \frac{1}{r+\mu}\, g'(0).
$$

It is easy to see that in the neighborhood of $\hat{S} = \frac{k}{\delta}$, the state S is positive. The factor $\frac{bg'(0)}{r+\mu}$ is positive, too. The exponent α takes a value $\alpha > 1$ in the U.S. cocaine epidemic's parameterization, there holds $A^\alpha < A$ when A is close to zero. Moreover, there holds $b\frac{k}{\delta}\frac{1}{r+\mu}g'(0) = 0.457061$. With this, we infer that there holds $\lim_{A\searrow 0, S\to\frac{k}{\delta}} \pi_1 = 0^{(-)}$. In the neighborhood of the no-use steady state $\hat{E}^3_{\hat{v}=0}$ the Lagrange Multiplier π_1 is negative and therefore it cannot be a candidate for the optimal long-run solution.

Substituting the expressions obtained for $\hat{\lambda}_A$ and $\hat{\lambda}_S$ into the Lagrange Multiplier π_2 from equation (4.22), we get

$$\pi_2 = A + b\,A^\alpha\,S\,\frac{v_{\max} - 1}{r + \mu}\,g'(v_{\max}). \tag{4.24}$$

Here, we evaluate $b\frac{k}{\delta}\frac{v_{\max}-1}{r+\mu}g'(v_{\max}) = -0.488117$. With this, we deduce $\lim_{A\searrow 0, S\to\frac{k}{\delta}} \pi_2 = 0^{(+)}$. In the neighborhood of the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$ the Lagrange Multiplier π_2 is positive and therefore it is a candidate for the optimal long-run solution.

Summarizing the analysis of boundary control steady states, the only candidates for optimality are the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$ and the high-use steady state $\hat{E}^1_{\hat{v}=v_{\max}}$. Please note that both involve upper boundary control $\hat{v} = v_{\max}$.

4.4 Stability Properties of the Feasible Steady States

The most stable situation that can be achieved in the four-dimensional canonical system is saddle point stability. The stable manifold is then two-dimensional. In terms of the Eigenvalues of the Jacobian Matrix the feature is characterized by two Eigenvalues (out of the four) that are negative or have negative real parts. Saddle point stability in the canonical system corresponds to stability in the optimized dynamical system.

4.4.1 Australia

For Australian IDU, there is a single feasible steady state at $\hat{A} = 0.333455$, $\hat{S} = 0.154617$, $\hat{\lambda}_A = -3.95724$, $\hat{\lambda}_S = -2.55002$ with boundary control $\hat{v} =$

$v_{\max}$. The linearized system around this steady state exhibits the Eigenvalues

$$
\begin{aligned}
\xi_1 &= 0.266251, \\
\xi_2 &= -0.226251, \\
\xi_3 &= 0.169598, \\
\xi_4 &= -0.129598.
\end{aligned}
$$

The Eigenvalues are real, and two of them are negative, the others positive. Hence, the fixed point is a saddle node in state-costate space, and a stable point in the optimized dynamical system.

The equilibrium number of users amounts to 333,455 injecting drug users in Australia. Users cause social costs and harms to themselves and to society in general. In Australia, good harm reduction tactics exist that are focused on harms borne by the users themselves. The model's solution indeed suggests that in steady state, the highest possible proportion $v_{\max} = 0.53$ of harms is eliminated. This means that all harms felt by the users are wiped out. At the steady state level of the state A, injecting drug use has grown so far that the reduction in the integrand of the objective function that can be achieved via full harm reduction $\hat{v} = v_{\max}$ is attractive. To benefit from this positive effects of harm reduction, the decision maker accepts the negative consequence of an increase in initiation. This downside is modeled by the function $g(v)$. For the Australian parameterization we evaluate $g(v_{\max}) = g(0.53) = 1.235$. This means that the fully expunged risks (for the users) associated with injecting lead to an increase in initiation by about 24% in steady state.

The resulting value of social costs in steady state is $\hat{A}(1 - \hat{v}) = 0.156724$. This number is an abstract value because the baseline harm per unit of use is normalized to $\kappa = 1$. The steady state number of initiates to drug use is $I(\hat{A}, \hat{S}, v_{\max}) = 0.03788$, meaning that in equilibrium, 37.880 susceptibles initiate injecting drug use every year. Initiation is the inflow that feeds the pool of users, whereas on the other hand μA gives the number of drug users that quit consumption. In steady state, this is $\mu \hat{A} = 0.1136 \cdot 0.333455 = 0.03788$, which is the same number as persons that flow into the pool. This is not surprising, because in steady state $\dot{A} = 0$ holds. Furthermore, we evaluate $\delta \hat{S} = 0.0952 \cdot 0.154617 = 0.01472$, thus 14.720 persons leave the pool of susceptibles in steady state. Summing up those who mature out and those who initiate use, we get $0.03788 + 0.01472 = 0.0526$, which is equal to the inflow k and what we expect due to the fact that $\dot{S} = 0$ in steady state.

4.4.2 United States

The Eigenvalues of the linearized system around the high-use equilibrium $\hat{E}^1_{\hat{v}=v_{\max}}$ with about 5.78 million users, about 6.32 million susceptibles and costates $\hat{\lambda}_A = -8.02684$, $\hat{\lambda}_S = -4.83059$ are two conjugate complex pairs

$$
\begin{aligned}
\xi_{1,2} &= 0.0996532 \pm 0.126644\,\mathrm{i}, \\
\xi_{3,4} &= -0.0596532 \pm 0.126644\,\mathrm{i},
\end{aligned}
$$

one of which has a negative real part. Hence this fixed point is of saddle point stability. It is a so-called saddle focus. Trajectories approaching this steady state will exhibit transient oscillations.

In equilibrium, harm reduction is pursued to the full extent, i.e. $\hat{v} = v_{\max}$, which is at about 17% for U.S. cocaine. The number of users $\hat{A} = 5.776702$ is the highest one among at the encountered steady states. Optimality of maximum control $\hat{v}_{\max}$ in such a steady state can be corroborated in the following way: The high number of users $\hat{A}$ directly transfers to high social costs in the objective function. By application of full harm reduction $\hat{v} = v_{\max}$ a 17% reduction can be achieved in the integrand of the objective function. Of course, there is again the downside of the increase in initiation. Here, it accounts for $g(\hat{v}_{\max}) = 1.09$, meaning that initiation goes up by 9%. There is an increase in participation, but in the trade-off between achieving lower social costs by 17% and an increase in initiation by 9%, the reduction in costs is the more attractive effect and makes the strategy $\hat{v} = v_{\max}$ optimal.

The integrand of the objective functional evaluates to a per year cost of $\hat{A}(1 - \hat{v}) = 4.771094$. Steady state initiation is $I(\hat{A}, \hat{S}, v_{\max}) = 0.959510$. This means that in the high-use equilibrium, each year 959.510 persons from the pool of susceptibles decide to take cocaine. Quitters from cocaine use account for the same number, $\mu\,\hat{A} = 0.1161 \cdot 5.776702 = 0.959510$, which is due to the fact that in steady state the number of users does not change, i.e. $\dot{A} = 0$. The number of persons who quit the pool of susceptibles is $\delta\hat{S} = 0.0605 \cdot 6.317187 = 0.382190$. Due to the fact that $\dot{S} = 0$ holds in steady state, the sum of those who mature out of the S-state and those who initiate must be equal to the constant k, that gives the constant inflow k to the pool of susceptibles. Indeed, we find that $I(\hat{A}, \hat{S}, v_{\max}) + \delta\hat{S} = 0.959510 + 0.382190 = 1.3417 = k$.

At the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$ with a steady state number of $\hat{S} = 22.176860$ million susceptibles, and costates $\hat{\lambda}_A = -4.00738$, $\hat{\lambda}_S = 0$ the

linearized system evaluates Eigenvalues

$$
\begin{aligned}
\xi_1 &= 0.2061, \\
\xi_2 &= -0.1661, \\
\xi_3 &= 0.1005, \\
\xi_4 &= -0.0605.
\end{aligned}
$$

They are real, two of them are positive and two of them negative. Thus, the no-use equilibrium of the U.S. cocaine epidemic is a saddle node. Here, a steady state with no use, $\hat{A} = 0$, will be approached in the long run. That means, that in the steady state itself, the decision maker will not incur any social cost stemming from cocaine abuse, because there is no cocaine problem. The steady state is a boundary control steady state where full harm reduction is applied, so we can expect that at least in a region around the steady state, where numbers of users are moderate, $v = v_{\max}$ will be applied, too. This optimal choice of full harm reduction may find its reason in the fact that for moderate levels of use, initiation is low enough that even with modest increases (see the function $g(v)$) there is little risk that use will explode. We will deal with detailed interpretations later, when the optimal paths are presented.

In steady state, there is no drug use, consequently there cannot be a positive number for initiation because the initiation function suggests that initiation can only happen when there are current users that recruit new users. Analogously, the outflow $\mu\hat{A}$ in steady state is equal to zero. For the dynamics of the susceptibles, we trivially evaluate the number of persons maturing out of this state as $\delta\hat{S} = \delta\frac{k}{\delta} = 1.3417 = k$.

4.5 Optimal Paths

This section is devoted to the determination of the optimal paths leading to the optimal steady states of the canonical system. The optimal equilibria have the property of saddle point-stability, meaning that there exists a two-dimensional stable manifold in the four-dimensional state-costate space. Only initial conditions $(A(0), S(0), \lambda_A(0), \lambda_S(0))$ located on the stable manifold exhibit trajectories that eventually end up in the steady state. The projection of the optimal solutions onto the (A, S)-plane gives the optimized phase portrait.

4.5.1 Australia

Recall that the steady state is located at $\hat{A} = 0.333455, \hat{S} = 0.154617$, with costates $\hat{\lambda}_A = -3.95724$, $\hat{\lambda}_S = -2.55002$. It is a boundary control steady state with optimal control $\hat{v}^* = v_{\max} = 0.53$.

The upper boundary control constraint $v_{\max} - v^* \geq 0$ from equation (4.2) is active there. Due to the complementary slackness condition (4.13), the Lagrange Multiplier π_2 has to be positive along an optimal solution. If the Lagrange Multiplier π_2 hits zero while computing backward, boundary control is no longer optimal and the system has to be switched to interior control v^* from equation (4.16). Further switches are necessary when control constraints become active, or when the Lagrange Multiplier π_1 hits zero when calculating at the lower bound $v^* = 0$.

Optimal Control and Optimized Phase Portrait

Figure 4.1 depicts optimal control as a function of states A and S. The darkest gray color indicates regions of the (A, S)-plane, where optimal control is given by the maximum harm reduction policy $v^* = v_{\max}$. The white part of the phase portrait gives the region where the pure use reduction regime is preferred, i.e. $v^* = 0$. Gray levels in between characterize different amounts of interior control. In more detail, the region where only some harm reduction interventions are implemented, i.e. $0 < v^* < \frac{1}{3} \cdot v_{\max}$, is depicted in light gray. The next darker gray color shows parts of the (A, S)-plane for which an intermediate harm reduction amount $\frac{1}{3} \cdot v_{\max} < v^* < \frac{2}{3} \cdot v_{\max}$ is optimal. The again darker third sliver shows where almost all possible harm reduction interventions are applied, but not yet to the full extent, meaning that there we find optimal control $\frac{2}{3} \cdot v_{\max} < v^* < v_{\max}$.

The lion's share of the (A, S)-plane is covered by boundary control regions. Next to the S-axis, the model suggests optimal control $v^* = 0$. There is a narrow segment for which interior control values are optimal, whereas next to this sliver we find the large region where the full harm reduction policy is optimal.

Furthermore, Figure 4.1 shows the projection onto the (A, S)-plane of some of the optimal trajectories. The little black arrows indicate the direction of convergence towards the steady state. The examples outline that the trajectories usually involve a lot of switches in optimal control.

Please note that although not presented here, the optimal paths exhibit negative costates.
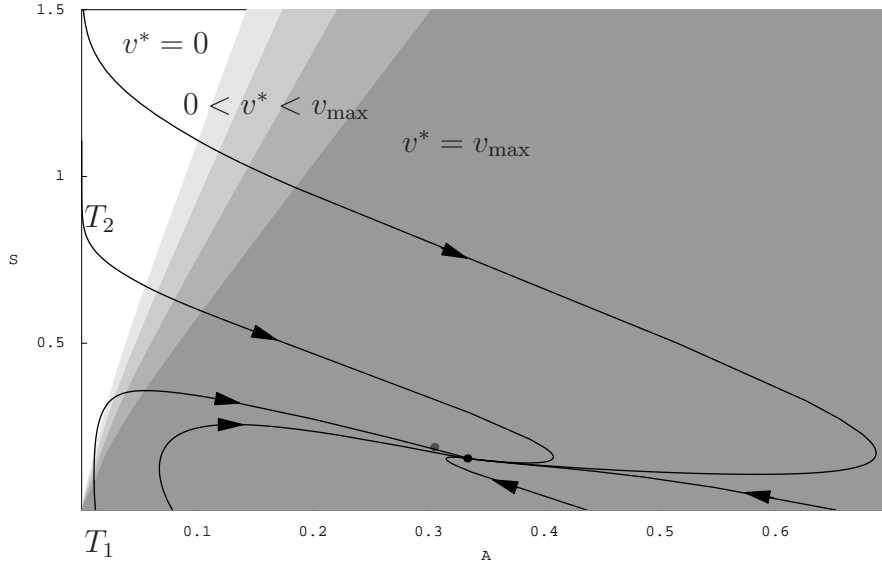
Figure 4.1: Optimal control as a function of A and S for IDU in Australia.

Implications

No matter what initial condition $(A(0), S(0))$ we encounter for the epidemic of IDU in Australia, the optimally controlled system always converges to a steady state with about 0.33 million users and 0.15 million susceptibles. It is shown as a black dot in Figure 4.1.

The gray dot at about 0.3 million users and 0.18 million susceptibles denotes the steady state $\hat{E}^1_{\hat{v}=0}$ of the system with $\hat{v} = 0$. The steady states are located very close to each other; there is a difference of only about 3,000 persons in the number of users and of about 30,000 susceptibles. Still, the analysis in section 4.3.1 shows that the lower boundary control steady state is not even a candidate for optimality. It is a steady state of the investigated system, going without harm reduction (i.e. applying $v = 0$) does not elicit increases in participation, and the number of users is lower than in the other possible fixed point. How can it happen that it is not optimal? The reason lies in the objective of our model. It comprehends the product of the number of users and harms that arise from drug abuse. Our aim is actually to minimize this product. Thus, the model integrates a use reduction approach and a harm reduction approach. At the steady state $\hat{E}^1_{\hat{v}=0}$ full harm is borne to society. At the steady state $\hat{E}^1_{\hat{v}=v_{\max}}$, with an only slightly higher number of users, the full reduction in harms felt by the users is achieved. The gray dot lies in the dark gray region, where optimal control is given by full harm

reduction. Hence, instead of staying at the steady state forever, it is better to apply harm reduction with full force. This example clearly shows that even if the system is close to a steady state of the use reduction policy, it is not necessarily the optimal choice to approach this steady state.

Furthermore, for the large region of the (A, S)-plane around the optimal long-run steady state $\hat{E}^1_{\hat{v}=v_{\max}}$, optimal control is full harm reduction $v^* = v_{\max}$. At those stages of the IDU problem, drug use has grown far and/or the pool of susceptibles is large. Due the random mixing and contagion effects modeled in our initiation function, we can expect that progression from the S-state to the A-state will be considerable in the future. Application of full harm reduction has adverse consequences and pushes initiation, but it is beneficial to cut down the social costs arising for a large number of users. In regions where use is still of modest size (white in Figure 4.1) the optimal control model suggests $v^* = 0$. Keeping in mind the results of the static analysis shown in section 3.6 for Australian IDU, that suggestion for optimal control seems counterintuitive. Figure 3.8 clearly shows that the static decision $v \equiv v_{\max}$ triggers a lower value of aggregated social cost than static control $v \equiv 0$, notably for any initial condition shown on the plot. The static analysis presumes that one of the two control options is chosen once and then applied for all time until the corresponding steady state is reached. In this static world of a one-time irrevocable control decision, full harm reduction always performs better, because at the later stages close to or even larger than the equilibrium levels of use, it significantly cuts down costs. Here, the optimal control approach delivers a more flexible policy. The optimal control model allows control to vary over time. This means harm reduction may progress through control's entire feasible domain $0 \leq v \leq v_{\max}$. Essentially, it is chosen in a way that optimizes the objective functional. Please note that in Figure 3.8, the performance of static full harm reduction is best compared to static pure use reduction in the regions that are shaded lighter, whereas the performance is worse for darker shading. Those darker slivers are located close to the S-axis. When we can choose control optimally from the entire spectrum $0 \leq v \leq v_{\max}$, at initial conditions shaded white in Figure 4.1 and close to the S-axis, optimal control is $v^* = 0$. This is because any amount $v > 0$ of harm reduction would adversely affect trajectories of use and trigger increases in aggregated harms/costs.

The gray regions of interior harm reduction v_i^* also stem from this great flexibility of control choice. Between the extremes of boundary control $v^* = 0$ and $v^* = v_{\max}$, harm reduction interventions have to be planned and implemented step by step. Harm reduction increases at stages of the epidemic, where the effect of reductions in harm are so beneficial that the increased

initiation is an adverse consequence, but still of acceptable size. Nevertheless, we also encounter trajectories along which optimal harm reduction v^* decreases at certain times. This means that the optimized world also allows reversing control when necessary. This happens when increases in use triggered by too high control v would lead to counterproductive increases in use.

Control is dynamically adjusted in order to apply it optimally with respect to the tension of reduced social $A(1-v)$ and the increase in initiation given by the function $g(v)$. Figure 4.1 shows that the region where interior control values $0 < v_i < v_{\max}$ are optimal is rather narrow in state space. The analyses of the phase portrait and of optimal control at different stages of the epidemic do not shed light on the important issue of the actual time structure of the optimal solution. The regions of the (A, S)-plane with maximum control and without harm reduction are divided by a narrow triangular domain of transitory control, hence in particular we are interested in how fast those regions are passed while the epidemic converges to its optimal fixed point. If initial optimal control is the pure use reduction policy $v^* = 0$, but later on a full harm reduction policy $v^* = v_{\max}$ is optimal, does the implementation and gradual incline up to the maximum extent happen within a decade or so? Or is harm reduction an intervention that can and should be applied quickly, probably within a few months only? If amounts of harm reduction decrease, how far and how fast shall harm reduction's implementation be reversed? Such questions and a whole lot of similar issues are of interest.

Furthermore, we assume a discount rate of $r = 0.04$ per year (see section 2.6), which stands for a rather farsighted decision maker. In concrete numbers, social costs that will occur in $t = 5$ years are factored into the current decision with a factor $e^{-rt} = e^{-0.04 \cdot 5} = 0.8187$. Costs one is confronted with in $t = 20$ years have still a factor of $e^{-rt} = e^{-0.04 \cdot 20} = 0.4493$. Although future costs are discounted, costs occurring in the rather far future are still factored into the current decision to a high degree. From this point of view, the detailed time structure of the optimal solutions is interesting, too.

Time Paths

To answer the above stated questions for IDU in Australia, the time paths for the trajectories labeled T_1 and T_2 in Figure 4.1 are depicted.

Trajectory T_1 emanates from $(A(0), S(0)) = (0.011935, 0)$. Optimal control $v^*(t)$ is shown in panel c) of Figure 4.2. Optimal control starts at $v^* = v_{\max}$, which is shown as a black solid line. Subsequently, declining

control represented by the dashed gray segment is optimal. Afterwards, the optimal choice is $v^* = 0$, which is shown in gray. But after some time, increasing control is optimal again, which is depicted as dashed black segment. Finally, optimal control stays at the upper bound $v^* = v_{\max}$ until the steady state is reached. It is again shown in black. The transitions between boundary control regimes are indeed quite short in terms of time. Along the gray dashed part of the transient, optimal control gradually decreases from $v^* = v_{\max}$ to $v^* = 0$ within $t = 2.4$ years. The black dashed transient shows that v^* increases from zero to full harm reduction in $t = 4.81$ years.

In panels a), b) and d), the segments of the resulting time paths for the states and costates are colored corresponding to the optimal control values along the trajectories. Panel a) shows that the number of users is slightly decreasing first, but then it starts to grow. The increase in use is moderate in the first years, but as soon as a critical mass is reached, use grows heavily within a decade. Finally, the steady state value will be approached without overshooting. Panel b) depicts the progression of the susceptibles S over time t. The pool of susceptibles grows at the early stages of the epidemic, overshoots its steady state value, reaches a peak after some $t = 15$ years, and then declines toward the steady state. Panel d) shows how the shadow prices of the users and susceptibles develop over time. Both costates λ_A and λ_S are negative, meaning that an increment in either A or S is always "bad". The costate of the number of users, λ_A, is located below the costate of S for the whole planning horizon, thus users are always judged worse than susceptibles, which is what we expected intuitively.

The tub-shaped form of optimal control v^* shown in panel c) of Figure 4.2 is very interesting, and there arises the question where this particular structure of optimal control stems from. The answer of this question is very much linked to initiation. In the current model, initiation is driven by the contact of current users and non-using susceptible individuals, thus we interpret the tub-shaped form of control complementing the above given information on the evolution of states with the according number of initiates over the course of the drug epidemic. Initiation $I(A, S, v)$ along the trajectory T_1 as a function of time is shown in Figure 4.3.

Given initial conditions like for trajectory T_1, i.e. no susceptibles and only a very modest number of users, the initiation term $I(A, S, v) = bA^{\alpha}Sg(v)$ is zero at the beginning. The dynamics of the susceptibles leads to an immediate increase in the number of people in the S-state, which can be seen on panel b) of Figure 4.2. Initiation grows then, but is at modest levels. It is so low, that initially the number of users drops slightly, although full harm reduction is applied, which additionally leads to a reduction of social cost
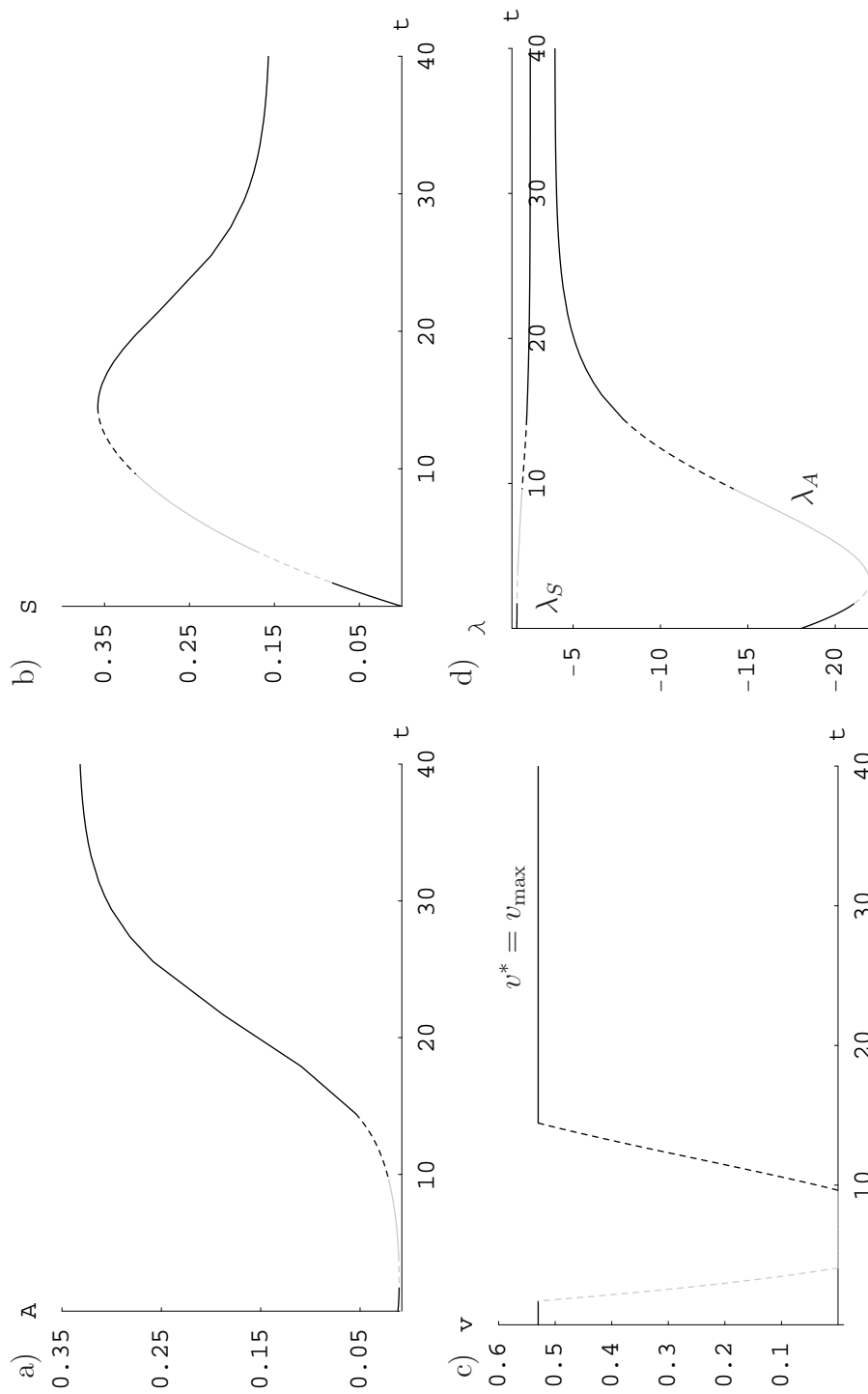
Figure 4.2: Optimal number of users A, susceptibles S, costates λ_S and λ_A, and the optimal amount of harm reduction v^* along trajectory T_1, as a function of time.
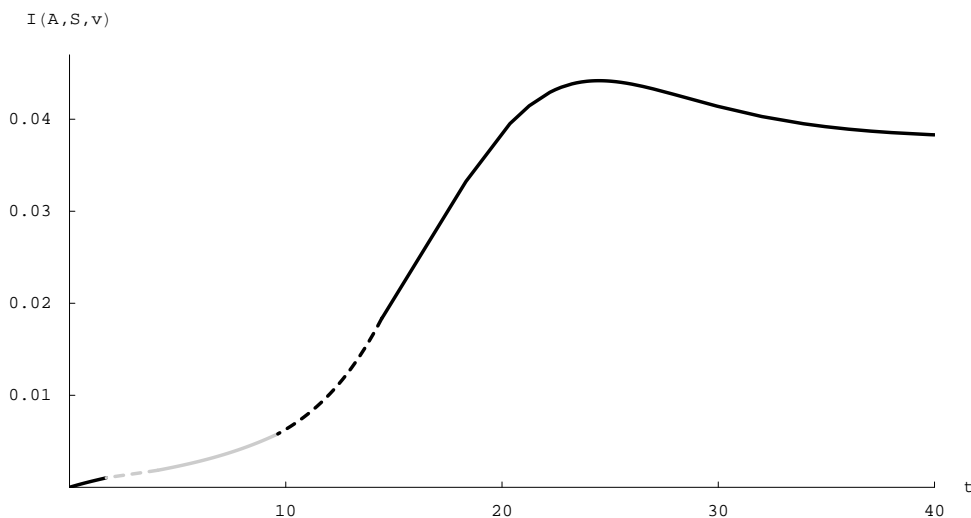
Figure 4.3: Initiation along the optimal trajectory T_1 for Australian IDU

counted in the objective function. Although costs arise only for a modest number of users, the full benefit of harm reduction $v(t)$ can be achieved and the increase in initiation is not much of a disadvantage at those levels of users A and susceptibles S. Nevertheless, the pool of susceptibles grows heavily in the first decade, the resulting combinations of users A and potential new consumers S reach levels at which harm reduction interventions might fuel up initiation in an undesired manner. This justifies the discontinuation of harm reduction measures shown, which finally lead to omission of application of the control variable. The pool of susceptibles is still growing at those stages of the epidemic, and by then use has reached levels at which the reduction of harms felt by the users is attractive again. Harm reduction is gradually implemented again, and although the numbers of new users is considerable at those time spans, control reaches the upper bound of the feasible domain and stays there until the steady state is reached. It is notable that the peak of susceptibles is reached when the final switch from interior control to full harm reduction happens. The pool of susceptibles is exploited afterwards, but use grows heavily. The justification for accepting the increases in initiation and use is the reduction of 53% of harms that can be achieved.

The above initial condition without susceptibles and with a low number of users is not very realistic. More suitable initial conditions for a drug epidemic are located at low numbers of users, but with a pool of susceptibles that is large. Assuming $A(0) = 10,000$ users, the corresponding initial value for the susceptibles located on trajectory T_2 in Figure 4.1 is $S(0) = 0.783634$. Starting optimal control there, the model suggests sticking to the policy
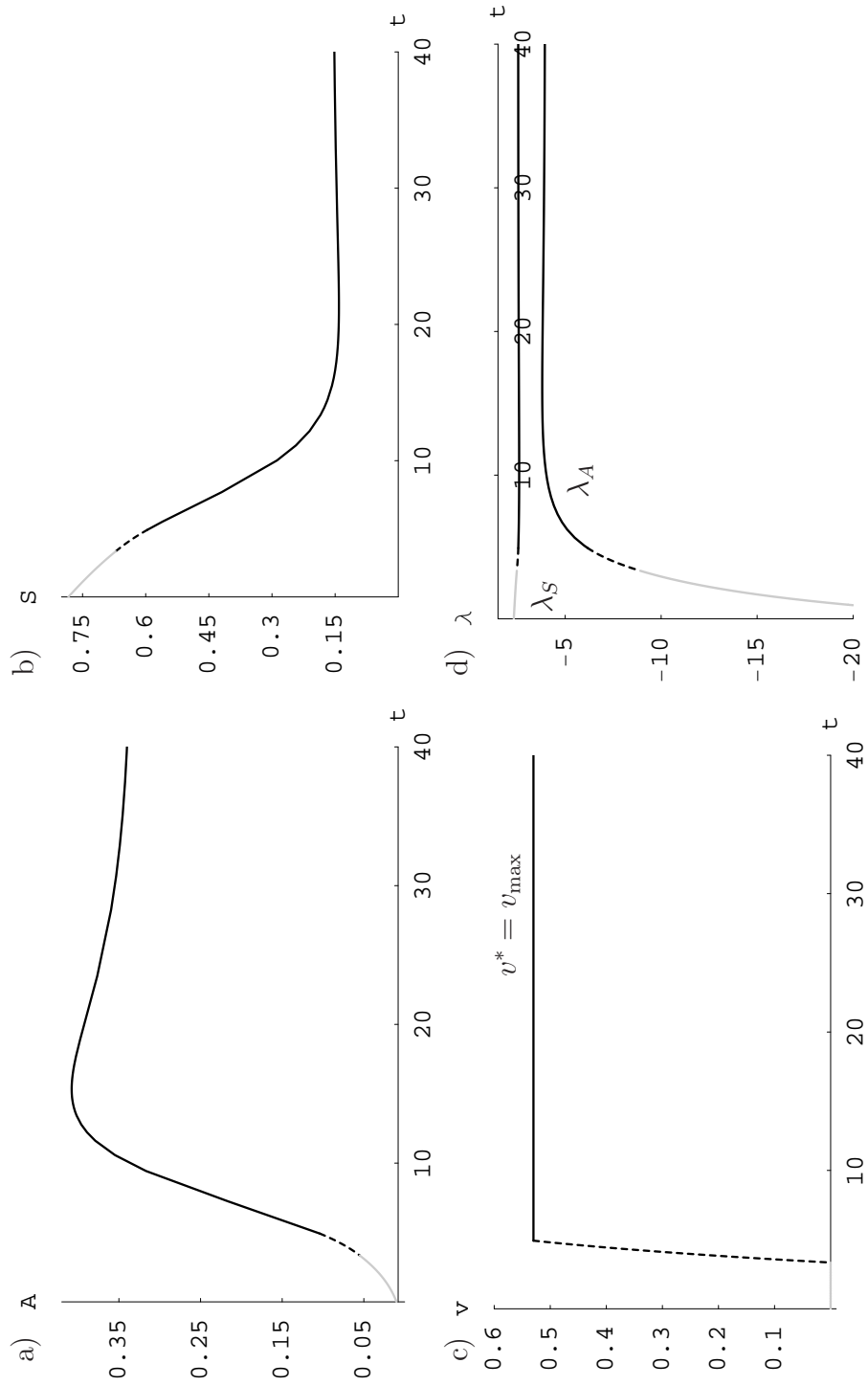
Figure 4.4: Optimal number of users A, susceptibles S, costates λ_S and λ_A, and the optimal amount of harm reduction v^* along trajectory T_2, as a function of time.

without harm reduction for $t = 3.34$ years. This is shown in panel c) by the gray segment. Afterwards, harm reduction is implemented relatively quickly. Namely, within $t = 1.58$ years optimal control progresses from $v^* = 0$ to $v^* = v_{\max}$ (black dashed part). Then, full harm reduction is applied until the steady state is reached, which is shown by the solid black segment in panel c).

Panel a) of Figure 4.4 shows that the number of users overshoots the steady state value $\hat{A}$. The evolution of the pool of susceptibles over time is shown on panel b). It is declining first and then grows slightly towards the steady state value $\hat{S}$. Panel d) plots the costates λ_A and λ_S on a shared panel. As above, the time paths nicely shows that the costates are indeed negative along the entire trajectory, and that the shadow price λ_A of an additional user is always higher in absolute value than the shadow price λ_S of an additional person in the susceptible state.

4.5.2 United States

Dominated steady state $\hat{E}^2_{\hat{v}=v_{\mathbf{max}}}$

In section 4.3.3 we heuristically excluded the intermediate-use steady state $\hat{E}^2_{\hat{v}=v_{\max}}$ from candidacy to optimality. The analysis of the basin of attraction of the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$ allows to give the mathematically proper argumentation by showing that this steady state is dominated. The black dot in Figure 4.5 depicts its projection $(\hat{A}^2_{\hat{v}=v_{\max}}, \hat{S}^2_{\hat{v}=v_{\max}}) = (0.727089, 20.180670)$ onto the (A, S)-plane. The trajectories shown in Figure 4.5 are the optimal paths approaching the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$. The gray segments of the trajectories show where $v^* = 0$. The black dashed parts stand for optimal control v^* that increases gradually from 0 to $v_{\max}$. The black segments indicate that $v^* = v_{\max}$, which characterizes optimal control in a large neighborhood of the no-use steady state.

Located at the initial condition $(A(0), S(0)) = (0.727089, 20.180670)$ there are two options. Under the full harm reduction regime $v = v_{\max}$, those state values form part of a steady state solution. Hence, one could stay there forever applying $\hat{v} = v_{\max}$ for all time. Alternatively, the location in the basin of attraction of the no-use steady state allows us to eventually approach this steady state. In order to converge to the no-use steady state, optimal control at the initial condition is $v^* = 0$.

The value of the objective functional for an optimal control problem with

general objective function F, state x, control v and costate λ is given by

$$
\begin{aligned}
\int_0^\infty e^{-rt} F\big(x(t), v(t)\big) \mathrm{d}t &= \frac{1}{r} H\big(x(0), v(0), \lambda(0)\big) \\
&= \frac{1}{r} H^0\big(x(0), \lambda(0)\big),
\end{aligned}
\tag{4.25}
$$

where H^0 is the Hamiltonian maximized with respect to control. The formula also holds true for models with boundary control (see e.g., Feichtinger & Hartl, 1986). Please note that in particular, the formula does not only hold true for optimal solutions, but for any solution x, v that satisfies the necessary optimality conditions.

To determine the objective functional's value for staying at the steady state $\hat{E}^2_{\hat{v}=v_{\max}}$ forever, we substitute the steady state values into formula (4.25) and get

$$
\int_0^\infty e^{-rt} F\big(\hat{A}(t), \hat{v}(t)\big) \mathrm{d}t = \frac{1}{r} H(\hat{A}, \hat{S}, v_{\max}, \hat{\lambda}_A, \hat{\lambda}_S) =
$$

$$
\frac{1}{0.04} H(0.727, 20.181, 0.174, 21.443, 1.205) = -15.013.
$$

The trajectory $(\tilde{A}(t), \tilde{S}(t), \tilde{\lambda}_A(t), \tilde{\lambda}_S(t))$ emanating from the initial condition $(\tilde{A}(0), \tilde{S}(0)) = (0.727089, 20.180670)$ in the (A, S)-plane, but converging to the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$ assigns initial shadow prices $\tilde{\lambda}_A(0) = -34.648$ to the drug use state and $\tilde{\lambda}_S(0) = -0.996$ to the pool of susceptibles. This path is the bold one among the trajectories shown in panel a) of Figure 4.5. Observing that control at the initial point of the bold trajectory is $v = 0$, formula (4.25) yields

$$
\int_0^\infty F(\tilde{A}(t), \tilde{v}(t)) \mathrm{d}t = \frac{1}{r} H(\tilde{A}(0), \tilde{S}(0), v(0), \tilde{\lambda}_A(0), \tilde{\lambda}_S(0)) =
$$

$$
\frac{1}{0.04} H(0.727, 20.181, 0, -34.648, -0.996) = -9.503.
$$

The objective of the given optimization problem is maximization of the negative objective functional. Staying at the fixed point with an intermediate level of users gives the objective functional's value -15.013, whereas the alternative path that eventually converges to the no-use steady state evaluates to $-9.503 > -15.013$. Hence, staying at the intermediate-use fixed point $\hat{E}^2_{\hat{v}=v_{\max}}$ is not optimal. At the corresponding level of users and susceptibles, the optimal solution is to converge to a steady state with zero users. Eradication of the cocaine problem is achieved by means of application of a

policy that can be described as "do pure use reduction first, but additionally implement harm reduction when use has sufficiently declined". The time paths for states, costates and control along the optimal path are shown in panels b) - e) of Figure 4.5.
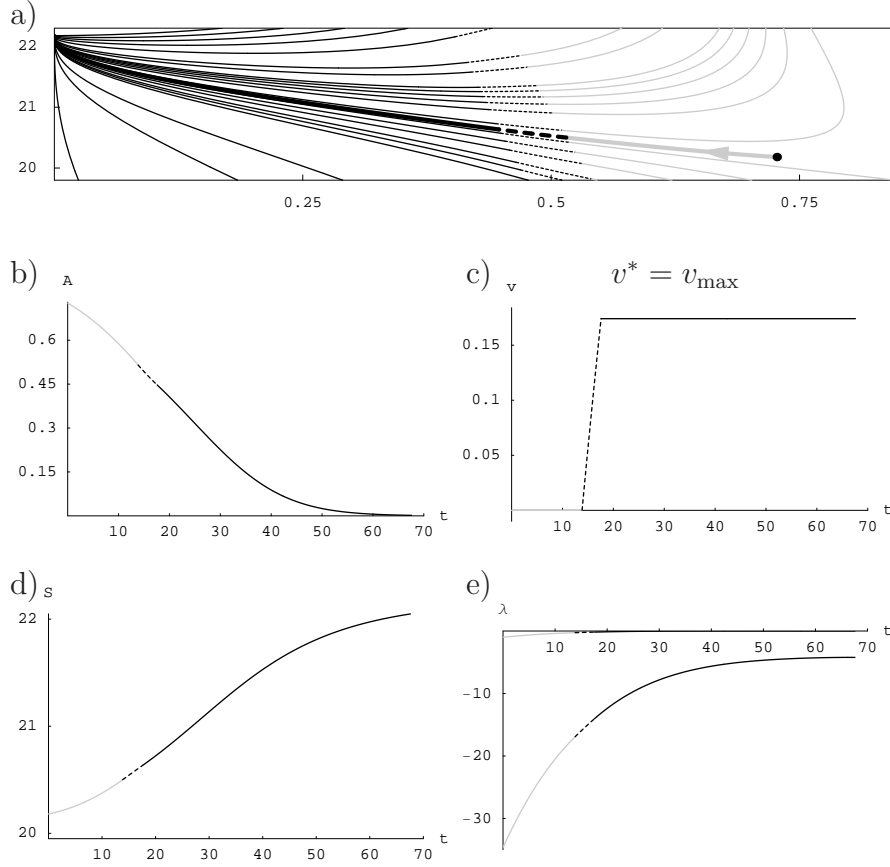

Figure 4.5: The intermediate-use steady state $\hat{E}^2_{\hat{v}=v_{\max}}$ shown as the black dot in panel a) is dominated. Staying forever at $\hat{E}^2_{\hat{v}=v_{\max}}$ is not optimal; the bold trajectory in panel a) leading to the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$ evaluates a better objective functional value. The optimal time path for state A is shown on panel b), for state S on panel d), and the costates λ_A and λ_S are depicted on panel e). Panel c) gives optimal control over time.

Optimal Control

Backward calculation of the optimal long-run steady states $\hat{E}^3_{\hat{v}=v_{\max}}$ and $\hat{E}^1_{\hat{v}=v_{\max}}$ leads to Figure 4.6. It shows optimal control on the (A, S)-plane

in terms of different gray shadings. The meaning of the distinct gray shadings is the same as used in Figure 4.1 for Australia. The darkest regions indicate that optimal control is given by control at the upper bound, i.e. $v^* = v_{\max}$. To the other extreme, the white part of the phase portrait marks the region of the (A, S)-plane where $v^* = 0$ is to be preferred. Regions shaded in gray levels between the extremes denote interior control. In detail, the region shaded in the lightest gray color is where $0 < v^* < \frac{1}{3} \cdot v_{\max}$, the next darker gray level colors regions where the optimal policy is between $\frac{1}{3} \cdot v_{\max} < v^* < \frac{2}{3} \cdot v_{\max}$. Finally, the remaining gray level marks harm reduction interventions $\frac{2}{3} \cdot v_{\max} < v^* < v_{\max}$.

Some of the trajectories of the optimized system are presented to outline the structure of the optimal phase portrait. The basins of attraction are separated from each other in an asymptotic way. This is because the saddle point structure of the dynamics of states A and S under static control $v \equiv 0$ is inherited to the optimized system. The intermediate-use steady state with boundary control $\hat{v} = 0$ is not feasible in the optimal control model. It exhibits a negative Lagrange Multiplier $\pi_1 < 0$ and can therefore not be a candidate for the optimal solution. Nevertheless, in the region around this steady state the optimal policy is given by $v^* = 0$. Consequently, the optimized state dynamics there follow the same rules as in the static system with $v \equiv 0$. The little black arrows indicate the direction of convergence to either one or the other optimal equilibrium.

Looking at the trajectories T_1 and T_2 in Figure 4.6, we see that the optimal policy may again be a sophisticated blend of boundary control and interior levels of harm reduction interventions. When use is initially high and the pool of susceptibles is of rather modest size, one optimally begins with a full harm reduction strategy done for some time, followed by a gradual cutback of control v. Then, a policy of pure use reduction is optimal, but later on harm reduction is re-implemented, exhibiting a monotonic increase until the upper bound is reached. Finally, full harm reduction remains the optimal policy until the steady state is reached.

Indifference Curve in the Basin of Attraction of the High-Use Steady State

Such advanced suggestions for optimal policy are interesting, but are not the fanciest feature our model reveals in terms of policy options. In a rather tiny region between $A = 3$ and $A = 3.5$ million users and between $S = 7$ and $S = 9.5$ million susceptibles, some trajectories overlapped. In such cases it has to be investigated which of the trajectories evaluate the best value of
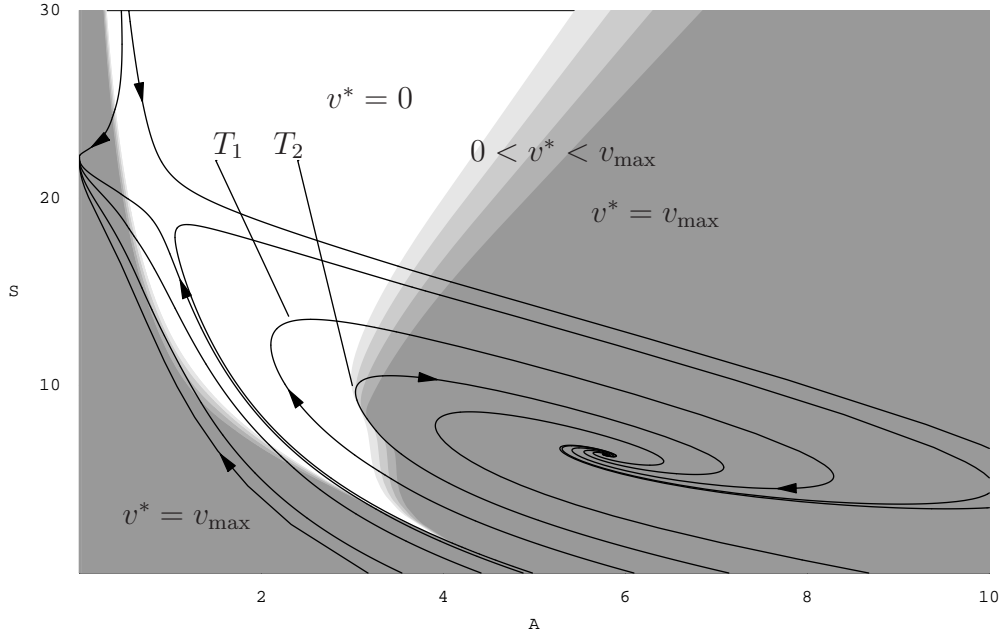
Figure 4.6: Optimal control in the phase portrait for the U.S. base case parameter set

the objective functional and so provides the optimal solution. In the current case, an indifference curve is encountered.

Figure 4.7 presents the indifference curve in the basin of attraction of the high-use steady state $\hat{E}^1_{\hat{v}=v_{\max}}$ of the U.S. cocaine epidemic as a red curve in the (A, S)-plane. The blue curve that emanates in the center and goes towards the upper right corner of the Figure gives the set of points in the (A, S)-plane where in reversed time, calculating backward from the steady state with $v^* = v_{\max}$, the Lagrange Multiplier π_2 hits zero and hence full harm reduction is no longer optimal. The green curve closely passing the point $(A, S) = (3, 11)$ depicts where boundary control $v^* = 0$ becomes optimal along the backward solution. The second blue curve shows where the Lagrange Multiplier π_1 hits zero. The green curve in the middle of the plot depicts where intermediate control relapses to the upper bound. Completing our description of the loci of switches, the green curve in the lower right corner of Figure 4.7 shows where increasing control achieves the value $v^* = v_{\max}$ in the backward continuation.

With respect to optimal control Figure 4.7 is divided into three regions. In the right segment of the Figure, optimal control uses full harm reduction, i.e. $v^* = v_{\max}$. At the left side, the other extreme in harm reduction in-
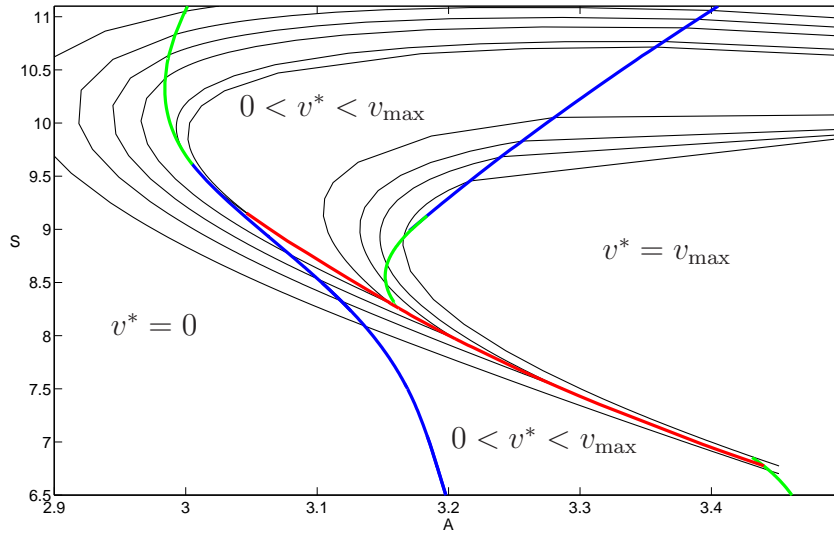
Figure 4.7: Indifference curve in the U.S. base case model.

terventions, namely to do no harm reduction at all, is optimal: $v^* = 0$. In between, we encounter a region of optimal interior amounts $0 < v_i^* < v_{max}$ of harm reduction. Harm reduction grows from the left to the right on the plot, but following the optimal trajectories, we also encounter declining control.

For most of the initial conditions in the basin of attraction of the high-use steady state, there is a unique path in terms of numbers of users and susceptibles that leads to the high-use steady state $\hat{E}_{\hat{v}=v_{max}}^1$. It is characterized by substituting the unique optimal control value into the canonical system. For some special initial conditions, there exist two different initial control values that induce distinct future control and consequently distinct paths on the (A, S)-plane, but both approach the same steady state and evaluate identical values of the objective functional. Thus, one cannot say which of the strategies is better; they are equally optimal and the decision maker is indifferent between them. Such special points are located on the red curve, and imply the label "indifference curve".

The black curves in Figure 4.7 depict some of the optimal trajectories. Most initial conditions were picked exactly on the red indifference curve. The right part of the indifference curve separates regions with interior control, the left part gives points of indifference where one option to initiate control is $v^*(0) = v_{max}$, and the alternative is to start control at some interior level $v^*(0) = v_i$. The future control choices and the resulting development of

states are quite different. Emanating from the red curve with decreasing control, one first aims for reductions in drug use. This choice means that at the beginning of the planning horizon, the system progresses through the region right below the red curve. Use declines there, and to keep it further at low levels, harm reduction declines, until one even switches to a pure use reduction regime. Under the classic control policy, the pool of susceptibles grows bigger and bigger. Via the random mixing effect in the term $b\,A^\alpha\,S$, initiation becomes high enough to trigger growth of the cocaine use state A. When the cocaine epidemic progresses to higher levels of use, harm reduction becomes attractive again. Optimal control v^* increases steadily and finally the optimal policy even switches to full harm reduction.

There are two possible scenarios for the alternative, depending of where on the red curve one starts. In both cases, the states progress through the region directly above the red curve at the beginning of the planning horizon. The distinction is that one either applies full harm reduction first, or that initially an interior amount of harm reduction is done. If one applies full harm reduction first, there will soon be a short intermezzo where control declines to some v_i^*. From there it increases again to full harm reduction, which is then optimal until the steady state is reached. Beginning with interior control $v^*(0) = v_i$, harm reduction declines first and reaches some minimum along the trajectory, but from there it ramps back up to full harm reduction and then the epidemic approaches the optimal steady state under full harm reduction.

The examples for possible paths emanating from the indifference curve show that levels of use are higher at the beginning when control interventions start at a higher level. What is not visible in Figure 4.7, is that the paths that lead initially to cutbacks in use, later on considerably overshoot the alternative. Suspending harm reduction at the beginning leads then to lower levels of use. This is because the increasing effect of $g(v)$ is omitted. Nevertheless, later on, numbers of users are higher. Differently, when harm reduction is applied for the critical initial conditions, use decreases first too, but the cutback is slowed down due to the risk compensation effect that increases initiation. Considering only the early times of the planning horizon, this seems a disadvantage, but at the later stages the strategy is rewarded by lower numbers of users. The tension between distinct control policies and the therewith induced numbers of users (decreases in use first, later an overshoot of the steady state value versus more moderate decreases in use first, but later a less pronounced overshoot of the steady state value of users $\hat{A}$) creates identical values in aggregated harm.

Time paths

To shed more light on this tension and the different developments over time that create the same amount of aggregated harm, Figure 4.8 depicts the two alternative developments for the initial condition $A(0) = 3.2611$, $S(0) = 7.7296$. Starting with full harm reduction $v(0) = v_{\max}$, the initial costates $\lambda_A(0) = -16.54077122$, $\lambda_S(0) = -5.778665364$ are assigned. The alternative involves initial control $v_i^*(0) = 0.0892$, and the costates at $t = 0$ are $\lambda_A(0) = -20.2462$, $\lambda_S(0) = -6.4042$.

Panel c) in Figure 4.8 shows the two distinct control options. The first one is to apply almost always full harm reduction with a rather moderate cutback of harm reduction in the middle of the first decade. The second one is a mix of use reduction and harm reduction, which also includes a time span where no harm reduction at all is done. Let us call the strategy mentioned first the "mostly harm reduction strategy" and the other one the "mixed strategy".

Please note, that the time span on panels a) and b) goes up to $t = 80$, whereas in panel c) time only ranges to $t = 20$. The possible control choices are distinct in the first $t = 15$ years only. For time $t > 15$ years, either option suggests full harm reduction. Hence, the short time span is sufficient to present the control alternatives. In terms of the evolution of users and susceptibles, the different control choices trigger lasting effects. To make the differences in stocks over time visible, the time scale of $t = 80$ is more appropriate. The colors in panels a) and b) associate the control information in panel c) with the development of states. The transient in black stems from the mostly harm reduction strategy, where the solid line indicates $v^* = v_{\max}$ and the dashed segment stands for the short intermezzo of interior harm reduction values being optimal. The mixed strategy induces the transients colored in gray. The dashed segments show the parts that stem from interior control, the solid line indicates that $v^* = 0$, and the dotted part of the time paths of the mixed strategy shows where $v^* = v_{\max}$ within the mixed policy option.

Panel a) shows nicely that the number of users is initially higher under the mostly harm reduction strategy. Use peaks earlier (black) than under the mixed strategy (gray). Moreover, the peak is higher when the mixed strategy is applied. Within the mostly harm reduction strategy, the slight cutback in control (dashed black line) happens when use would grow too strong due to the risk compensation effect $g(v)$ that drives initiation. Furthermore, the black solid and gray dotted curves show how the number of users evolves under $v^* = v_{\max}$. Along those parts, we also encounter time segments for

which the number of users is declining. Along the entire first black segment, $v^* = v_{\max}$ is chosen, and use ebbs back. Those examples show that even if we model that harm reduction interventions increase initiation, the full harm reduction strategy does not hamper declines in use.

Panel b) of Figure 4.8 shows the evolution of the pool of susceptibles. Again, the mixed strategy induces an overshoot in the state. This is what we expect, because the trajectory stemming from the mixed strategy progresses through the (A, S)-space above the trajectory induced by the strategy with mostly harm reduction.

In the basin of attraction of the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$ the regions where transitions between boundary control values occur are quite narrow in (A, S)-space. In the basin of attraction of the high-use steady state $\hat{E}^1_{\hat{v}=v_{\max}}$ the transitory regions are broader, but still rather small compared to the large regions where boundary control is optimal. For Australian IDU, we found similarly narrow spaces of optimal interior control, and the time paths in Figures 4.2 and 4.4 revealed that the transitions are also quick in terms of time that is needed to switch from one bound to the other. We infer from Figures 4.5 and 4.8 that it is similar when optimally applying harm reduction for the U.S. cocaine epidemic. Control measures are introduced quickly when it is advantageous to exploit the social cost cut in the objective function. When the epidemic's course is affected adversely due the increased initiation, the harm reduction policy is optimally reversed quickly, too. Essentially, in such situations, it is not always necessary to completely shy away from harm reduction. Sometimes it is enough to reduce the amount of harm reduction interventions to some intermediate value, but soon increase it to the full extent again (see the black control strategy in panel c) of Figure 4.8).

The different regions where pure use reduction and full harm reduction are the optimal control policy suggested by the current model are not congruent to the gray and white regions we identified in section 3.6, when conducting the static analysis of $v \equiv 0$ versus $v \equiv v_{\max}$. As in the case for Australia, this is due the more flexible approach of the optimal control model. In the static world of a one-time decision, we simply compared aggregated harm along the trajectories resulting from either static option.

The most striking effect concerns regions that are shaded white and located close to the gray region in Figure 3.7. For any initial condition located there, static full harm reduction on the entire way to a steady state (either the high-use of the no-use steady state) triggers less social cost than to never apply harm reduction. In the optimal control result, a large part of this domain is now optimally controlled with $v^* = 0$. In the static world the de-
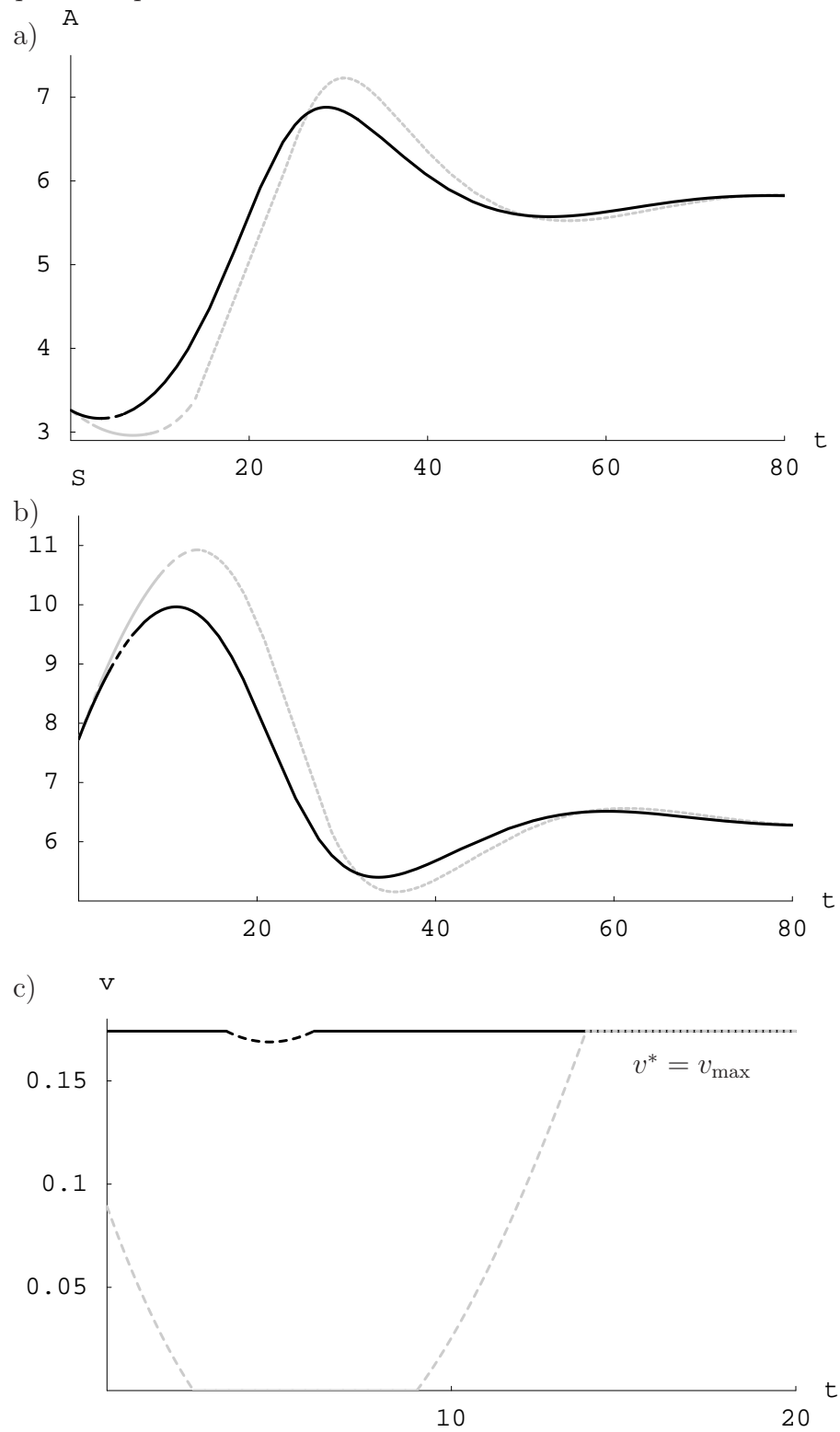
Figure 4.8: Time paths for A, S and v for an initial condition on the indifference curve occurring for the U.S. base case parameterization

cision taken at $t = 0$ cannot be altered later. The optimal control approach allows for flexible adjustments in control, all aimed at minimization of total social cost. Picking an initial condition in the critical region that is shaded white on both Figure 3.7 and 4.6, the optimal solution in the world of flexible adjustment of control consists in $v^* = 0$ at some initial time spans, but later on optimal control progresses to interior and full harm reduction. At the intermediate stages shaded white on the gray level portrait in Figure 4.6 any amount $v > 0$ of harm reduction negatively affects the course of the cocaine epidemic. Consequently, if dynamic choice of control is possible, harm reduction is omitted there. Nevertheless, the model suggests that harm reduction may be applied for higher numbers of users. There, even the modest share of v_{max} at about 17% leads to significant benefits because it fully cuts down the harms felt by the users, such that to society overall, there is a remainder of only 83% of baseline social cost. Furthermore, harm reduction can be applied safely at the early stages of the epidemic. Although $v^* > 0$ leads to increases in initiation, the contagious effect of the U.S. cocaine epidemic is low at low levels of use, hence there is no risk that use might explode when harm reduction is applied.

4.6 Conclusions from the Base Case Model

In the preceding sections, an optimal control model was analyzed in order to determine within the context of that model how the mechanism of harm reduction should be optimally applied in the case of the U.S. cocaine epidemic and for the prevalence of IDU in Australia. The main conclusion of the model is that if drug consumption can vary over time, drug control interventions are optimally varied, too. A very smart feature of the model presented here is that it consists in only one control variable v. This control has a lower bound $v = 0$, which is essential in the optimal solutions and implicitly stands for application of the traditional drug control interventions of law enforcement, prevention and treatment. The model tries to provide an answer or some advice to the participants in the discussion whether use reduction or harm reduction was the right policy.

In the Australian parameterization of the model, harm reduction tactics are optimal for almost all possible initial conditions. Only if numbers of users are very low, it is better to omit harm reduction. Still, when the number of users increases only slightly, harm reduction shall be applied. Very soon, harm reduction is optimally done with full force.

For the U.S. cocaine epidemic parameterization, the policy recommen-

dation is less straightforward. Multiple steady states occur. This creates a domain of interior levels of use where harm reduction is omitted in the optimal control solution. For high levels of use, or when use is rather low, the optimal strategy consists in full application of harm reduction.

For both the U.S. and Australian parameterization, the most probable initial conditions exhibit optimal trajectories that involve switches in optimal control. The results presented in this thesis show once more that for a problem that varies over time, the optimal control interventions can be quite distinct over time, too. Furthermore, we can deduce that control interventions that pursue different aims (use reduction versus harm reduction) may both have merits over the course of a drug epidemic, but not necessarily always at the same time. An indifference curve occurs for the high-use steady state of the U.S. cocaine epidemic parameterization. For initial conditions located there, there exist two possibly quite different options that are equally optimal.

To the extent that one can generalize the results from these stylized models, the answer to the opponents in the discussion may be that neither the one nor the other strategy is unequivocally the best. Perhaps both are good, important, and advantageous, but possibly not at all times. Use reduction tactics are traditional and most have been evaluated to be effective. Nevertheless, at certain stages of the dynamic evolution of a drug problem in a society, it maybe fruitful with respect to minimizing social costs and harms borne to society, to apply harm reduction mechanisms. Proponents of use reduction, and of harm reduction, need not demonize the ideas and downplay the arguments of the other. Rather, they might accept that the ideas of the other party may be fruitful and beneficial in at least certain circumstances, and try to reconcile their perceptions in order to build models that will yield better drug control strategies for the future.

Chapter 5

Sensitivity and Bifurcation Analysis

In optimal control problems, the investigated dynamic systems usually have a lot of parameters. Exact values are mostly not available. Sensitivity and bifurcation analysis are the appropriate tools to investigate the effect of changes in parameters. Sensitivity analysis seeks to answer the question whether and how results change when parameters are slightly modified. Bifurcation theory tries to explain disruptions like destruction or creation of steady states in the system when the parameters vary.

5.1   Sensitivity Analysis

This section assesses the sensitivity of the optimal steady state solutions under parameter changes. Departing from the optimal steady states for the base case parameterizations for Australian IDU and U.S. cocaine use, a one-dimensional 1%-increase sensitivity analysis is conducted. This means that one of the base case parameter values is multiplied with 1.01, while the other parameters are kept at their original value, and then the steady state values are re-computed. The effects are expressed as percentages. The subsequent Tables 5.1 and 5.2 present the effect of the 1%-increase of each parameter value on the steady state values of states, $\hat{A}$ and $\hat{S}$, of costates, $\hat{\lambda}_A$ and $\hat{\lambda}_S$, and the Lagrange Multiplier of the active constraint.

The optimal steady states in the base case are boundary control steady states with control at the upper bound, i.e. $\hat{v}^* = v_{\max}$. Thus, the Tables 5.1 and 5.2 contain the Lagrange Multiplier π_2. The percentage change in the

Lagrange Multiplier π_2 is investigated to find out which parameter changes are most likely to induce that the boundary control steady state is no longer a candidate for the optimal solution.

Interpreting the changes in costate values, we have to keep in mind that the costates are negative. The Tables 5.1 and 5.2 list the change in the absolute value of the costates. Consequently, a positive percentage in the columns for the costates means that the absolute value of the costate increases, which means that the value of the costate decreases. In the interpretation of the costates as shadow prices of an infinitesimal increment in the state, a positive value means that an increase in the state is internally valued worse under the new parameters than in the original system. When a shadow price is shifted to the right on the negative axis, meaning that its absolute value decreases, the internal validation is less negative. This manifests in a negative value in the Tables.

5.1.1 Australia

Table 5.1 gives the results of the sensitivity analysis for the unique optimal steady state $\hat{E}^1_{\hat{v}=v_{\max}} = (0.333455, 0.154617, -3.95724, -2.55002)$ of the base case parameterization for Australian IDU.

In each line, the symbol of the single parameter multiplied with 1.01 is listed, followed by the induced changes in steady state values given as percentages. E.g., if the inflow k to the pool of susceptibles goes up by 1%, the steady state value $\hat{A}$ is increased by 1.318% relative to the base case. The number of susceptibles increases by 0.181%, $|\hat{\lambda}_A|$ decreases by 0.213%, and the absolute value of costate $\hat{\lambda}_S$ increases by 0.187%.

The strongest impacts (more than proportional) on the equilibrium values are encountered when parameters k, μ and $v_{\max}$ are varied. When k increases by one percent, $\hat{A}$ increases by more than 1.3%, π_2 by more than 1.4%. The 1%-increase of μ reduces the steady state value $\hat{A}$ by more than 1.3%, the Lagrange Multiplier π_2 is decreased by more than 1.4% compared to its base case value. The increase in the upper bound $v_{\max}$ by 1% leads to decreases of about 1.1% in $|\hat{\lambda}_A|$ and $|\hat{\lambda}_S|$. This means that increments in both A and S are judged negatively, but not entirely as bad as in the base case.

The modification of $v_{\max}$ by +1% has strong impacts on the costate values in steady state, but 1%-increases in the other parameters of the function $g(v)$, which are c_m, c_s, ω, and η, yield only modest changes in the steady state values.

	$\hat{A}$	$\hat{S}$	$|\hat{\lambda}_A|$	$|\hat{\lambda}_S|$	π_2
k	1.318	0.181	-0.213	0.187	1.408
b	0.365	-0.940	-0.246	0.217	0.468
α	-0.352	0.905	0.536	0.0868	-0.478
μ	-1.305	0.817	-0.445	-0.846	-1.424
δ	-0.369	-0.051	0.193	-0.171	-0.449
c_m	-0.059	0.152	0.0399	-0.035	0.018
c_s	0.059	-0.153	-0.040	0.035	-0.018
v_{max}	0.059	-0.153	-1.167	-1.093	0.182
w	0.059	-0.153	-0.040	0.035	-0.018
η	0.078	-0.200	-0.053	0.046	0.006

Table 5.1: Effects of a 1%-increase in parameter values on the steady state values of $\hat{E}_{\hat{v}=v_{\max}}$ and the active Lagrange Multiplier π_2 for the Australian base case parameterization.

The consequences of the increases in parameters in terms of increased or decreased equilibrium stocks of states and the corresponding Lagrange Multiplier are mostly as one would intuitively expect thinking of the dynamics $\dot{S}$ and $\dot{A}$ (see equations (4.4) and (4.5)).

Increasing k means that more people stream into the pool of susceptibles every year. Thus, we expect a higher $\hat{S}$, and indeed the corresponding percentage displayed in Table 5.1 is positive. The increased inflow to susceptibility makes the initiation term $bA^\alpha S$ larger. This increased flow from S to A leads to a higher steady state value $\hat{A}$. For the Lagrange Multiplier π_2 an increase is detected. Thus, the higher inflow to the pool of susceptibles assures in a certain sense that the full harm reduction steady state remains a candidate for optimality. A fearful decision maker could expect that the negative consequences of harm reduction modeled in $g(v)$ were more likely to trigger an explosion in use, when the constant inflow k to the pool of susceptibles was higher. Hence, he or she would most possibly shy away from a full harm reduction policy. Table 5.1 shows that at least for the high-use steady state and for a slight increase of 1% in k, the indicator for optimality of this high-use steady state with full harm reduction is shifted to the right along the positive axis.

If the epidemic is more virulent, i.e. when the parameter value of b is larger, $\hat{A}$ increases and $\hat{S}$ decreases, just as one would expect regarding the initiation term $I(A, S, v) = bA^\alpha S g(v)$ which is an inflow to the pool of users and an outflow from the pool of susceptibles.

The increase in the parameter α leads to less users $\hat{A}$ in steady state, whereas the steady state number $\hat{S}$ of susceptibles is higher. This is understandable due to the fact that for power functions $f(A) = A^\alpha$ with $0 < \alpha < 1$, an increase in α pulls the concave function closer to the line $f(A) = A$ on the interval $A \in [0, 1]$. In the current context this means, that the contagious effect of current users on susceptibles is reduced. The new system then reaches a steady state with a lower number of users.

Increasing μ by 1%, the constant outflow μA is directly reduced, which decreases the state A. In turn, initiation decreases such that the steady state value of users is lower, as one would intuitively expect. Further, the equilibrium value $\hat{S}$ is increased.

The increase in δ, which is the fraction of susceptibles that flow out of the state S ("maturing out"), leads to a decrease in the number of susceptibles in steady state. Due to the smaller pool of susceptibles from which new users can be recruited, the equilibrium number of users is decreased as well.

Thinking of the Lagrange Multiplier π_2, fearful argumentations could expect that increases in the infectivity proportionality constant b bear greater danger under a full harm reduction strategy, because the increases attributable to application of harm reduction modeled by $g(v)$ multiply the increased b in the initiation term $I(A, S, v) = b\,A^\alpha S g(v)$. Nevertheless, using $1.01 \cdot b$ the steady state with full harm reduction is still optimal and the value of the Lagrange Multiplier π_2 even increases, thus this modification in the parameters does not push the boundary control steady state closer to the borderline where it ceases being a candidate for optimality. Contrary to this, when α is decreased, which pushes A^α closer to a linear form, the Lagrange Multiplier π_2 decreases. The same happens, when the outflow μ of drug use or the outflow δ of the pool of susceptibles is increased. The results indicate that when people quit spans of susceptibility quicker, when the contagion effect is less pronounced, and when desistance is higher, the full harm reduction solution is still optimal for slight changes. Nevertheless, the decreases in π_2 indicates that if the direction of the change stays the same under larger increases in the parameters values, the dynamics are affected in a way that might eventually rule out the full harm reduction steady state from candidacy for optimality. In a system with more susceptibles (increased k) or an increased proportionality constant of contagion, b, full harm reduction tactics in the steady state are the appropriate tool.

As stated above, the 1%-changes of the parameters c_m, c_s, ω and η of the function $g(v)$ have only a low impact on the steady state values, and are not analyzed in depth here.

5.1.2 United States, High-Use Steady State

Table 5.2 summarizes the effects of an increase by 1% in the parameter values derived for the U.S. cocaine epidemic on the high-use steady state $\hat{E}^1_{\hat{v}=v_{\max}} = (5.776702, 6.317187, -8.02684, -4.83059)$.

The most severe changes happen when the parameters k, α or μ are increased. Their 1%-increases induce changes of more then 3% in the Lagrange Multiplier π_2 at the steady state. More than proportionate, but less striking increases or decreases are encountered for the increase of k ($\hat{A}$ and $\hat{\lambda}_2$), b ($\hat{S}$ and $\hat{\lambda}_A$), α ($\hat{A}$, $\hat{S}$ and $\hat{\lambda}_A$), μ ($\hat{A}$, $\hat{S}$ and $\hat{\lambda}_A$) and δ (π_2).

The changes induced by multiplication of the parameters of the function $g(v)$ with 1.01 are very moderate only. Particularly so for the 1%-increase in the upper boundary value for harm reduction, $v_{\max}$, for which there resulted over-proportional effects for the system run for Australian IDU.

	$\hat{A}$	$\hat{S}$	$|\hat{\lambda}_A|$	$|\hat{\lambda}_S|$	π_2
k	1.793	-0.9915	-1.642	-0.559	3.198
b	0.506	-1.270	-1.062	-0.361	1.404
α	1.380	-3.465	-1.929	-0.027	3.431
μ	-1.790	2.027	1.124	-0.014	-3.391
δ	-0.515	0.290	0.857	0.292	-1.232
c_1	-0.042	0.105	0.089	0.030	0.301
c_2	0.042	-0.106	-0.089	-0.030	-0.304
v_{max}	0.042	-0.106	-0.300	-0.241	0.206
w	0.042	-0.106	-0.089	-0.030	-0.304
η	0.046	-0.115	-0.097	-0.033	-0.293

Table 5.2: Effects of a +1%-change in parameter values on the steady state with a high number of users $\hat{E}^1_{\hat{v}=v_{\max}}$ in the U.S. base case parameterization.

5.1.3 United States, No-Use Steady State

The presentation in a Table is omitted, because the no-use steady state $\hat{E}^3_{\hat{v}=v_{\max}}$ located at $(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (0, 22.176860, -4.00738, 0)$ is not subject to many changes. It is obvious, that the steady state value $\hat{A} = 0$ is insensitive to parameter changes. The steady state value of susceptibles $\hat{S} = \frac{k}{\delta}$ only depends on the two parameters k and δ. Looking back on the convergence of the Lagrange Multiplier π_2 (see equation (4.22)) discussed in section 4.3.4,

the expressions involved in equation (4.24) depend continuously on the parameters. Hence, the property $\lim_{A \searrow 0, S \to \frac{k}{\delta}} \pi_2 = 0^{(+)}$ persists. Consequently, the no-use steady state with $\hat{v} = v_{\max}$ remains a candidate for optimality under small variations of the parameter values.

5.2 Bifurcation Analysis

Bifurcation theory detects and explains disruptions in the dynamical behavior under parameter change. Examples for such disruptions are the creation or annihilation of steady states or the exchange of stability properties of equilibria.

Thinking of a certain parameter p of a parameter-depending dynamical system, a parameter value where a bifurcation occurs is called a bifurcation value or critical value of the system and will be denoted by p_c.

A most fundamental type of bifurcation, in which fixed points of a system are created or destroyed, is the so-called saddle node bifurcation. Alternative denominations for the catastrophic fold (or tangent) bifurcation in literature are abundant. A nice denomination for such a creation event is "blue sky bifurcation", which refers to the magical emergence of a pair of fixed points "out of the blue sky". This type of bifurcation is indeed detected in the current model with the U.S. parameterization.

The following section 5.2.1 illustrates the annihilation event of steady states in the two-dimensional system (4.4) and (4.5). In the remainder of this Chapter, the four-dimensional system (4.17)-(4.20) is analyzed with respect to bifurcation points. The bifurcation analysis presented here is limited in the sense that we only focus on two parameters, which are b and μ.

A good introduction to the field of bifurcation theory is given in Grass et al. (2008). For more comprehensive elucidations on the theoretical and numerical details, the reader is referred to good textbooks on the specific subject (e.g., Guckenheimer & Holmes, 1983; Kuznetsov, 1998).

5.2.1 Blue Sky Bifurcation in the Two-Dimensional System

The results from section 3.7 suggest that decreasing the infectivity proportionality constant b for the U.S. cocaine epidemic induces a change in the

vector field of the two-dimensional system (4.4), (4.5). The case of a decreased parameter value of b might indeed play a role, because there are hypotheses that the virulence of the U.S. cocaine epidemic has dropped in its later stages (Tragler et al., 2001; Caulkins et al., 2004; Johnson et al., 1996).

Remember that the boundary control steady states are located at the intersection of the isoclines $\dot{A} = 0$ and $\dot{S} = 0$. The inner branch of the isocline $\dot{A} = 0$ is given by $S = \frac{\mu A}{bA^\alpha g(v)}$. As shown in Figure 3.3 it is a downward sloping line in the U.S. case ($\alpha > 1$). Recall furthermore that in section 3.1 we concluded that any steady state of the system is located along the line $\hat{S} = \frac{k - \mu \hat{A}}{\delta}$, which is the black line in Figure 3.3. Looking at its equation $S = \frac{\mu A}{bA^\alpha g(v)}$ it is clear that the isocline $\dot{A} = 0$ is pushed further away from the origin, when b decreases. The intermediate-use steady state of the system is shifted to the right along the line, whereas the high-use steady state moves to the left along the line. If the critical value b_c is reached, a borderline case is met. The inner branch of the isocline $\dot{A} = 0$ is tangent to the line $\hat{S} = \frac{k - \mu \hat{A}}{\delta}$. The intermediate-use steady state and the high-use steady state collide there. When b decreases further, they do not exist any longer. The property of tangentiality is helpful in order to compute the critical parameter values where blue sky bifurcations occur in the two-dimensional system with boundary control.

In mathematical terms, the tangentiality means that the derivatives with respect to A of the equations $S = \frac{\mu A}{bA^\alpha g(v)}$ and $\hat{S} = \frac{k - \mu \hat{A}}{\delta}$ are equal. Additionally the property of a steady state has to be given. Thus, we arrive at a set of three equations

$$k - \delta S - b\,A^\alpha\,S\,g(v) = 0, \tag{5.1}$$

$$b\,A^\alpha\,S\,g(v) - \mu A = 0, \tag{5.2}$$

$$\frac{\mu(1 - \alpha)}{bg(v)}A^{-\alpha} = -\frac{\mu}{\delta}. \tag{5.3}$$

We substitute the base case parameters k, δ, α, μ and the parameter values for $g(v)$ into the system (5.1)-(5.3), and solve it simultaneously to obtain the critical parameter value b_c and the corresponding steady state $(\hat{A}_c, \hat{S}_c)$.

In the uncontrolled case, this yields the critical steady state $\hat{A}_c^{\hat{v}=0} = 2.901002$, $\hat{S}_c^{\hat{v}=0} = 14.212291$, and the critical parameter value $b_c^{\hat{v}=0} = 0.006434$.

In the system with full harm reduction the blue sky bifurcation occurs at $\hat{A}_c^{\hat{v}=v_{\max}} = 2.901002$, $\hat{S}_c^{\hat{v}=v_{\max}} = 14.212291$, and the critical parameter value is $b_c^{\hat{v}=v_{\max}} = 0.005885$.
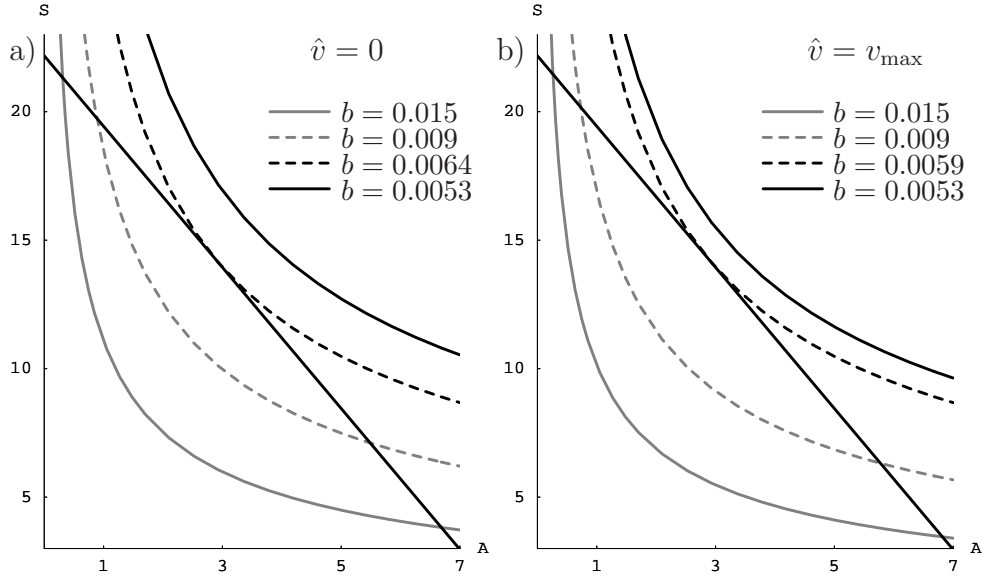
Figure 5.1: Illustration of the annihilation event of intermediate- and high-use steady state when b declines for the U.S. cocaine epidemic

Figure 5.1 sheds light on the annihilation event when b changes. In both panels, the line $\hat{S} = \frac{k - \mu \hat{A}}{\delta}$ is shown in black on the (A, S)-plane, whereas the different curves show the inner branch of the isocline $\dot{A} = 0$ for different values of b.

Panel a) depicts the case when there is no harm reduction, i.e. $\hat{v} = 0$. The gray dashed curve is for the base case value $b = 0.009$, the gray solid curve illustrates the case of the higher value $b = 0.015$, where the steady states are located further away from each other. The black dashed curve depicts the borderline case of $b_c^{\hat{v}=0} = 0.006434$ where the saddle and the node of the two-dimensional system collapse into a single steady state. The black solid curve shows the case $b = 0.0053$, where infectivity is so low, that fixed points with $\hat{A} > 0$ do not exist.

Panel b) shows the same phenomenon for full harm reduction control $\hat{v} = v_{\max}$. The gray dashed curve gives the inner branch of the isocline $\dot{A} = 0$ for the base case value $b = 0.009$, the gray curve illustrates the location of the equilibria in the case $b = 0.015$, the black dashed curve depicts the borderline case $b_c^{\hat{v}=v_{\max}} = 0.005885$, and the black solid curve shows the case $b = 0.0053$, where infectivity is so low, that the two fixed points with intermediate and high levels disappear.

Note that for all those variations of b, the no-use steady states of the two-dimensional system remain unchanged at $\hat{E}_{\hat{v}=0} = \hat{E}_{\hat{v}=v_{\max}} = (\hat{A}, \hat{S}) = (0, \frac{k}{\delta})$.

5.2.2 Bifurcation Analysis in the Four-Dimensional System

Figure 5.1 displays isoclines and steady states on the (A, S)-plane to visualize the collision and annihilation of the intermediate-use saddle and the high-use focus. Bifurcation diagrams provide the same information, but display the parameter that undergoes the variation and the corresponding steady state values. The steady states of the canonical system (4.17)-(4.20) have four components, thus in a two-dimensional plot, only one can be displayed, which is $\hat{A}$ in what follows.

For the simple investigation whether annihilations or creations happen for variation of a parameter value, analysis of the two-dimensional system like conducted in the preceding section is sufficient. Nevertheless, the canonical system (4.17)-(4.20) of the optimally controlled system is richer and more conclusions can be drawn. The computations then also provide the appropriate costate values $\hat{\lambda}_A$ and $\hat{\lambda}_S$. We determine the Lagrange Multipliers for the boundary control steady states. This allows for insight for which parameter values the boundary control steady states are candidates for optimality and for which not.

Additionally, the bifurcation analysis is conducted for the steady states with steady state control $\hat{v}$ from equation (4.16). This investigation provides the information for which parameter values there exist steady state solutions with feasible interior control $0 < \hat{v}_i < v_{\max}$.

Bifurcation Analysis with Respect to the Parameter μ for Australian IDU

In the case of Australian IDU, where the epidemic spreads with an exponent $\alpha < 1$, a single candidate for optimality is encountered. It is a steady state with upper boundary control $\hat{v}^* = v_{\max}$. Any trajectory that emanates in the first quadrant of the (A, S)-plane will eventually reach this steady state. Due to the shape of the isoclines, we expect that under parameter variation, there only occurs a shift of the boundary control steady state along the line $\hat{S} = \frac{k - \mu \hat{A}}{\delta}$, but that additional steady states do not emerge.

Figure 5.2 shows the bifurcation plot when μ varies. At the base case parameter value $\mu_{BC} = 0.1136$ there is vertical dashed line colored black.
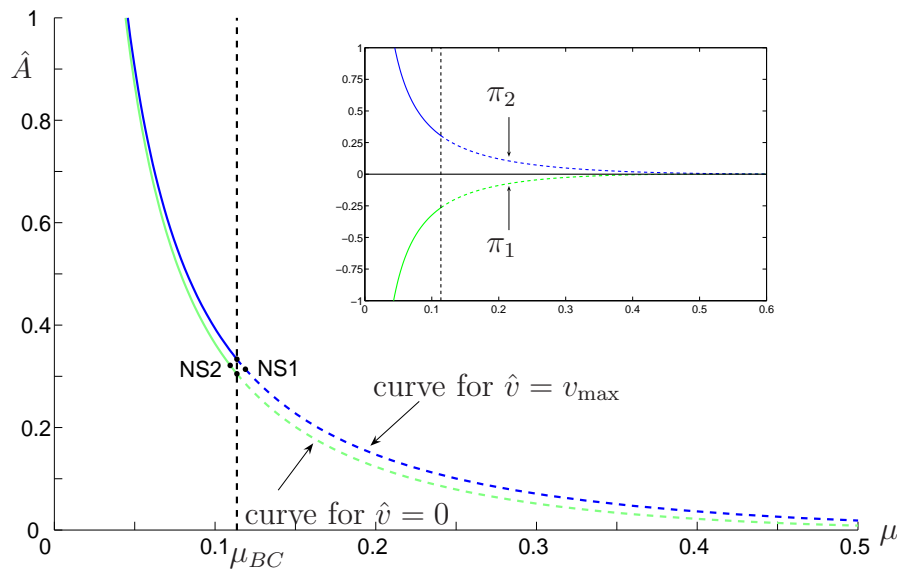
Figure 5.2: Bifurcation diagram with respect to the parameter μ for the steady states $\hat{A}$ with boundary control in the case of Australian IDU

The bifurcation analysis emanates from the black dots located at this line. The green solid line shows how the steady state value with $\hat{v} = 0$ changes when μ decreases. The blue solid line shows how the decrease in μ affects the steady state value of drug use when $\hat{v} = v_{\max}$. The dashed lines show the consequences of an increase in the value of the parameter μ.

The bifurcation lines show that when the outflow μ of drug use is reduced, the steady state value in either boundary control system increases. To the contrary, when more people cease use every year, the steady state numbers decline.

The black dot labeled NS1 located on the blue dashed bifurcation line indicates that at $\mu = 0.1189 > \mu_{BC}$ a bifurcation point is detected in the system with full harm reduction $\hat{v} = v_{\max}$. The corresponding steady state of the canonical system is located at $\hat{A} = 0.3139$, $\hat{S} = 0.1605$, $\hat{\lambda}_A = -3.8777$, $\hat{\lambda}_S = -2.4521$. It is a neutral saddle, and the Jacobian Matrix evaluated at this steady state exhibits the Eigenvalues

$$\begin{aligned}
\xi_1 &= 0.232076, \\
\xi_2 &= -0.192076, \\
\xi_3 &= 0.192076, \\
\xi_4 &= -0.152076.
\end{aligned}$$

In the system without harm reduction, i.e. $\hat{v} = 0$, we encounter the neutral saddle NS2 at $\mu = 0.1094 < \mu_{BC}$. The according steady state values are $\hat{A} = 0.321376$, $\hat{S} = 0.183081$, $\hat{\lambda}_A = -9.052919$, $\hat{\lambda}_S = -5.313421$. The Jacobian Matrix evaluated at this steady state has the Eigenvalues

$$\begin{aligned}
\xi_1 &= 0.211193, \\
\xi_2 &= -0.171193, \\
\xi_3 &= 0.171193, \\
\xi_4 &= -0.131193.
\end{aligned}$$

The crucial property of neutral saddles is that there is a pair of Eigenvalues with opposite signs that sum up to zero. In the present cases, the pairs are made up by ξ_2 and ξ_3 in each case. At the bifurcation line that is crossed here, nothing dramatic happens. The steady states for parameter values close to the critical parameter still exist and are of saddle point stability. Therefore, neutral saddles are not analyzed further.

The subplot of Figure 5.2 shows the Lagrange Multipliers π_1 and π_2 evaluated at the corresponding steady state as a function of the parameter μ. In the neighborhood of the base case parameter $\mu_{BC} = 0.1136$ the Lagrange Multiplier π_1 is negative, whereas π_2 is positive. Only for parameter values as high as $\mu \approx 0.48$ the Lagrange Multiplier π_1 switches to positive values. Increasing μ even further to about $\mu \approx 0.65$ leads to $\pi_2 < 0$. This means when the desistance from IDU is high, the no harm reduction policy becomes a candidate for the optimal solution.

As stated in section 4.2.1, for the base case parameter values for Australian IDU there do not exist steady states with v^* from equation (4.16). Given the previously described change of the signs of the Lagrange Multipliers for values of μ around $\mu = 0.6$, the system with that parameter value was analyzed with respect to bifurcations and feasibility of control. There exists a steady state with $\hat{v} = 0.4131$ at $\hat{A} = 0.0070$, $\hat{S} = 0.5083$, $\hat{\lambda}_A = -3.8480$, $\hat{\lambda}_S = -0.2220$. From there, the bifurcation analysis was conducted. The red curve in Figure 5.3 shows the value of $\hat{A}$ as a function of μ. Feasibility of the
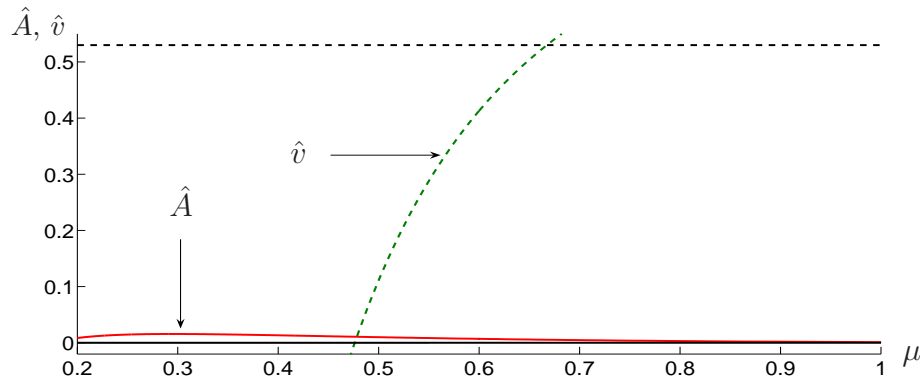
Figure 5.3: Analysis of steady states with control $\hat{v}$ from equation (4.16) under variation of the parameter value of μ in the case of Australian IDU.

steady state solution is given only for those values of the parameter μ, for which the green dashed curve representing control as given by equation (4.16) is located between the horizontal black lines. The black solid line indicates the lower bound $v = 0$, whereas the black dashed line shows $v = v_{\max}$.

From Figure 5.3 we infer that for values between $\mu \approx 0.47785$ and $\mu \approx 0.66725$ there exists a steady state with feasible control $0 < \hat{v}_i < v_{\max}$. Those values are close to the values for which the Lagrange Multipliers at the steady states of both boundary control systems are positive. Although from a mathematical point of view the investigation of the system with such values of the parameter μ is interesting, we omit those investigations due to the fact that quitting rates between $\mu = 0.48$ and $\mu = 0.67$ are quite high compared to the base case value $\mu_{BC} = 0.1136$.

Bifurcation Analysis with Respect to the Parameter b for Australian IDU

The results from the bifurcation analysis for IDU in Australia with respect to the parameter b are shown in Figure 5.4. When the value of b declines compared to the base case value $b_{BC} = 0.5112$, the steady state number of users in both boundary control scenarios declines, which is represented by the solid lines (green for $\hat{v} = 0$, blue for $\hat{v} = v_{\max}$). For increased values of b, the steady state numbers grow, which is visible from the dashed lines.

Neutral saddles are encountered, which are labeled NSm1, NSm2 (system with $\hat{v} = v_{\max}$, blue bifurcation curve), NSn1 and NSn2 (system under $\hat{v} = 0$, green bifurcation line). Those steady states are not investigated in detail.
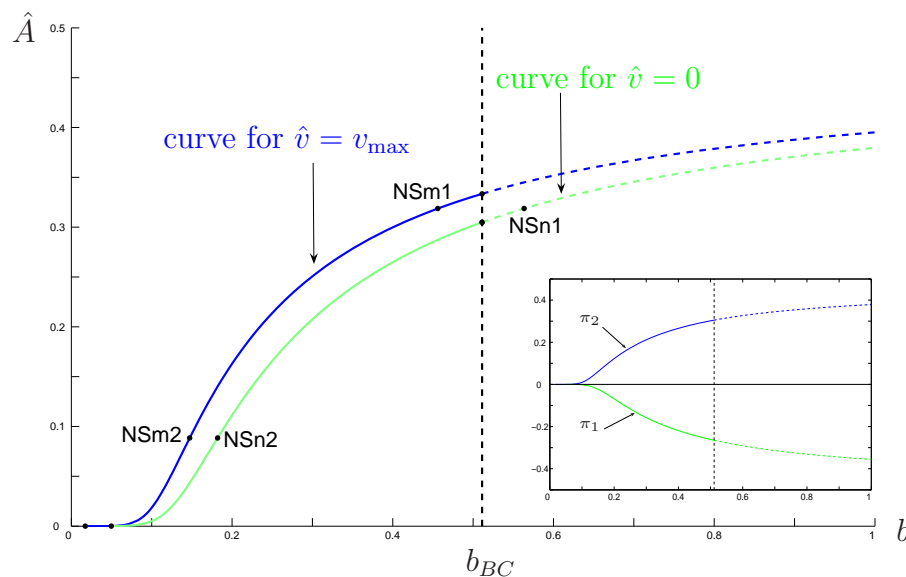
Figure 5.4: Bifurcation diagram with respect to the parameter b for the unique interior steady states $\hat{A}$ with boundary control for IDU in Australia

The Lagrange Multipliers π_1 and π_2 along the bifurcation lines are depicted in the little subplot of Figure 5.4. For any value of b the upper boundary control steady state is a candidate for optimality $(\pi_2 > 0)$, whereas the Lagrange Multiplier π_1 for the lower boundary control steady states is negative and hence this steady state is not a candidate.

It was investigated whether there exist steady states with v^* from equation (4.16) for values of the parameter b different from the base case parameter value $b_{BC} = 0.5112$. Such steady states do not exist.

Bifurcation Analysis with Respect to the Parameter μ for U.S. cocaine

Figure 5.5 shows the bifurcation plot with respect to the parameter μ for the steady states in the case of the U.S. cocaine epidemic. The blue curve gives the bifurcation line under full harm reduction $\hat{v} = v_{\max}$, the light green curve is for the system without harm reduction $\hat{v} = 0$. The solid lines depict the steady state numbers at the high-use steady states as a function of the parameter μ, the dashed lines show the steady state numbers of the intermediate-use equilibria when μ changes. The dark green curve is the bifurcation plot for the steady state with control $\hat{v}$ from equation (4.16). At

the base case parameter value $\mu_{BC} = 0.1661$, there is a vertical dashed line in black, denoting where the computations for the bifurcation analysis begin.

The bifurcation lines show that when μ decreases the high-use steady state's number of users increases, whereas the number of users in the intermediate-use steady state decreases. When the parameter value μ increases, the high-use steady state number of users decreases, whereas $\hat{A}$ increases at the intermediate-use steady state.

Along any of the bifurcation lines, there occurs a blue sky bifurcation. The critical parameter values are $\mu_c^{\hat{v}=0} = 0.2060$ for the light green bifurcation line of the system with lower boundary control $\hat{v} = 0$, $\mu_c^{\hat{v}=v_{\max}} = 0.2181$ for the blue bifurcation line of the system with upper boundary control $\hat{v} = v_{\max}$, and $\mu_c^{\hat{v}_i} = 0.4822$ for the dark green bifurcation line of the system with $\hat{v}$ from equation (4.16). The corresponding steady states of the four-dimensional system are located at $\hat{A} = 2.3396$, $\hat{S} = 14.2123$, $\hat{\lambda}_A = -176.9738$, $\hat{\lambda}_S = -44.6426$ for $\hat{v} = 0$, at $\hat{A} = 2.2096$, $\hat{S} = 14.2123$, $\hat{\lambda}_A = -227.6043$, $\hat{\lambda}_S = -57.4144$ for $\hat{v} = v_{\max}$, and at $\hat{A} = 0.721935$, $\hat{S} = 16.423249$, $\hat{\lambda}_A = -0.409003$, $\hat{\lambda}_S = -0.071234$, with control $\hat{v} = 1.04056$ which is not feasible.

Apart from these blue sky bifurcation points labeled BSn, BSm and BSi on Figure 5.5, other bifurcation points are found. Several neutral saddles are encountered which are not discussed in detail. The occurring Hopf bifurcation points Hn and Hm are of greater interest. This type of bifurcation denominates the event that a pair of conjugate complex Eigenvalues is purely imaginary. In the phase portrait around such a steady state periodic solutions occur.

As before, the Lagrange Multiplier of the active boundary control constraints is evaluated along the bifurcation lines for the boundary control steady states. The resulting information is provided in the subplot of Figure 5.5. The Lagrange Multiplier π_1 at the intermediate-use steady states of the uncontrolled system (dashed branch of the light green curve) is negative. Symmetrically in a certain sense, along the dashed branch of the blue curve which gives the bifurcation line of the intermediate-use steady state with upper boundary control, a positive Lagrange Multiplier $\pi_2 > 0$ is computed. We expect that this steady state will be located in the basin of attraction of the optimal no-use steady state and that one can show that it is dominated analogously to what turned out in section 4.5.2 for the base case, but this has to be investigated in detail for any concrete parameter value.

The solid branch of the light green curve, denoting the high-use steady states under the pure use reduction regime, evaluates to negative Lagrange
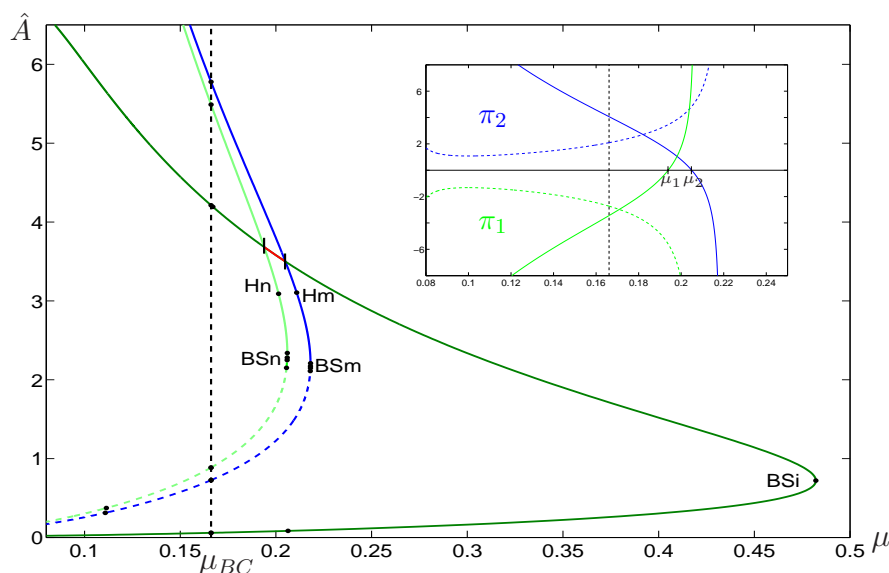
Figure 5.5: Bifurcation diagram with respect to the parameter μ for the steady state values $\hat{A}$ for the U.S. cocaine epidemic

Multipliers π_1 for any $\mu < \mu_1 = 0.1938$. For rates of desistance higher than this, the high-use steady state with $\hat{v} = 0$ has $\pi_1 > 0$. Along the upper branch of the blue bifurcation line the steady states exhibit a positive Lagrange Multiplier $\pi_2 > 0$ for parameters $\mu < \mu_2 = 0.2048$. For fractions of quitters from drug use higher than this, the Lagrange Multiplier π_2 of the upper boundary control steady state with high-use is negative. The parameter values μ_1 and μ_2 are indicated by the black little bars along the blue and light green bifurcation line in Figure 5.5.

Along the dark green bifurcation line for the system with control taken from equation (4.16), the respective control values $\hat{v}$ were computed. Steady states evaluating interior control values $0 < \hat{v} < v_{\max}$ are feasible. The red segment of the dark green curve shows that exactly for parameter values $\mu_1 < \mu < \mu_2$ there exist steady states with interior control values. Furthermore, for parameter values between $\mu = 0.038822$ and $\mu = 0.039225$ control evaluates to feasible values. The corresponding segment of the dark green bifurcation curve is not visible in Figure 5.5. Those steady states are not analyzed here because the corresponding quitting rates are low.

We now continue the discussion of the Hopf bifurcations points. The critical parameter values are $\mu_c^{\hat{v}=0} = 0.2014$ in the system without harm

reduction and $\mu_c^{\hat{v}=v_{\max}} = 0.2108$ in the system with full harm reduction. The steady state for $\mu_c^{\hat{v}=0} = 0.2014$ is located at $\hat{A} = 3.0908$, $\hat{S} = 11.8889$, $\hat{\lambda}_A = -28.7570$, $\hat{\lambda}_S = -9.8495$. The steady state values for $\mu_c^{\hat{v}=v_{\max}} = 0.2108$ are $\hat{A} = 3.1052$, $\hat{S} = 11.3552$, $\hat{\lambda}_A = -19.7673$, $\hat{\lambda}_S = -7.2063$.

From Figure 5.5 we infer that the Hopf bifurcation point Hm evaluates to a Lagrange Multiplier $\pi_2 < 0$. Thus, it is not a candidate for optimality and therefore further investigation is omitted.

The Hopf bifurcation steady state Hn in the system without harm reduction occurs for the parameter value $\mu_c^{\hat{v}=0} = 0.2014$, which happens to lie on the arc of the light green bifurcation line where π_1 is positive. For the critical parameter value, there also exists a steady state with interior control. In order to determine what happens there, a detailed analysis of the system for the critical parameter value has to be conducted. Although this is an interesting endeavor, it is omitted in this thesis.

Bifurcation Analysis with Respect to the Parameter b for U.S. cocaine

Figure 5.6 shows the bifurcation diagram for the steady states when the proportionality constant b for the virulence of the U.S. cocaine epidemic varies. The black dashed line at $b_{BC} = 0.009$ represents the base case parameter value.

The blue and light green curves show the same feature presented as in Figure 5.1, but shed light on the situation from another point of view. Figure 5.1 shows how the intermediate- and high-use steady states on the (A, S)-plane move closer and closer when b declines, and that those steady state finally cease to exist when the parameter passes a critical value. In Figure 5.6 the steady state value $\hat{A}$ is shown as a function of the parameter b. The light green dashed line depicts $\hat{A}$ at the intermediate-use steady state of the system with $\hat{v} = 0$, the solid line shows the high-use steady state number of users in the uncontrolled system. Analogously, the blue dashed line presents the intermediate-use number $\hat{A}$ when $\hat{v} = v_{\max}$, whereas the solid blue line shows the corresponding high-use steady state value of A. When the value of b declines, the values along the high- and intermediate use branch come closer and closer to each other, until they finally collide at the blue sky bifurcation points labeled BSn and BSm in Figure 5.6. At the critical parameter values $b_c^{\hat{v}=0} = 0.006434$ (indicated by the vertical light green dashed line) the blue sky bifurcation happens in the uncontrolled system. The corresponding steady state values are $\hat{A} = 2.901002$, $\hat{S} = 14.212291$, $\hat{\lambda}_A = -81.315631$,
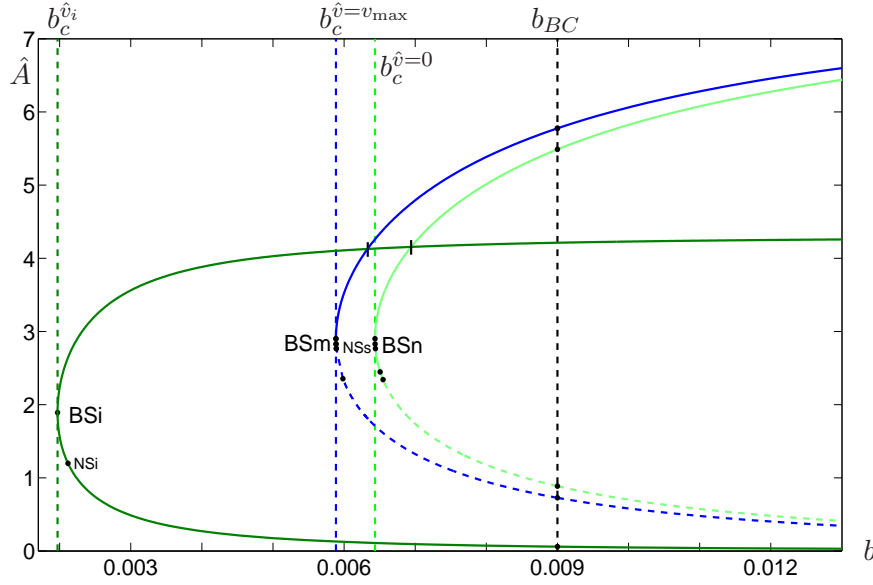
Figure 5.6: Bifurcation diagram with respect to the parameter b for the steady states $\hat{A}$ for the U.S. cocaine epidemic

$\hat{\lambda}_S = -20.512316$. In the system with $\hat{v} = v_{\max}$, we get $b_c^{\hat{v}=v_{\max}} = 0.005885$ which is highlighted by the blue dashed vertical line, and $\hat{A} = 2.901002$, $\hat{S} = 14.212291$, $\hat{\lambda}_A = -67.160206$, $\hat{\lambda}_S = -16.941532$. The critical parameter values $b_c^{\hat{v}=0}$ and $b_c^{\hat{v}=v_{\max}}$ are of course exactly the ones derived from equations (5.1)-(5.3) in section 5.2.1.

With respect to the boundary control steady states, Figure 5.6 directly implies the following conclusion over the existence of such steady states. Please note that in both boundary control systems we encounter the no-use steady state at $(\hat{A}, \hat{S}) = (0, \frac{k}{\delta})$, independent of the value of b.

$$\text{for} \begin{cases} b < 0.005885 & \text{2 steady states at } \hat{A} = 0 \\ b \in (0.005885, 0.006434) & + \text{ 2 steady states with } \hat{v} = v_{\max} \\ b > 0.006434 & + \text{ 2 steady states with } \hat{v} = 0 \end{cases} \quad (5.4)$$

The dark green curve in Figure 5.6 gives the bifurcation plot for the steady states with $\hat{v}$ derived from equation (4.16). When the parameter value of b declines relative to the base case value, there happens a blue sky bifurcation. The critical parameter value is $b_c^{\hat{v}_i} = 0.001971$, with steady state values $\hat{A} = 1.891598$, $\hat{S} = 16.983563$, $\hat{\lambda}_A = -1.455114$, $\hat{\lambda}_S = -0.226214$. Please note that the corresponding control value is $\hat{v} = 1.01861 > v_{\max}$,

which is not feasible. The respective point in the $(b, \hat{A})$-plane on Figure 5.6 is labeled BSi.

At the base case parameter value b_{BC}, a steady states with control $0 < \hat{v}_i < v_{\max}$ was not found (see section 4.2.2). There arises the question whether there exist values of the parameter b for which $0 < \hat{v}_i < v_{\max}$ holds. The answer is shown graphically in Figure 5.7. We first discuss the signs of the Lagrange Multiplier at the boundary control steady states. The according information is given in the left panel of Figure 5.7. The light green dashed line shows that along the intermediate-use branch of the system with $\hat{v} = 0$, there always holds $\pi_1 < 0$. Along the high-use branch (light green solid curve), $\pi_1 < 0$ for large values of b, but at $b = 0.006942$ the Lagrange Multiplier π_1 becomes positive. Symmetrically in a certain sense, along the intermediate-use branch of the system with upper boundary control (blue dashed curve), there holds $\pi_2 > 0$, whereas along the high-use branch (blue solid curve) there holds $\pi_2 > 0$ as long as the value of b is large enough. At $b = 0.006332$ the Lagrange Multiplier π_2 changes its sign. Evaluating $\hat{v}$ along the dark green bifurcation line results in the dark green dashed and red solid line in the right subplot of Figure 5.7. The light green and blue lines show the Lagrange Multipliers π_1 and π_2 at the high-use steady states as a function of the parameter b, the dark green dashed line shows steady state control $\hat{v}$ from equation (4.16). The red line denotes interior control values $0 < \hat{v}_i < v_{\max}$. It shows that steady states with interior control $0 < \hat{v}_i < v_{\max}$ exist for parameter values between $b = 0.0063373$ and $b = 0.0069363$, i.e., those steady states occur for the values of b for which both boundary control steady states exhibit a positive Lagrange Multiplier.

Please note that the simple analysis conducted here only gives a statement on the existence of steady states. The investigation of the dimension of the stable manifold of the steady states, the determination whether dominated equilibria exist among the possibly multiple candidates for optimality, and the analysis whether DNSS curves or indifference curves occur has to be conducted for each parameter value in particular.

Along the bifurcation lines there occur neutral saddles. They are labeled NSn, NSm and NSi in Figure 5.6, but due to the fact that they are neutral saddles, the are not analyzed in more detail. The three dots located close to BSn and BSm points resulting from the continuation of the bifurcation line, they do not denote bifurcation points.
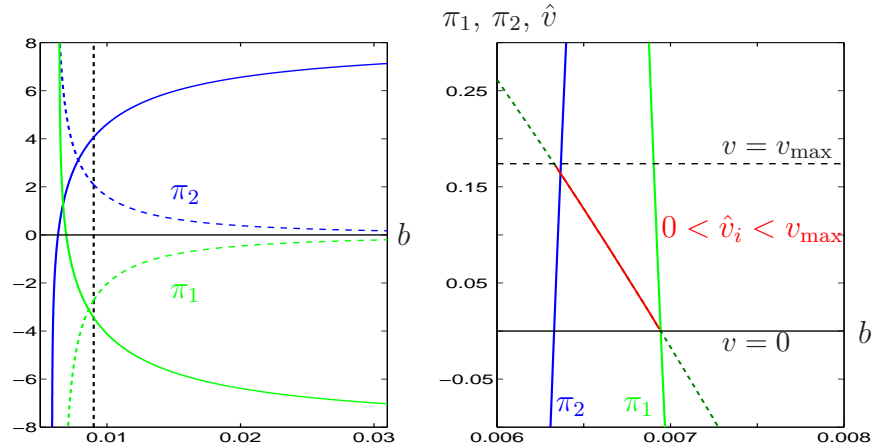
Figure 5.7: Lagrange Multipliers π_1 and π_2, and control $\hat{v}$ when b varies for the U.S. cocaine epidemic

5.2.3 Conclusions from the Bifurcation Analysis

The shifts in boundary control steady state values for U.S. cocaine and Australian IDU correspond to rational expectations how $\hat{A}$ would evolve under variations of b and μ.

The additional investigation of the existence and feasibility of steady states with interior control $0 < \hat{v}_i < v_{\max}$ showed that when the Lagrange Multipliers of the boundary control steady states computed for the U.S. cocaine epidemic change their sign, there exist such steady states with feasible control $0 < \hat{v}_i < v_{\max}$.

Nevertheless, the information derived here does not allow drawing direct conclusions with respect to the optimal phase portrait. For statements on the issue of optimality, a detailed investigation of the system under a new parameter value is necessary. The first section of the subsequent Chapter 6 is devoted to the example of the important problem of a lower parameter value b for the U.S. cocaine epidemic.

Chapter 6

Variations of the Model

The parameterizations derived for the model cannot be expected to reflect exact values. Hence, a sensitivity analysis was conducted and bifurcations were investigated, which gave interesting and important insights. Furthermore, the base case model neglects control cost and innovators into drug consumption. This Chapter presents results from models where parameter values or functional forms are modified.

6.1 Decreasing the Virulence of the U.S. Cocaine Epidemic

There are hypotheses that the virulence of the U.S. cocaine epidemic has dropped at the later stages of the epidemic (Tragler et al., 2001; Caulkins et al., 2004; Johnson et al., 1996). The results of section 3.7 allude to a change in the vector field of the system when the contagion proportionality constant b decreases. This motivated the bifurcation analysis with respect to the parameter b in section 5.2. It revealed that there happens a blue sky bifurcation. When b falls down to about two thirds of its base case value, the intermediate-use and high-use steady states of the system do not exist any longer. Checking the respective Lagrange Multipliers along the bifurcation lines revealed that for a certain interval around the critical parameter values, both boundary control steady states with a high number of users exhibit a positive Lagrange Multiplier. Hence both are candidates for optimality.

Consequently, it is interesting to determine the solution to the optimal control model for a lower value of the parameter b. All other parameters

are kept at their base case values derived for the U.S. cocaine epidemic (see Table 2.1).

The parameter $b = 0.0065$ is appropriate to create a system that is as rich as outlined above. Looking back on the existence overview from equation (5.4), we observe that $b = 0.0065$ is large enough, such that six steady states with boundary control exist. Furthermore, for the new parameter value the high-use steady state of the system without harm reduction (upper branch of the green bifurcation line in Figure 5.6) is located on the arc where $\pi_1 > 0$. On the other hand, the new $b = 0.0065$ is larger than $b_2 = 0.006332$, where the Lagrange Multiplier π_2 becomes negative. Hence, the high-use steady state with upper boundary control is still among the candidates for optimality, and the high-use steady state without harm reduction joins the pool of candidates.

6.1.1   Analysis of Steady States

Searching for Steady States with Interior Control

Before going into depth with the analysis of the boundary control steady states, the existence of steady states with interior control has to be assessed. Two steady states are encountered. The first one is located at

$$(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (0.107571, 21.8815, -0.9175, -0.0074)$$

with control $\hat{v} = 1.04678 > v_{\max}$, so it is not feasible. The second one is located at

$$(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (4.13633, 10.8208, -18.4676, -7.1501)$$

with control $\hat{v} = 0.1271 \in [0, v_{\max}]$. It is feasible, but the Jacobian Matrix of the system evaluated at this steady state exhibits the Eigenvalues

$$\begin{aligned}
\xi_{1,2} &= 0.02 \pm 0.114257\,\mathrm{i}, \\
\xi_3 &= 0.107323, \\
\xi_4 &= -0.0673225.
\end{aligned}$$

It has only a one-dimensional stable manifold, and can thus be excluded.

Boundary Control Steady States with $\hat{v} = 0$

Next, the three boundary control steady states with $\hat{v} = 0$ are investigated. The first one is located at

$$(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (0, 22.1769, -4.8520, 0)$$

with Lagrange Multiplier $\pi_1 = 0$, where analogously to the investigations in section 4.3.4 there holds $\lim_{A\searrow 0, S\to \frac{k}{\delta}} \pi_1 = 0^{(-)}$.

The second one is the intermediate-use steady state at

$$(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (2.46965, 15.3965, -813.324, -170.433) =: \hat{P}_{\hat{v}=0}.$$

It exhibits a positive Lagrange Multiplier $\pi_1 = 121.997$, but the Jacobian Matrix evaluated at the steady state has the Eigenvalues

$$\begin{aligned}
\xi_1 &= 0.0718859, \\
\xi_2 &= 0.0378254, \\
\xi_3 &= -0.0318859, \\
\xi_4 &= 0.00217463.
\end{aligned}$$

Only one of them is negative, thus there is only a one-dimensional stable manifold leading to the steady state $\hat{P}_{\hat{v}=0}$. As a candidate for optimality it is hence of no relevance. The following investigation shows that the one-dimensional stable manifold still plays a role in the optimal phase portrait.

The third fixed point with boundary control $\hat{v} = 0$ is the high-use steady state located at

$$(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (3.35206, 12.9739, -40.8588, -12.2265) =: \hat{C}_{\hat{v}=0}$$

with a positive Lagrange Multiplier $\pi_1 = 4.17192$. The Eigenvalues of the Jacobian Matrix are

$$\begin{aligned}
\xi_{1,2} &= 0.0451663 \pm 0.0383409\,\mathrm{i}, \\
\xi_{3,4} &= -0.00516634 \pm 0.0383409\,\mathrm{i}.
\end{aligned}$$

One of the two pairs of conjugate complex Eigenvalues has a negative real part, which classifies the steady state as fixed point of saddle type having a two-dimensional stable manifold. Hence, it is a candidate for an optimal long-run steady state, which motivates the notation C.

Boundary Control Steady States with $\hat{v} = v_{\mathbf{max}}$

We proceed with the three boundary control steady states with upper boundary control. The numerically determined no-use steady state is located at

$$(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (0, 22.1769, -4.0074, 0) =: \hat{C}_{\hat{v}=v_{\max}}$$

with a Lagrange Multiplier $\pi_2 = 0$. Analogously to the investigations in section 4.3.4 it can be shown that there holds $\lim_{A \searrow 0, S \to \frac{k}{\delta}} \pi_2 = 0^{(+)}$. This means that when conducting the backwards calculation from points close around the no-use steady state, the Lagrange Multiplier is positive and the computed solutions are optimal. With respect to the costate λ_S the argumentation is also analogous to section 4.3.4. There holds $\lim_{A \searrow 0, S \to \frac{k}{\delta}} \lambda_S = 0^{(-)}$, which corresponds to our expectation of negative costates. The Eigenvalues of the linearized system around the steady state $\hat{C}_{\hat{v}=v_{\max}}$ are

$$
\begin{aligned}
\xi_1 &= 0.2061, \\
\xi_2 &= -0.1661, \\
\xi_3 &= 0.1005, \\
\xi_4 &= -0.0605.
\end{aligned}
$$

They are real and two of them are negative, thus the saddle node $\hat{C}_{\hat{v}=v_{\max}}$ is a candidate for optimality, too.

The intermediate-use steady state of the system with full harm reduction is encountered at

$$
(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (1.65192, 17.6416, 45.0157, 6.0329) =: \hat{D}^1_{\hat{v}=v_{\max}}.
$$

It has a positive Lagrange Multiplier $\pi_2 = 7.62217$, but the costate values are positive in this equilibrium. This does not correspond to the expectation that the shadow prices of users and susceptibles in the model should be negative. It can easily be shown that remaining at the steady state $\hat{D}^1_{\hat{v}=v_{\max}}$ is dominated by a trajectory that emanates from the steady state's projection to the (A, S)-plane, but leads to the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$. The fact that it is dominated is the reason for the label D.

The high-use steady state of the four-dimensional system with $\hat{v} = v_{\max}$ is

$$
(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (4.33609, 10.2723, -15.4585, -6.35265) =: \hat{D}^2_{\hat{v}=v_{\max}}.
$$

It exhibits a positive Lagrange Multiplier $\pi_2 = 0.67549$. The Jacobian Matrix evaluated at this point has Eigenvalues

$$
\begin{aligned}
\xi_{1,2} &= 0.0587654 \pm 0.0752474\,\mathrm{i}, \\
\xi_{3,4} &= -0.0187654 \pm 0.0752474\,\mathrm{i}.
\end{aligned}
$$

The high-use steady state $\hat{D}^2_{\hat{v}=v_{\max}}$ is of saddle type and principally represents another candidate for optimality. Nevertheless, the subsequent analysis reveals that it is dominated by the no-use steady state $\hat{C}_{v=v_{\max}}$.

6.1.2 Determination of Optimal Steady States

This is a typical case of multiple candidates for the optimal long-run steady state(s). The information on the dominated steady states $\hat{D}^1_{\hat{v}=v_{\max}}$ and $\hat{D}^2_{\hat{v}=v_{\max}}$ is given in advance to make notation clear, but actually it turns out only later during the backward calculation. This backward calculation to analyze the structure of the phase portrait around the fixed points is now presented in several steps.

Possible Basin of Attraction of $\hat{C}_{\hat{v}=0}$

The first step is the backward calculation around the high-use steady state of the system without harm reduction, $\hat{C}_{\hat{v}=0}$. Please note that it is depicted as a blue dot in Figure 6.1. The trajectories determined by backward calculation with control $v = 0$ are depicted in light gray in Figure 6.1. Above we found that the neighboring steady state $\hat{P}_{\hat{v}=0}$ (the blue cross in Figure 6.1) exhibits a one-dimensional stable manifold. The gray pocket shown in Figure 6.1 is the projection of that manifold onto the (A, S)-plane. Parts of the gray trajectories shown in Figure 6.1 will turn out to be indeed optimal, whereas for certain domains of the regions in the pocket there will exist trajectories that lead to another steady state an perform better in terms of maximization of the negative objective functional.

Dominated Steady States

The second step is the backward calculation from the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$. Figure 6.2 shows that the projections of the steady states $\hat{D}^1_{\hat{v}=v_{\max}}$ and $\hat{D}^2_{\hat{v}=v_{\max}}$ onto the (A, S)-plane happen to fall in the basin of attraction of the no-use steady state $\hat{C}_{v=v_{\max}}$. The steady states are shown as green crosses. Figure 6.2 depicts the trajectories emanating from the (A, S)-values of $\hat{D}^1_{\hat{v}=v_{\max}}$ and $\hat{D}^2_{\hat{v}=v_{\max}}$, respectively, and ending up in the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$ as black curves. Making use of formula (4.25), one finds that remaining at the intermediate-use steady state $\hat{D}^1_{\hat{v}=v_{\max}}$ forever yields a value of the objective functional of $J = -34.1087$, whereas the path leading to the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$ results in $J^* = -22.4205$. Staying forever at the fixed point with a high numbers of users at $\hat{D}^2_{\hat{v}=v_{\max}}$ yields $J = -89.5316$, compared to the value $J^* = -81.3576$ resulting for the trajectory that converges to the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$. Hence, the intermediate-use steady state $\hat{D}^1_{\hat{v}=v_{\max}}$ and high-use steady state $\hat{D}^2_{\hat{v}=v_{\max}}$ are both dominated by the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$.
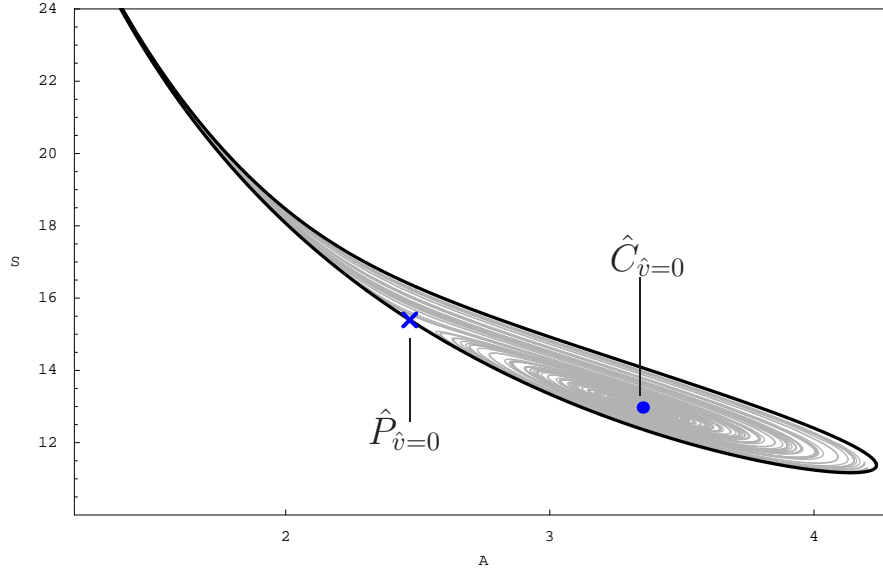
Figure 6.1: Trajectories with $v = 0$ around the high-use steady state $\hat{C}_{\hat{v}=0}$, enclosed by a pocket created by the one-dimensional stable manifold of the steady state $\hat{P}_{\hat{v}=0}$. Note that some of the gray trajectories are not optimal, which turns out only later.

The red dots shown in Figure 6.2 are the steady states from the un-restricted maximization problem that were not feasible. The blue symbols (dots and cross) show the steady states with lower boundary control $\hat{v} = 0$, whereas the green crosses depict the steady states with upper boundary control $\hat{v} = v_{\max}$. The curves in blue and green indicate where switches in the system happen. When interpreting Figure 6.2, please note that the Figure only provides the information on the switches necessary for this part of the analysis; the large region in the upper right of Figure 6.2 still has to be analyzed.

For the two trajectories (black) shown in Figure 6.2, optimal control at the initial points and during the first part of the planning horizon is $v^* = 0$. At the green curve, the Lagrange Multiplier π_1 of the active constraint hits zero, meaning that the boundary control solution $v = 0$ is no longer optimal. The optimized system switches to interior control. Harm reduction v then gradually increases, until optimal control v^* hits the upper bound at $v^* = v_{\max}$. This is represented by the blue curve. There, the optimized system switches to upper boundary control, which is optimal until the steady state $\hat{C}_{\hat{v}=v_{\max}}$ with zero users is finally reached.
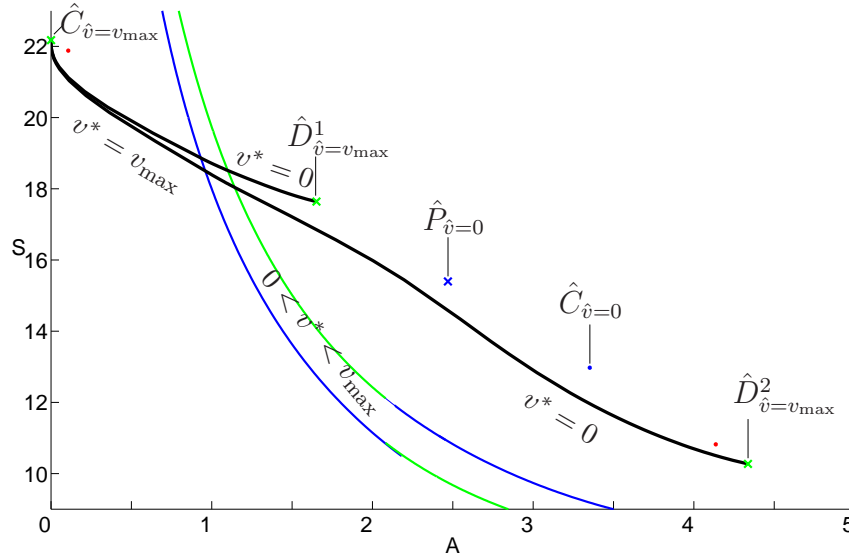
Figure 6.2: The no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$ dominates the steady states $\hat{D}^1_{\hat{v}=v_{\max}}$ and $\hat{D}^2_{\hat{v}=v_{\max}}$.

Indifference Curves and DNNS Curve

The next steps of the backward calculation reveal first an indifference curve for trajectories that eventually approach the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$. In Figures 6.3 and 6.4 it is shown as a blue curve.

Proceeding with the backward calculation, trajectories that eventually lead to the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$ overlap the pocket around the high-use steady state $\hat{C}_{\hat{v}=0}$. Investigation of the Hamiltonian in the critical region reveals a DNSS curve. The DNSS curve is shown in black in Figures 6.3 and 6.4. The occurrence of the DNSS curve reveals that parts of the gray trajectories shown in Figure 6.1 are not optimal. A DNSS curve is different from an indifference curve in view of the fact that emanating from a point on the DNSS curve, the decision maker is indifferent between convergence to different steady states. In the current case, the black curve acts as a separatrix between convergence to the high-use steady state $\hat{C}_{\hat{v}=0}$ and convergence to the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$. The curve is the borderline where two alternatives perform equally optimal in terms of the value of the objective functional. There exists no unique solution; both options are equal in terms of aggregate cost. The curve is a hairline case, for which the decision maker can choose which one of the steady states he or she wants to approach. Lo-
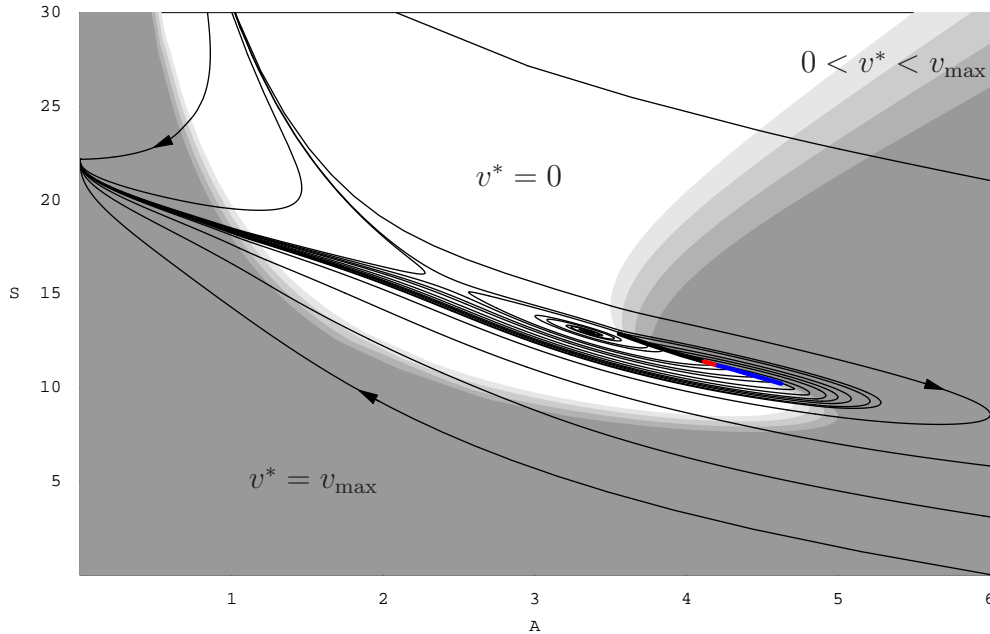
Figure 6.3: Optimal control and optimal phase portrait for the U.S. cocaine epidemic with decreased b

cated a bit to the left or to the right of the DNSS curve, there exists a unique optimal solution.

The final finding when investigating the optimal phase portrait for the case of reduced infectivity of the U.S. cocaine epidemic is that some trajectories wind back around the region where trajectories converge to the high-use steady state $\hat{C}_{\hat{v}=0}$. There exists another indifference curve, which is shown in red in Figures 6.3 and 6.4.

6.1.3 Interpretation of the Indifference Curves and the DNSS Curve

Figures 6.3 and 6.4 depict optimal control as a function of the states A and S. Like in the kindred Figures 4.1 and 4.6 shown for the base case parameterizations, the white region indicates where on the (A, S)-plane optimal control is given by the pure use reduction regime, i.e. $v^* = 0$. At the other extreme, the darkest gray region shows where $v^* = v_{\max}$. The other gray regions stand for optimal interior control values $0 < v_i^* < v_{\max}$. The shading is leveled depending on whether control interventions amount up to one third, two thirds or the full extent of the available harm reduction interventions.
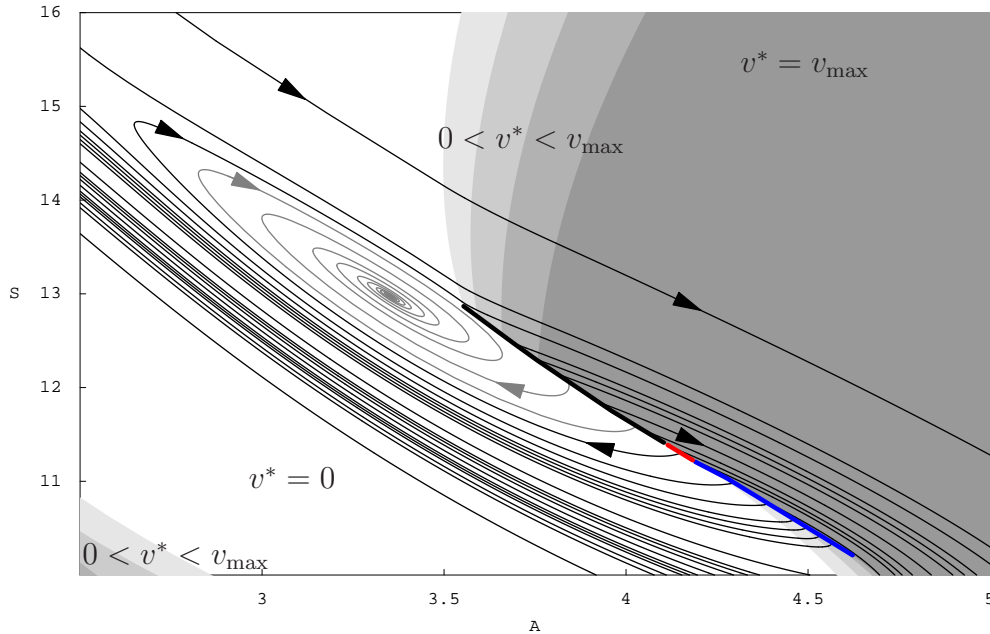
Figure 6.4: Detailed plot of optimal control around the DNSS curve and indifference curves

Exposition of the optimal development of states takes a back seat in that Figures. Only some of the optimal trajectories are depicted in order to show more clearly the structure of the optimal phase portrait. The little arrows are of particular importance, because they help to understand how the states A and S evolve.

Figure 6.4 zooms in to the region where the indifference curves (red and blue) and DNSS curve (black) are located. The curves provide initial conditions for which a decision maker is indifferent between two possibly quite distinct control strategies inducing different courses of the drug epidemic. Indifference stems from the fact that the two strategies evaluate the same value of aggregate social cost. This allows for interesting interpretations. It should be emphasized that indifference curves and DNSS curves are not merely an abstract mathematical feature. To the contrary, they can yield important insights. Grass et al. (2008) bring a brilliant statement on the issue of DNSS curves:

"Multiplicity means that for given initial states there exist multiple optimal solutions; thus the decision-maker is indifferent about which to choose. This explains why such initial states are called points of indifference. In contrast, history-dependence occurs when the optimal solution depends on the

problem's temporal history.

The existence of multiple equilibria has been long-recognized in physics, and more recently has provided an important enrichment of economic control models. In policy-making, it may be crucial to recognize whether or not a given problem has multiple stable optimal equilibria and, if so, to locate the thresholds separating the basins of attraction surrounding these different equilibria. At such a threshold different optimal courses of actions are separated. Thus, in general, starting at a threshold, a rational economic agent is indifferent between moving toward one or the other equilibrium. Small movements away from the threshold can destroy the indifference and motivate a unique optimal course of action. Among the economic consequences of such an unstable threshold is history-dependence (also denoted as path-dependence): the optimal long-run stationary solution toward which an optimally controlled system converges can depend on the initial conditions." (Cited from Grass et al. (2008), p.237.)

In the subsequent interpretations on the indifference curves and the DNSS curve, the following, general insight is important. The drug control policy alternatives that are separated by the curves, and the trajectories of use that stem from those policy options, show a trade-off either between different effects in the short run or between effects in the short and in the long run. As a basic notion for such ideas, please note the following: For any of the trajectories that emanates from the curves, the one strategy always involves a lower initial value $v(0) = v^L$ of harm reduction, whereas the other strategy has a higher initial amount $v(0) = v^H$ of harm reduction. When the option with v^L is chosen, levels of use decline first. When initiating control with $v(0) = v^H$, numbers of users increases first. The detailed interpretation of trade-offs is now given for each curve separately.

Indifference Curve #1

The blue indifference curve contrasts two different scenarios for the development of a drug epidemic in the short run. In the long run, the policies follow an identical control pattern and the cocaine epidemic approaches the no-use steady state. Taking the decision in favor of the higher initial control v^H, which is $v^H = v_{max}$ in this case, the decision maker accepts larger numbers of users at the beginning. The benefit from this is the reduction of social cost to a fraction $1 - v_{max}$ along the first part of the transient. The other option is to choose the lower initial control value v^L, which means in this case either to do no harm reduction first or to begin at some interior level, but let harm reduction decline to $v^* = 0$ in the following time span. That decision leads

to an immediate reduction in the number of users, but initially the decision maker cannot benefit from the reduction in social costs. The tension between reduction of users, but high cost in the objective function on the one side, and reduced social cost in the objective function, but increased in use on the other side, contributes to understanding why the different strategies are finally equivalent in terms of social cost aggregate over the entire planning horizon.

An essential feature of the first strategy is that in the initial phase, where control is at the upper bound $v = v_{\max}$, the number of users increases at the beginning, but the pool of susceptibles is exploited in this phase. Remembering the dynamics of use (see equations (4.4) and (4.5)), the outflow μA from the drug use state is large, when A is large. The initiation term $b\,A^\alpha\,S\,g(v)$ models the inflow to the A-state. It is driven by the contagious effect of current users on susceptibles. Nevertheless, when use is high, but the number of susceptibles is low enough, initiation is rather modest. At such circumstances, the inflow to state A might not be sufficiently high to keep levels of use growing. This is the case here, although harm reduction is applied with full force. The dynamics themselves lead to the turn in the trajectory that is the cornerstone for the following continuous decrease in the number of users.

What happens here is similar to what we have found for the base case parameterization for the U.S. cocaine epidemic (see section 4.5.2). Namely, even though the model assumes that harm reduction increases initiation relative to the "no harm reduction"-alternative, this does not hamper declines in use when the full harm reduction strategy is applied.

DNSS Curve and Time Paths

The black curve in Figures 6.3 and 6.4 is the DNSS curve. The idea is in principle the same as for the indifference curve: For a certain initial condition in (A, S)-space, there are two trajectories that yield the same value of the objective functional. However, the important difference is that in the case of a DNSS curve the trajectories approach different steady states.

For initial conditions located on the black curve, the decision maker can impose a pure use reduction regime, i.e. $v^* = 0$. Under this control choice, the system approaches the steady state $\hat{C}_{v=0}$ with $\hat{A} = 3.35206$ million users and $\hat{S} = 12.9739$ million susceptibles. Two of the corresponding trajectories are shown in gray in Figure 6.4. The little arrows indicate that the system spirals clockwise into the boundary control steady state $\hat{C}_{v=0}$. This option

chooses the lower one of the available initial control values, i.e. $v(0) = 0 = v^L$, whereas $v^H > 0$. This is a simple strategy in the sense that optimal control v^* does not change over time. Corresponding to what we found above, use declines first. Later, it curls back to an increasing number of users. Finally, it spirals into the steady state.

The alternative option to depart from an initial condition located on the DNSS curve involve some non-zero level of harm reduction $0 < v^H \leq v_{\max}$. This yields the trajectories that emanate from the black curve and go to the right, which means that use increases first. Later on, the number of users declines and after some switches in the control policy, and the no-use steady state $\hat{C}_{v=v_{\max}}$ is reached. Please note that the black trajectories representing the alternatives to the gray trajectories without harm reduction in Figure 6.4 both start with full harm reduction $v(0) = v_{\max}$.

Panel b) of Figure 6.5 shows the time paths for the numbers of users along the two gray trajectories in Figure 6.4 that converge to the high-use steady state $\hat{C}_{\hat{v}=0}$. The corresponding control $v^* = 0$ is shown in panel d). Panels a) and c) of Figure 6.5, in contrast, show the alternative control options as a function of time and the course of the epidemic stemming from them. Panel c) shows a tub-shaped form of the mixed control strategy. The number of users (panel a)) converges to the no-use steady state $\hat{C}_{v=v_{\max}}$ when this more complex control strategy is applied. It is given by maximum boundary control at the early and at the late stages, minimum boundary control at some intermediate time spans, and quick transitions between those extremes.

Apart from the fact that the mixed control strategy that induces convergence to the no-use steady state $\hat{C}_{\hat{v}=v_{\max}}$ is so very different from the pure use reduction regime represented in panel d), the underlying time spans outline that the paths leading to the high-use equilibrium $\hat{C}_{v=0}$ take much longer to converge to the steady state than the alternatives do. Panels a) and b) illustrate graphically the difference of either an increase or a decrease of A during the first phase.

Please note that the feature described above is very interesting. The DNSS curve that occurs in the current two-state model of drug epidemics provides the possibility that departing from the DNSS curve to the left (which in the present model means that use declines) lets the system eventually converge to a high-use steady state. Vice versa, emanating to the right and allowing for increasing numbers of users, the system eventually ends up in a steady state without use. Summarizing, starting left leads to a right equilibrium, whereas starting right lets the system end up left.

Beyond this, please note that whereas the mixed strategy causes a single

a) A

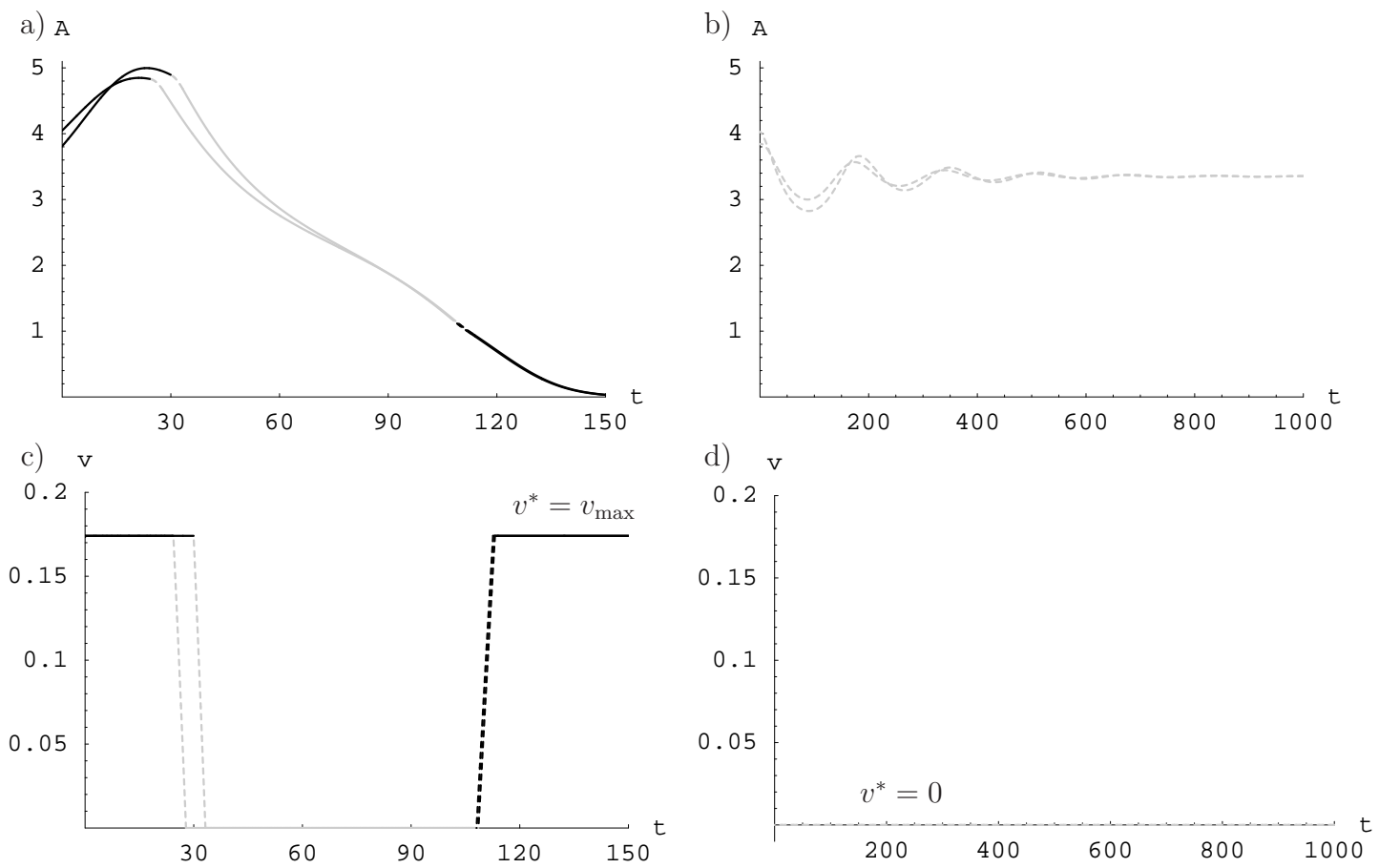b) A

c) v

$v^* = v_{\max}$

d) v

$v^* = 0$

Figure 6.5: Time paths for initial conditions located on the DNSS curve

peak in use that is followed by a continuous decline to zero use, the pure use reduction regime leads to damped oscillations in the number of users. After every 200 years, there is a new peak in use, although over time the peaks are less and less pronounced. The axes of panels a) and b) are equally scaled, hence we infer that the pure use reduction regime only traverses a narrow interval of numbers of users, whereas the difference in minimum and maximum use is much larger when the mixed control strategy is used.

At the hairline case of the DNSS curve, the decision maker is indifferent between the scenarios "reductions in use first, high-use steady state in the long run" and "increases in use first, no-use steady state in the long run". It is a hairline case, but also a separatrix. For any small movement away from the black curve in Figures 6.3 and 6.4, there is a unique optimal strategy. Slightly to the left on the A-axis, convergence to the steady state $\hat{C}_{v=0}$ under a pure use reduction regime is less costly than any strategy that involves harm reduction interventions. For only slightly higher numbers of users and/or susceptibles, the pure use reduction policy is no longer an option. The optimal suggestion is to apply harm reduction. Numbers of users increase in the short run. This may be interpreted as a downside, but the argument that sheds a different light on the supposed downside is that in the long run, a steady state where drug use is eradicated is approached.

Indifference Curve #2

The interpretation of the indifference curve depicted in red mainly takes the same line as the above interpretations. The first option is to choose a strategy with fully implemented harm reduction first. It increases use at the early stages, but then the A-state starts to decline. Subsequently, the control policy involves the usual switches to optimally approach the no-use steady state $\hat{C}_{v=v_{\max}}$. If the decision maker chooses his second option, which consists in the pure use reduction regime, this leads to decreasing numbers of users as an immediate effect. Along the associated trajectories, the pool of susceptibles grows. Although use declines, many new users are recruited out of the large pool of susceptibles. This tips the epidemic over to increasing drug use. Very soon, the number of users reaches a size for which it is favorable to do harm reduction in order to moderate the social cost borne by society. Finally, the same number of users as at the initial condition $A(0)$ is reached again, but now with a higher level of susceptibles. Then, one follows the same optimal control policy than along the alternative path. Compared to the number of users along the alternative path, one easily detects an overshoot in the number of users.

The strategy that begins without harm reduction postpones the increase in use at the expense of a higher peak in the number of users later in the course of the epidemic. The considerable decline in use at the early stages combined with the overshoot at later states is opposed to immediate increases in use that lead to a less pronounced peak.

6.1.4 Conclusions

Regarding the dominated steady states $\hat{D}^1_{v=v_{\max}}$ and $\hat{D}^2_{v=v_{\max}}$, we identified conditions for which staying at a fixed point under application of full harm reduction yields higher aggregate social costs than applying a mixed policy. Figure 6.2 shows that in the case of either of the dominated steady states, the mixed policy is made up by a pure use reduction regime first, followed by increasing control v^*. Later, the full harm reduction strategy is pursued. Along the entire transient, the number of users declines. Harm reduction is implemented only when the drug epidemic has dropped down to levels of use at which there is no more danger that use could be affected in an adverse manner due to a coupling of effects of $g(v)$ and the feedback from current users on initiation.

For the U.S. cocaine epidemic's base case parameterization, section 4.5.2 revealed that the intermediate-use steady state of the system with full harm reduction was dominated by the optimal no-use steady state where full harm reduction is applied. This result is persistent under the new parameterization with decreased b. In the base case, the high-use steady state of the system with full harm reduction $\hat{v} = v_{\max}$ was the second optimal long-run steady state. This property vanishes when the proportionality constant for contagion, b, declines from its base case value $b = 0.009$ to $b = 0.0065$, which represents a decline of 27.8%. In the current parameterization, the high-use steady state is dominated by the no-use steady state, too.

Although there are dominated steady states, we still encounter two optimal long-run steady states. In the base case, the basins of attraction of the two optimal long-run steady states are non-overlapping. In the present case, there exists a DNSS curve. It has been described and interpreted in detail above. Comparison of Figures 6.3 and 4.6 shows that the optimal control structures have many parallels, though there are also some differences.

In terms of the optimal phase portrait there are big differences between the base case b and the new value. When b is larger (base case) the no-use steady state can only be approached if the initial condition involves a number of users that is low enough. If use has grown beyond certain

thresholds given by the stable manifold of the intermediate-use steady state of the two-dimensional system (see Figure 3.4), the optimal solution is always to approach a steady state at endemic levels of use. Very different to this, in the optimal phase portrait stemming from the new parameterization, there is only a tiny region for which a steady state with a high number of users is achieved. For most of the initial conditions on the (A, S)-plane, use should be eradicated.

The conclusion from this difference is an apparent need to find the answer to the question "Has the infectivity of cocaine in the U.S. indeed declined, and if so, which new parameter value for b best represents this decline in the SA Harm Reduction Model?" Remember that the current example reduces $b = 0.009$ by 27.8% to $b = 0.0065$. The answer may have a crucial impact - not so much on the principal policy recommendation, because in a rough generalization both the base case and the current case suggest full harm reduction at low and very high levels of use. However, at intermediate levels the value of the parameter b definitely affects which steady state is approached in the long run. For a decision maker in the U.S. it makes a big difference whether a drug epidemic that has grown to some 4-6 million users can be eradicated by a smart mix of control interventions, or if that mix of interventions is only able to let the epidemic approach a steady state with more than 5 million users.

6.2 Modeling Innovators into Drug Use

The previous results involve the possibility that cocaine use in the U.S. is fully eradicated in the long run. Experts often comment on that feature stating that any steady state with zero drug use is rather unrealistic. They argue that it is impossible to totally suppress drug use in a free society. Their opinion is that there will always be a number of people, albeit perhaps only a modest number, who consume drugs. Related to such comments, we present now a slight update of the parameters of the initiation function which shifts the no-use steady state of the base case model for the U.S. cocaine epidemic to the right. With the new parameter set, the lowest possible value of $\hat{A}$ amounts to some thousand cocaine users.

In the preceding sections of this thesis, initiation into drug use was exclusively driven by the social interaction between current drug users and persons being susceptible to drug use. Although most new drug consumers are "recruited" due to such social network effects, some new users enter the drug using population due to some internal reason. They start to take a

drug because they are curious, by shifting from other drugs, or due to some other internal impetus that is not influenced by current users. Following the spirit of Bass (1969) and harking back to the jargon of diffusion models, this group of initiates is called "innovators". That label contrasts them with the "imitators" who are introduced to consumption by a friend, sibling or acquaintance who is already using that drug. The base case initiation function (2.3) presented in section 2.4 incorporates the coefficient of innovation, τ. Nevertheless, the base case parameterizations had $\tau = 0$. Here, the same initiation function is used

$$I(A, S, v) = (\tau + bA^\alpha)\, S\, g(v), \quad \text{but now with} \quad \tau > 0.$$

The other base case parameter values listed in Table 2.1 are not adjusted; the coefficient of innovation τ is introduced additionally. This actually increases the initiation term slightly, but the parameterizations used here are anyway by no means stalwart, so we accept this inaccuracy.

Application of Pontryagin's Maximum principle to the model with τ is very similar to the base case. The details are omitted, but the canonical system with innovators which now reads

$$
\begin{aligned}
\dot{A} &= (\tau + b\,A^\alpha)\, S\, g(v) - \mu\, A, \\
\dot{S} &= k - \delta\, S - (\tau + b\,A^\alpha)\, S\, g(v), \\
\dot{\lambda}_A &= 1 - v + (r + \mu)\lambda_A + b\alpha A^{\alpha-1}S(\lambda_S - \lambda_A)\, g(v), \\
\dot{\lambda}_S &= (r + \delta)\lambda_S + (\tau + b\,A^\alpha)\,(\lambda_S - \lambda_A)\, g(v).
\end{aligned}
$$

The Hamiltonian maximizing condition yields interior optimal control

$$v^* = \frac{\Phi}{-c_2\omega}\left(\left(\frac{\Psi_\tau\Phi}{c_2\omega\eta}\right)^{\frac{1}{-1+\eta}} - 1\right), \tag{6.1}$$

where the former Ψ from equation (4.15) has to be replaced by

$$\Psi_\tau := \frac{A}{(\tau + b\,A^\alpha)\, S\, (\lambda_A - \lambda_S)}.$$

In the harm reduction versus use reduction debate, several fears about the possible downsides of harm reduction are articulated. Due to possible risk compensation effects, there is the risk of an increase in initiation and the possibility that current users will consume more. A more subtle fear is reflected in claims that "harm reduction might send the wrong message" (cf. MacCoun, 1998). Although this rather addresses legalization issues, one

can interpret it with respect to the reputation of a drug. Harm reduction interventions could make a drug more popular and spread it independent from the social networks. Using the original initiation function from the base case, a susceptible non-user's decision to initiate depends on that person's immediate social environment. Looking at the process of initiation into drug abuse in more detail, it is clear that such direct social factors may give strong incentives to start consumption. Nevertheless, the overall reputation of the drug in society may play a vital role, too. If a drug is portrayed as harmless in movies or news media, this might induce a non-user to try the particular drug, though there are no friends of the individual "seducing" him/her to do so. The innovators τS are exactly such individuals who are susceptible to drug use and try drugs although none of their associates directly encourages that desire. In this sense, the introduction of $\tau > 0$ can be seen as an attempt to pay attention to the fears that harm reduction might send a wrong message.

6.2.1 United States

In the model, τ gives a constant fraction of susceptibles that initiate drug use for other reasons than contagion by a current user. In the LH-model for cocaine abuse in the U.S., Behrens et al. (1999) assume a value of 50,000 people flowing in to light cocaine use in the U.S. each year. In this number, immigration of cocaine users to the U.S. is included, which cannot be considered in the current case, because the innovator group stems from the pool of those among the U.S. population who are susceptible. This is a reason why the number of 50,000 should be toned down for the purpose of the current model. Nevertheless, we first used those 50,000 people and the steady state number $\hat{S}$ of susceptibles from the uncontrolled high-use steady state to determine a preliminary τ. Solving

$$\tau \hat{S} g(v) \overset{v=0}{=} \tau \hat{S} = 7.10762 \cdot \tau = 0.05$$

for τ results in $\tau \approx 0.007$. For this parameter value, the isoclines $\dot{A} = 0$ and $\dot{S} = 0$ of the dynamics of states A and S for the U.S. cocaine epidemic have one intersection, whereas in the base case we encountered multiple equilibria. Using the lower parameter $\tau = 0.0007$ for the calculations results in three steady states in each boundary control system, as was the case in the base case.

Analysis of Steady States

Of course, first one has to search for fixed points of the canonical system with optimal control v^* from equation (6.1). Afterwards, the boundary control steady states are investigated. The computed equilibria $\hat{A}$, $\hat{S}$, $\hat{\lambda}_A$, $\hat{\lambda}_S$ together with corresponding control $\hat{v}$ and/or Lagrange Multipliers π_1 and π_2 of active constraints are summarized in Table 6.1. The three steady states with lower boundary control $\hat{v} = 0$ exhibit negative Lagrange Multipliers π_1. Hence, they are not among the candidates for the optimal solutions. Among the steady states with upper boundary control $\hat{v} = v_{\max}$, only the high-use steady state has a positive Lagrange Multiplier π_2. Thus, this high-use steady state is a candidate for the optimal long-run solution. The first line of Table 6.1 identifies the second candidate. It exhibits interior control $\hat{v}_i = 0.1409$ and has a relatively low number of users $\hat{A}$.

$\hat{v}$	$\hat{A}$	$\hat{S}$	$\hat{\lambda}_A$	$\hat{\lambda}_S$	π_1	π_2	classification
0.1409	0.1999	21.6281	-11.332	-0.170	-	-	candidate
-0.7259	4.1827	10.6933	-13.557	-34.528	-	-	not feasible
0	0.1575	21.7444	-10.102	-0.1195	-0.034	-	not optimal
0	0.6496	20.3934	132.557	6.6294	-7.062	-	not optimal
0	5.5077	7.0556	-10.691	-6.023	-3.492	-	not optimal
$v_{\max}$	0.2186	21.5767	-12.305	-0.203	-	-0.027	not optimal
$v_{\max}$	0.4297	20.9972	-69.727	-2.281	-	-2.257	not optimal
$v_{\max}$	5.7915	6.2767	-7.944	-4.798	-	4.102	candidate

Table 6.1: Equilibria for the U.S. cocaine epidemic when innovation to drug use is modeled.

In order to simplify notation, we denote the interior control steady state listed in the first line of Table 6.1 as $\hat{I}_\tau$ and the full harm reduction steady state listed in the last line of Table 6.1 as $\hat{M}_\tau$.

To determine their stability properties, the linearized system is investigated. The Jacobian Matrix at the interior control steady state $\hat{I}_\tau$ has Eigenvalues

$$
\begin{aligned}
\xi_{1,2} &= 0.0979727 \pm 0.00505549\,\mathrm{i}, \\
\xi_{3,4} &= -0.0579727 \pm 0.00505549\,\mathrm{i}.
\end{aligned}
$$

Among the two pairs of conjugate complex Eigenvalues, one has a positive and the other one a negative real part. Thus, the fixed point is of saddle type exhibiting a two-dimensional stable manifold.

The Eigenvalues of the Jacobian Matrix evaluated at the steady state $\hat{M}_\tau$ with upper boundary control are

$$\begin{aligned}
\xi_{1,2} &= 0.100986 \pm 0.127217\,\mathrm{i}, \\
\xi_{3,4} &= -0.0609861 \pm 0.127217\,\mathrm{i}.
\end{aligned}$$

This indicates that it also has the desired property of saddle point stability and exhibits a two-dimensional stable manifold. Please recall that in the base case solution, the high-use steady state with boundary control $\hat{v} = v_{\max}$ was located at $(\hat{A}, \hat{S}) = (5.7767, 6.31719)$. Here, the optimal high-use steady state $\hat{M}_\tau$ also has upper boundary control and is encountered at $(\hat{A}, \hat{S}) = (5.7915, 6.2767)$. Introducing $\tau = 0.0007$ shifts the high-use steady state only slightly to the lower right in the (A, S)-plane. Compared to the base case values we find an increase of about 0.25% in the number of users and a decrease of about 0.64% in the number of susceptibles in steady state.

Optimal Control

As with the U.S. cocaine epidemic's base case parameterization (see section 4.5.2), the basins of attraction of the two optimal long-run steady states are non-overlapping. Notably, the basin of attraction of the high-use steady state $\hat{M}_\tau$ is quite similar to that of the high-use, upper boundary control, optimal long-run steady state in the base case. Again, an indifference curve can be identified. Its location and the structure of the trajectories are very similar to the base case, so the detailed exposition is omitted.

In the investigation of the SA harm reduction model, the low-use steady state $\hat{I}_\tau$ is the first fixed point that exhibits non-boundary control and indeed has a two-dimensional stable manifold. At that steady state of the optimized system, some harm reduction is done, but the interventions are not realized to the maximum possible extent.

Figure 6.6 shows optimal control as a function of A and S around the low-use steady state $\hat{I}_\tau$ in detail. Some trajectories (black) are shown to hint at the optimal phase portrait. Additionally, the arrows indicate the direction of convergence. The locus where the trajectories meet is exactly the interior control fixed point $\hat{I}_\tau$ at $(\hat{A}, \hat{S}) = (0.1999, 21.6281)$ with steady state control $\hat{v}^* = 0.1409$. The region in state space where the optimal policy is full harm reduction $v^* = v_{\max} = 0.17408$ is shaded in the darkest gray, whereas regions with optimal pure use reduction policy $v^* = 0$ are left white. In between, optimal control is leveled in steps of $\frac{1}{5} \cdot v_{\max}$ now, whereas former kindred Figures had only three levels for interior control. Nevertheless, as in the
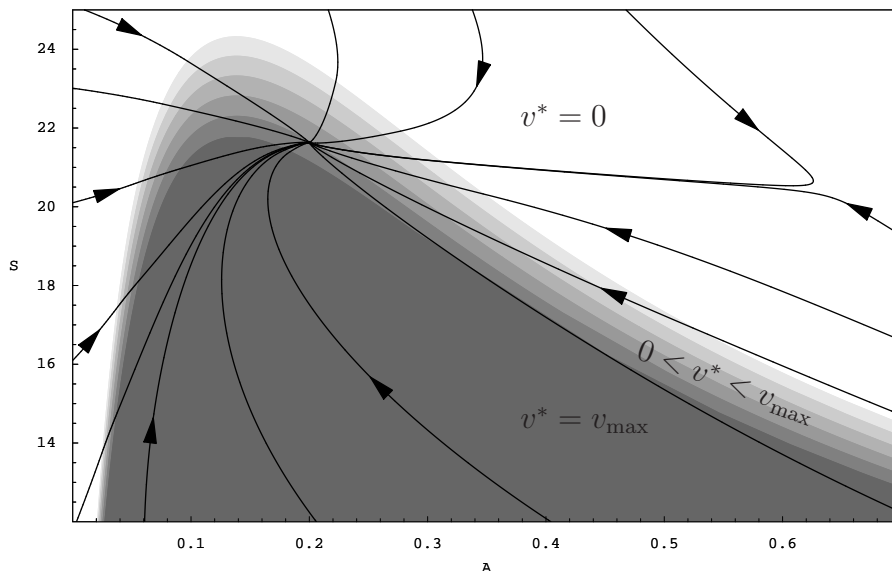
Figure 6.6: Optimal control as a function of A and S around the low-use interior-control steady state $\hat{I}_\tau$.

previous Figures of that kind, the shading is darker the higher the optimal amount v^* of harm reduction is.

When looking at the information on the optimal phase portrait provided by Figure 6.6, please remember that the basins of attraction of the optimal long-run steady states $\hat{I}_\tau$ and $\hat{M}_\tau$ are separated in an asymptotic way. The trajectories in the upper right corner of Figure 6.6 show paths in the basin of attraction of $\hat{I}_\tau$ that are quite close to the separatrix. Initial conditions located slightly on the left of those trajectories fall in the basin of attraction of the high-use steady state $\hat{M}_\tau$. In the part of the basin of attraction of $\hat{M}_\tau$ that is close to the separatrix, optimal control is $v^* = 0$. Hence, the white shading of the upper right corner of Figure 6.6 is appropriate. Nevertheless, one has to be aware that initial conditions located above the two trajectories depicted there belong to the basin of attraction of the other optimal long-run steady state.

Interpretation of Results

In the previously investigated base case parameterization of the U.S. cocaine epidemic and the case of a less virulent epidemic, we had an optimal no-use steady state with upper boundary control $\hat{v}^* = v_{\max}$. In large regions around this steady state, full harm reduction was optimal, too. Although the

model assumes that harm reduction tactics increase initiation (see function $g(v)$ explained in section 2.5.2) harm reduction could be done safely with full force when use was rare. Using $\tau = 0$ the inflow to the pool of users is driven exclusively by the recruitment of new users by current users. Hence, for low numbers of users A there was no risk of exploding increases in use, because the recruitment effect was then only modest. Here we encounter $\tau > 0$, which means that a certain fraction of the susceptibles flows to the drug use state independent of the contagious spread caused by current users. This alters the dynamics and destroys the possibility of no-use steady states. It creates a steady state with a low, non-zero number of users, because no matter whether there are current users or not, some of the susceptibles always start drug consumption.

The shift in the equilibrium number of users is not the only effect of the innovators. The optimal long-run steady state with the low number of users is no longer a boundary control steady state with full harm reduction. The optimal steady state strategy is now control $\hat{v}^* = 0.1409$, which is nevertheless close to the upper bound at $v_{\max} = 0.17408$.

Another observation is that while in the former results control was always increasing just before the full harm reduction region around the steady state was reached, for the current parameterization there exist trajectories along which control is decreasing in the time spans right before the steady state is reached, and others along which control is increasing while the optimized dynamics converge towards the steady state.

Furthermore, regarding the structure of optimal control on the (A, S)-plane, the optimal policy for stages of the epidemic with very low numbers of users and many persons in the susceptible state is not given by full harm reduction as it was the case with the base parameterization of the U.S. cocaine epidemic (see Figure 4.6). The optimal control solution for the new parameterization suggests a no harm reduction policy for those stages of the epidemic. That difference is triggered by the innovators. In the absence of innovators, the stylized model suggests that harm reduction can be safely applied there with full force. In the base case, numbers of users are declining at those stages of the epidemic. Now, there are innovators who push the steady state with the lowest number of users away from zero users to a modest steady state number of users. Beyond this, for low numbers of users, the A-state is increasing along the optimal trajectories. Applying harm reduction at those stages of prevalence induces more susceptibles to become innovators. This leads to further increases in use, which affects the course of the epidemic negatively. Only when a critical mass of drug users is reached, and additionally the pool of susceptibles is of an appropriate size that does

not exceed about 24 million people, will harm reduction start to show its merits. The darkest region of Figure 6.6 is located very close by. There, full harm reduction is the optimal policy. Figure 6.6 shows that in the case with innovators and modest numbers of users, regions of very different recommendations for optimal control are located very close to each other. Please note that in the dark gray region with $v^* = v_{\max}$, there also occur decreases in A. The trajectories that progress through that gray region switch to interior control shortly before the steady state is reached. Those paths are characterized by harm reduction that declines towards the steady state value.

As far as one can generalize the results from the stylized model for the U.S. cocaine epidemic to other epidemics, the differences we found and interpreted above point to a clear need for a decision maker who is confronted with a drug epidemic similar to the U.S. cocaine epidemic, to investigate in all detail whether the initiation process is driven by contagion only or whether there are also innovators. The policy recommendation in the cases $\tau = 0$ and $\tau = 0.0007$ are rather different for low levels of use. Neglecting innovators if the real parameter value was some $\tau > 0$ leads to ill-timed application of harm reduction at low levels of use, which exacerbates the drug epidemic.

For the current U.S. cocaine epidemic, the question of whether $\tau = 0$ or $\tau = 0.0007$ might be of less importance. The policies suggested by the current two-state one-control model differ for low numbers of users, but the current cocaine epidemic has progressed to levels of use far beyond such stages. The base case and the current parameterization differ only marginally in the numbers of A and S at the optimal high-use steady state, the structure of optimal control at high levels of use and the optimized phase portrait around the high-use steady state. The UNODC published a current number of 7,097,000 cocaine users in the United States in the most recent World Drug Report (UNODC, 2008). Hence, if optimal harm reduction interventions were to start today, current initial conditions are located in the basin of attraction of the high-use steady state, where innovators do not make a big difference.

6.2.2 Australia

The fraction $\tau = 0.0007$ was also used to update the initiation term for Australian IDU. This means that each year 0.07% of the susceptibles start a career of drug use due to some impetus that is not affected by current users.

The results of the optimal control model change only marginally. Hence, the exposition is limited to the most basic information. The optimal steady

state $\hat{E}^{\tau}_{\hat{v}=v_{\max}}$ is now located at

$$
\begin{pmatrix} \hat{A} \\ \hat{S} \\ \hat{\lambda}_A \\ \hat{\lambda}_S \end{pmatrix} = \begin{pmatrix} 0.333887 \\ 0.154102 \\ -3.94972 \\ -2.54937 \end{pmatrix}.
$$

The optimal base case steady state $\hat{E}^1_{\hat{v}=v_{\max}}$ is located at $(\hat{A}, \hat{S}, \hat{\lambda}_A, \hat{\lambda}_S) = (0.333455, 0.154617, -3.95724, -2.55002)$ (see section 4.3.1). We identify an increase in steady state use by 0.13%, whereas the steady state number of susceptibles is 0.33% lower than in the base case. The optimized phase portrait and gray level plot of optimal control on the (A, S)-plane are not presented, because there is no visible difference compared to Figure 4.1.

6.3 Linear Harm Reduction Cost Term in the Objective Function

Most optimal control models in the field of drug problems consider the costs that arise from applying drug control interventions. Modeling classic interventions such as treatment, law enforcement, or prevention, the annual budgets the government assigns to those programs are then simply added to the objective function. A similar approach for harm reduction is a less straightforward endeavor, because as explained in preceding sections of this thesis, harm reduction is rather an attitude than a program with a budget. Nevertheless, we here assess what happens to the optimal solutions of the model without innovators, i.e. $\tau = 0$, when control costs are considered. The most basic way to do so is to use a linear form for the function $c(v)$ in the generalized objective function from section 2.5.1, i.e. $c(v) = cv$ with a positive constant c. The new Lagrangian function is then given by

$$
\mathcal{L} = -(A(1 - v) + cv) + \lambda_A \, \dot{A} + \lambda_S \, \dot{S} + \pi_1 v + \pi_2(v_{max} - v).
$$

The new parameter c occurs in the objective function, but not in the state dynamics. In the derivatives of the Lagrangian $\mathcal{L}$ with respect to the states A and S, the cost term cv vanishes. Those two features imply that the new canonical system does not change compared to the base case except for the expression v^* that results from the Hamiltonian maximizing condition. Looking for steady state solutions with interior control, $H_v = 0$ is used. In

the resulting expression for the optimal control, the new parameter c occurs. It is given by

$$v^* = \frac{\Phi}{-c_s\omega}\left(\left(\frac{\Psi_c\Phi}{c_s\omega\eta}\right)^{\frac{1}{-1+\eta}} - 1\right), \tag{6.2}$$

where Ψ from equation (4.15) is modified to

$$\Psi_c := \frac{A - c}{b\,A^\alpha\,S\,(\lambda_A - \lambda_S)}. \tag{6.3}$$

The existence of steady state solutions with interior control was investigated in several of the cases that will be discussed later in this section. We forestall the information that the steady states derived with v^* from equation (6.2) are either not feasible (control out of bounds) or have only a one-dimensional stable manifold.

Consequently, the following analysis concentrates on the results for the boundary control steady states, which are the same as presented in section 4.1. In order to determine whether such a boundary control steady state is a candidate for the optimal long-run solution, the Lagrange Multiplier associated with the active constraint is the first variable of interest. The necessary conditions for optimality (see section 4.1 in the formulation for the base case) state that for an optimal boundary control solution, the corresponding Lagrange Multiplier has to be non-negative. The equations for the Lagrange Multipliers π_1 and π_2 are achieved by setting the derivative of the Lagrangian Function $\mathcal{L}$ with respect to control v equal to zero and solving for π_1 or π_2, respectively. Those expressions are affected by the new cost term cv. We investigate now what happens to the Lagrange Multipliers when c varies.

For boundary control steady states without harm reduction, i.e. $\hat{v} = 0$, the upper control constraint is not active, thus $\pi_2 = 0$. Setting the derivative $\mathcal{L}_v$ equal to zero then yields

$$\pi_1 = c - A + bA^\alpha S(\lambda_S - \lambda_A)g'(0). \tag{6.4}$$

Checking for positivity at a steady state value $(\hat{A}, \hat{S}, \hat{\lambda}_S, \hat{\lambda}_A)$ we get

$$\hat{\pi}_1 > 0 \Leftrightarrow c > \hat{A} - b\hat{A}^\alpha\hat{S}(\hat{\lambda}_S - \hat{\lambda}_A)g'(0). \tag{6.5}$$

Analogously, the Lagrange Multiplier π_2 for the upper boundary constraint is

$$\pi_2 = -c + A - bA^\alpha S(\lambda_S - \lambda_A)g'(v_{\max}). \tag{6.6}$$

Positivity at a steady state, i.e. $\hat{\pi}_2 > 0$, is given as long as there holds

$$c < \hat{A} - b\hat{A}^\alpha \hat{S}(\hat{\lambda}_S - \hat{\lambda}_A)g'(v_{\max}). \tag{6.7}$$

We substitute the boundary control steady states derived in the base case into the above equations (6.5) and (6.7) and conclude for which values of c they are candidates for an optimal fixed point solution. For Australian IDU, the four-dimensional canonical system has a unique saddle point steady state with $\hat{A} > 0$. The analysis here is conducted for this steady state, both for upper and lower boundary control. For the U.S. cocaine epidemic, where we encountered multiple steady states, the analysis focuses on the high-use steady states. In the base case, optimal control at and around the steady states used the highest possible amount of harm reduction. Those results were derived under the assumption that harm reduction is either free of any cost or that costs are so modest that they can be neglected. The simple linear function $c(v) = cv$ helps to derive insights whether and how the optimal control structure changes when costs are taken into account.

Australia

We first analyze the Australian parameterization. The lower boundary control fixed point in the base case is $\hat{E}_{\hat{v}=0}$. It is located at states $\hat{A} = 0.304916$, $\hat{S} = 0.188672$, and has costates $\hat{\lambda}_A = -8.92372$, $\hat{\lambda}_S = -5.13915$. The associated Lagrange Multiplier is negative $\pi_1 = -0.265058$. Hence, the steady state is not a candidate for the optimal solution in the base case. The upper boundary control steady state $\hat{E}_{\hat{v}=v_{\max}}$ is located at $\hat{A} = 0.333455$, $\hat{S} = 0.154617$, $\hat{\lambda}_A = -3.95724$, $\hat{\lambda}_S = -2.55002$. It evaluates to $\pi_2 = 0.304802 > 0$, which identified it as an optimal solution in the base case.

Substituting the concrete steady state values into equations (6.5) and (6.7), we get the relations

$$\pi_1 = -0.265058 + c > 0 \quad \Leftrightarrow \quad c > 0.265058,$$
$$\pi_2 = 0.304802 - c > 0 \quad \Leftrightarrow \quad c < 0.304802.$$

From those relations we conclude the following: For low costs $c < 0.265058$, the Lagrange Multiplier π_1 of the pure use reduction equilibrium is negative, whereas the Lagrange Multiplier π_2 of the full harm reduction steady state is positive. Hence, the latter one is the single candidate for optimality. For a high cost parameter $c > 0.304802$, the steady state without harm reduction control is the only optimal solution, because $\pi_1 > 0$, but $\pi_2 < 0$. For intermediate costs, $0.265058 < c < 0.304802$, the Lagrange Multipliers of both
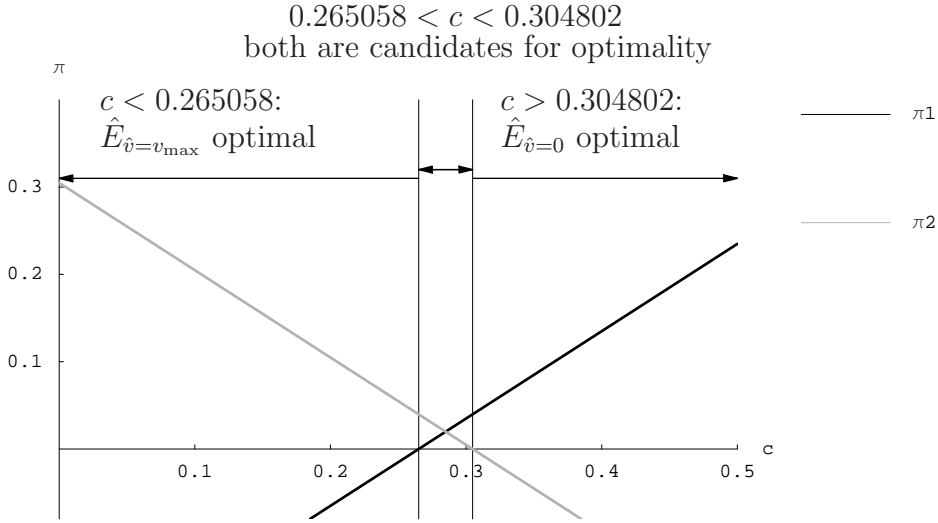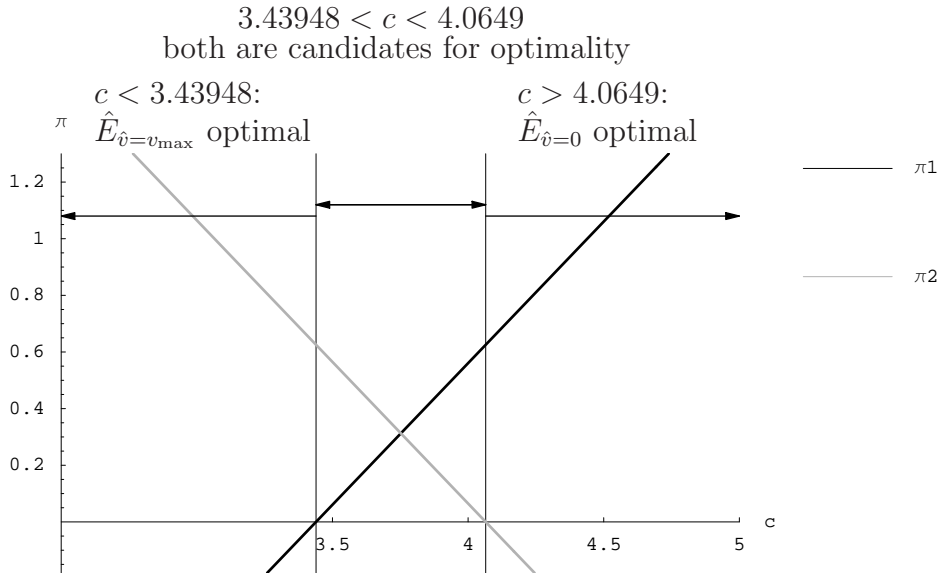
Figure 6.7: The value of the cost parameter c determines which boundary control steady states are candidates for optimality in the model for Australian IDU.

boundary control steady states are positive. In principle, both are candidates for optimality. Figure 6.7 summarizes those results graphically. In the case of two candidates, investigating whether both of them have a two-dimensional stable manifold is necessary. If yes, then one has to analyze whether one is dominated by the other or if there exists a DNSS curve separating their basins of attraction.

United States

Before conducting the analogous analysis of Lagrange Multipliers π_1 and π_2 for the U.S. cocaine epidemic we first recall the high-use steady states from the base case. $\hat{E}^1_{\hat{v}=0}$ is located at $\hat{A} = 5.4888$, $\hat{S} = 7.10762$, $\hat{\lambda}_A = -10.8414$, $\hat{\lambda}_S = -6.07868$. In the base case this steady state was not a candidate for optimality, because it evaluates $\pi_1 < 0$. The upper boundary control steady state $\hat{E}^1_{\hat{v}=v_{\max}}$ is located at $\hat{A} = 5.7767$, $\hat{S} = 6.31719$, $\hat{\lambda}_A = -8.02684$, $\hat{\lambda}_S = -4.83059$, and has a positive Lagrange Multiplier $\pi_2 > 0$ in the base case. Hence, it was a candidate for the optimal long-run solution and turned out as the optimal solution that is approached when initial levels of use are high.

Substituting those steady state values into equations (6.5) and (6.7) re-

Figure 6.8: The value of the cost parameter c determines which boundary control steady states are candidates for optimality in the model for U.S. cocaine.

veals the following: For low cost parameters $c < 3.43948$, the high-use steady state with full harm reduction is a candidate for the optimal solution, whereas the high-use steady state without harm reduction is not. For high parameter values $c > 4.0649$, the situation is the other way round: $\hat{E}^1_{\hat{v}=v_{\max}}$ is no longer among the pool of candidates, because it evaluates $\pi_2 < 0$, but $\hat{E}^1_{\hat{v}=0}$ is a candidate due to its $\pi_1 > 0$. For cost values in an intermediate range $3.43948 < c < 4.0649$, both boundary control steady states are candidates for the optimal long-run solution. As in the related cases above, further investigations have to be conducted for such intermediate values of c. Figure 6.8 gives the graphical overview for the U.S. case.

6.3.1 Cost Parameter Scenarios for Australia

The particular cases investigated are $c = 0.14$, $c = 0.28$, and $c = 0.5$. The base case analyzed in section 4.5.1 is the special case $c = 0$.

Steady States

From the findings on the influence of the cost parameter c on the steady state values of the Lagrange Multipliers, in the case $c = 0.14$, we conclude

that the steady state with $\hat{v} = v_{\max}$ is the single candidate for optimality, whereas in the case with $c = 0.5$, the equilibrium with $\hat{v} = 0$ is the only one for which the backward calculation has to be done. In both cases there exist indifference curves. Their detailed exposition is omitted for the sake of focusing on the more global differences in optimal control when different values for the control cost c are assumed.

For the cost parameter $c = 0.28$, the above analysis of Lagrange Multipliers reveals that both steady states, $\hat{E}_{\hat{v}=v_{\max}}$ and $\hat{E}_{\hat{v}=0}$, are candidates for the optimal long-run solution. In such cases, DNSS curves can occur. They separate the basins of attraction of the steady states at points at which a decision maker is indifferent which equilibrium to approach, because any of the alternatives causes the same cost. Such a DNSS curve was detected in the scenario analyzed in section 6.1. In the current case such a curve does not occur. Instead, we have to deal with dominated steady states. Staying at the steady state $\hat{E}_{v=0}$ located at $(\hat{A}, \hat{S}) = (0.304916, 0.188672)$ in the (A, S)-plane forever results in the following value of the objective functional where we make use of equation (4.25):

$$J = \frac{1}{0.04} H(0.3049, 0.1887, 0, -8.9237, -5.1391) = -7.623.$$

Emanating from $(A(0), S(0)) = (0.304916, 0.188672)$, the steady state $E_{\hat{v}=v_{\max}}$ located at $(\hat{A}, \hat{S}) = (0.333455, 0.154617)$ can be reached. The corresponding policy assigns the interior harm reduction value $v(0) = 0.2496$ for initial control. Hence, the objective functional's value for the resulting trajectory evaluates to

$$J^* = \frac{1}{0.04} H(0.304, 0.188, 0.2496, -4.266, -2.538) = -7.6001.$$

Our objective is to maximize the negative aggregate cost value, hence we conclude that the trajectory that yields J^* is preferred. Despite the fact that harm reduction interventions cause costs in this scenario, it bears less cost to approach the equilibrium with full harm reduction where social cost is cut down to $1 - v_{\max}$ of its baseline value, than to stay at the pure use reduction steady state forever and without such a reduction in social costs. In the basin of attraction of the dominating steady state $\hat{E}_{\hat{v}=v_{\max}}$ an indifference curve is found, but the detailed presentation is omitted.

Optimal Control

Figure 6.9 provides an overview over optimal control on the (A, S)-plane for the different control cost values. Like in the kindred Figures 4.1 and
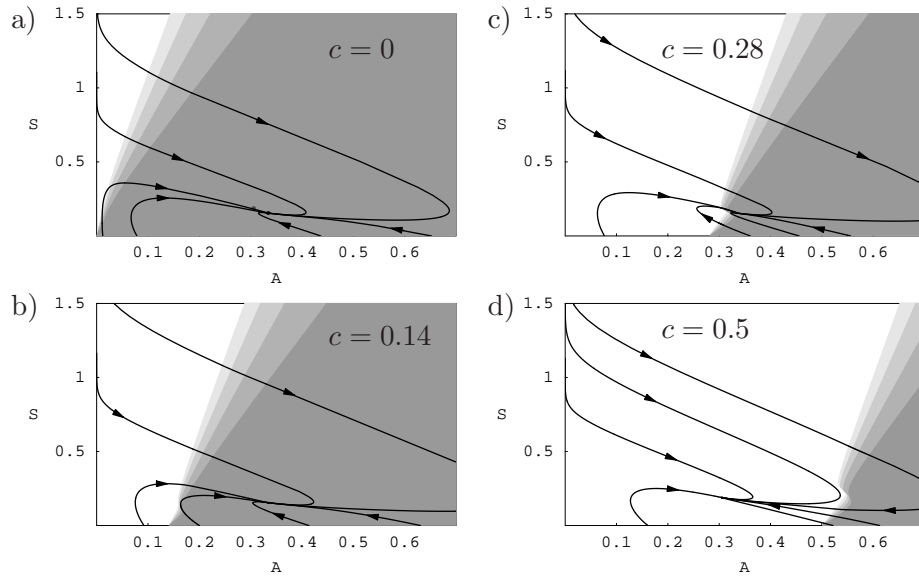
Figure 6.9: Comparison of optimal control on the (A, S)-plane for different harm reduction control cost values c in Australia

4.6, the shading indicates how much harm reduction is optimal. At white parts of the (A, S)-plane, optimal control is given by $v^* = 0$. To the other extreme, optimal control $v^* = v_{\max}$ is indicated by the darkest gray color. The intermediate gray levels indicate where harm reduction interventions are implemented to some interior extent, which is either $0 < v < \frac{1}{3} \cdot v_{\max}$, $\frac{1}{3} \cdot v_{\max} < v < \frac{2}{3} \cdot v_{\max}$, or $\frac{2}{3} \cdot v_{\max} < v < v_{\max}$. To insinuate the flow of the optimal system subject to the different values of c, any of the panels shows some trajectories. The little arrows indicate the direction of convergence towards the optimal long-run steady state.

Panel a) in Figure 6.9 redraws the optimal control structure from the base case, which is equivalent to the case $c = 0$ in the current formulation. The other panels show optimal control on the (A, S)-plane for the cases with $c > 0$.

The result for $c = 0.14$ is depicted in panel b). What can immediately be seen is that optimal control is structurally very similar to the base case. For a large region around the optimal long-run equilibrium, full harm reduction is the best choice. When the number of users is rather modest, sticking to a pure use reduction strategy is the optimal policy option. At intermediate stages

of the epidemic, we encounter transitory control. It is either declining (when harm reduction triggers negative effects in initiation) or increasing (when the reduction in social costs in the objective function that is attributable to harm reduction is attractive). The triangular form of the region with optimal interior control values $0 < v_i^* < v_{\max}$ is broken up. With respect to the optimal trajectories, this leads to the existence of an indifference curve. But much more interesting than this particular hairline case curve is a more general observation: The regions with interior amounts of harm reduction $0 < v_i^* < v_{\max}$ and with full harm reduction $v^* = v_{\max}$ are more or less only shifted to the right along the A-axis by an increment of 0.14.

The optimal control structure for the system with cost $c = 0.28$ is presented in panel c). This is the case where among two candidates for optimality, $\hat{E}_{\hat{v}=v_{\max}}$ is the dominant steady state. The arrangement of the differently shaded regions is again shifted further to the right. Compared to the base case, the increment for the right-shift on the A-axis is now 0.28. The fundamental division into the differently colored regions is like in the previous cases, even though the crack in the triangular shape of the regions with optimal interior control $0 < v_i^* < v_{\max}$ is more pronounced now.

Finally, the case of a single candidate for optimality that exhibits $\hat{v} = 0$ was investigated. Panel d) presents optimal control as a function of A and S for this case with control cost $c = 0.5$. At the optimal long-run steady state, harm reduction is omitted, because control cost is too high there. Furthermore, the no harm reduction policy is the optimal policy choice for most initial conditions shown on the panels of Figure 6.9, which range from 0 to $700,000$ users ($A = 0.7$ on the plot) for $A(0)$ and from $S(0) = 0$ to $S(0) = 1.5$ million for the susceptibles. Nevertheless, when the epidemic grows beyond a certain threshold of users, interior amounts of harm reduction and full force application of the control tool become the optimal strategies, because they ameliorate the harms felt by the considerable number of users.

The similarity of the structure of the regions with boundary and non-boundary control in the different cases is striking. The light gray sliver for the control range $0 < v < \frac{1}{3} \cdot v_{\max}$ is always the narrowest one, whereas the sliver with $\frac{2}{3} \cdot v_{\max} < v < v_{\max}$ is always the broadest one. Most striking, from one case to the other, they do not seem to change in broadness.

Comparison of Costs

In order to compare the aggregate costs for each of the values of c when emanating from a fixed initial condition, $(A(0), S(0)) = (0.3, 1)$ was chosen.

The initial value for A is quite close to, but below the steady state numbers of users $\hat{A}_{v=0} = 0.304916$ and $\hat{A}_{v=v_{\max}} = 0.333455$. The initial value for the S-state is a rather arbitrary number. Primarily, its choice is based on the idea that at the beginning of the epidemic, there should be a considerable number of susceptibles from which initiates can be recruited.

In the system with $c = 0$, optimal control at the initial condition is full harm reduction, i.e. we have $v^*(0) = v_{\max}$. The steady state $\hat{E}_{\hat{v}=v_{\max}}$ is reached with upper boundary control along the entire planning horizon. At the initial state $A(0) = 0.3$, $S(0) = 1$, the optimal trajectory assigns the costates $\lambda_A(0) = -4.3929$ and $\lambda_S(0) = -2.7504$. Making use of formula (4.25) yields

$$ J^*|_{c=0} = \frac{1}{r} H(0.3, 1.0, 0.53, -4.3929, -2.7504) = -6.0368. $$

The optimal trajectory in the system with $c = 0.14$ is emanating from the initial condition $(A(0), S(0), \lambda_A(0), \lambda_S(0)) = (0.3, 1.0, -4.4018, -2.7487)$. The optimal initial control value is the interior amount $v^*(0) = 0.4118$ of harm reduction. From there, control ramps up to the maximum, and then the system converges to the steady state with full harm reduction. The resulting value of the objective functional is

$$ J^*|_{c=0.14} = \frac{1}{r} H(0.3, 1.0, 0.4118, -4.4018, -2.7487) = -7.8911. $$

In the system with $c = 0.28$, the initial states $(A(0), S(0)) = (0.3, 1.0)$ yield optimal costate values $(\lambda_A(0), \lambda_S(0)) = (-4.7577, -2.7285)$. Optimal initial control is given by $v^*(0) = 0$. Pure use reduction is optimal for some initial time span, but then more and more harm reduction is done. Finally, it is applied with full force until the steady state is reached. The value of the objective functional for the resulting trajectory is given by

$$ J^*|_{c=0.28} = \frac{1}{r} H(0.3, 1.0, 0, -4.7577, -2.7285) = -9.7242. $$

When the harm reduction cost parameter c is high, i.e. $c = 0.5$, the optimal costates for $(A(0), S(0)) = (0.3, 1.0)$ are $\lambda_A(0) = -6.1137$, $\lambda_S(0) = -3.3296$. Optimal control at the initial point is $v^*(0) = 0$, then use increases to levels where implementation of harm reduction measures is beneficial to the system. Harm reduction is introduced, then gradually increases until its maximum possible value $v^* = v_{\max}$ is reached, and under this full harm reduction policy, the turn from increasing to decreasing numbers of users is

done. When numbers of users are sufficiently low, harm reduction is gradually reduced and finally abolished. For the remaining time until the steady state is reached, the optimal policy is pure use reduction. The value of the objective functional along this trajectory with a rather sophisticated blend of boundary and transitory control values is

$$J^*|_{c=0.5} = \frac{1}{r} H(0.3, 1.0, 0, -6.1137, -3.3296) = -11.3457.$$

Figure 6.10 shows the different values of control cost c investigated for Australian IDU and the associated value of the objective functional along the optimal trajectory that emanates from the initial condition at $A(0) = 0.3$ million IDUs and $S(0) = 1$ million susceptibles. The discounted, aggregate social cost J^* along the optimal path leading to the optimal long-run steady state is higher, when c is large. This fact is not astonishing. In Figure 6.9, we can observe that optimal control on the (A, S)-plane is in principle always governed by the idea: "Do harm reduction when use is high, because then it reduces social cost via the objective function. Do not apply harm reduction when use is low, because then it adversely affects the system dynamics due to risk compensation effects." Solely, the distinction whether use is "high" or "low" enough depends on the value of c. The objective function now puts social cost $A(1-v)$ into relation to the control cost term cv. This demands for a trade-off. The number of users for which the reduction in the objective function becomes attractive despite the adverse effect modeled by $g(v)$ is higher when c is high. For low levels of use and c being large, the benefit of the reduction in the term $A(1-v)$ is wiped out by the control cost value cv. The new tension between cost for the control mechanism and the fact that the control can reduce social costs, leads to the shift in the optimal control regions as identified in Figure 6.9.

Thoughts about Realistic Values of c

The above findings are of somewhat theoretic nature, because the social cost parameter normalized to $\kappa = 1$ (see section 2.5.1) and the control cost parameter c are fairly abstract numbers. In what follows we try to derive further insights on a realistic magnitude of the parameter c.

Following Moore (2005), the Australian governments spent \$ 3.2 billion on drug control in the year 2002-03. Australia's drug budget is divided into "proactive" policies, which account for \$ 1.3 billion, whereas \$ 1.9 billion are spent "reactively". The latter term means that interventions deal with the consequences of drug use. A detailed list of expenditure components reveals
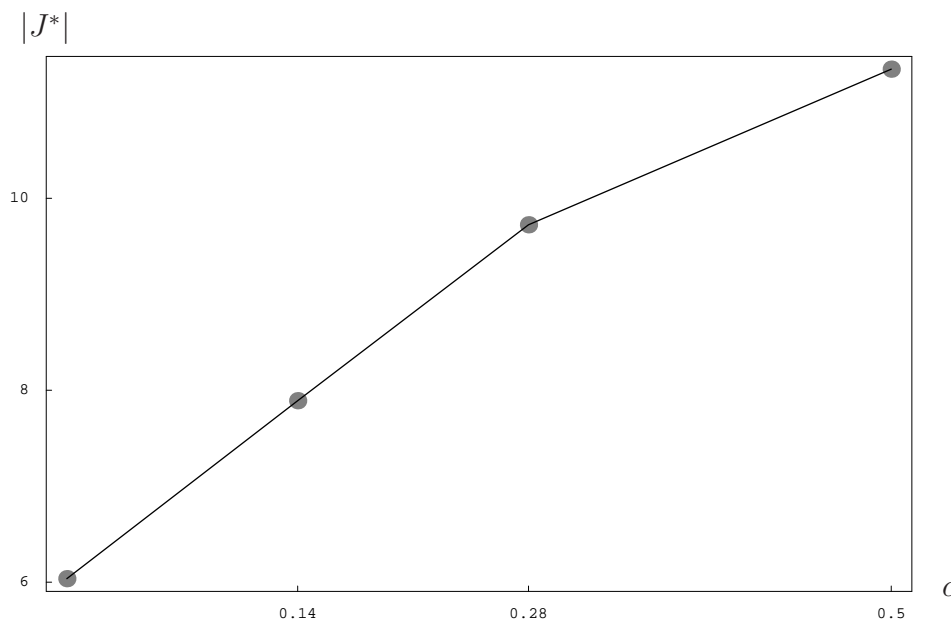
Figure 6.10: $|J^*|$ evaluated for the optimal trajectories emanating from $(A(0), S(0)) = (0.3, 1)$ for different values of control cost c in Australia

that needle and syringe programs received \$ 4.6 million of federal money, and \$ 33.7 million were funded by state or territory, respectively. Hepatitis C education and family support received funds accounting for \$ 6.5 million out of the federal budget. Summing up those items, Moore (2005) concludes that \$ 44.8 million were spent in the year 2002-03 on harm reduction measures. Hence, for the following calculations we denote $B = 44.8 \cdot 10^6$.

Harm reduction is one of the important pillars of Australia's strategies to control drug consumption and problems stemming from it. We assume that over one year an average level v_c of harm reduction measures is applied. For this level, we explore two extreme cases: an optimistic assumption of full harm reduction $v_c = 0.53$ and the pessimistic assumption that the spending of the \$ 44.8 million resulted in a reduction of only 5%, i.e. $v_c = 0.05$. This yields an optimistic and a pessimistic value for the parameter c.

There is a little subtlety of the objective function we have to consider when deriving those values. The social cost stemming from drug use is quantified by the constant $\kappa = 1$. It is a normalized, abstract number that was used for simplification in the base case model, where it did not matter if the outcome was an abstract number or dollar cost. For the current investigation κ and c have to be synchronized.

Remember first that κ gives the social cost per unit use. In the model, the unit $A = 1$ means that there is one million users. Moreover, the proxy for the harms felt by the users were health-related social costs of drug abuse listed in COI studies. In Australia, the social cost parameter $c_s = 39,255$ \$ per user per year can be directly used. Taking into account the unit of $A = 1$ million of users, we arrive at the non-normalized parameter $\tilde{\kappa} = 39.255 \cdot 10^9$.

Next, we derive $\tilde{c}$ for the non-normalized dollar cost of harm reduction interventions. Due to the assumption that v_c is the average amount of harm reduction applied over one year, we do not take into account discounting of costs and simply equate $B = \tilde{c} \cdot v_c$. The pessimistic assumption $v_c = 0.05$ results in the pessimistic estimate $\tilde{c}_p = \frac{B}{v_c} = \frac{44.8 \cdot 10^6}{0.05} = 8.96 \cdot 10^8$. The optimistic assumption $v_c = 0.53$ evaluates to the optimistic estimate $\tilde{c}_o = \frac{B}{v_c} = \frac{44.8 \cdot 10^6}{0.53} = 8.45 \cdot 10^7$. Normalizing those values with the help of $\tilde{\kappa}$, we arrive at $c_p = \frac{\tilde{c}_p}{\tilde{\kappa}} = \frac{8.96 \cdot 10^8}{39.255 \cdot 10^9} = 0.0228$ and $c_o = \frac{\tilde{c}_o}{\tilde{\kappa}} = \frac{8.45 \cdot 10^7}{39.255 \cdot 10^9} = 0.0022$.

Both the optimistic estimate $c_o = 0.0022$ and the pessimistic $c_p = 0.0228$ are far away from the critical parameter $c = 0.265058$, where the feature of the full harm reduction steady state being the unique candidate for optimality is endangered. In panels a) and b) in Figure 6.9, optimal control in the cases $c = 0$ and $c = 0.14$ is displayed. We infer that even in the pessimistic scenario of $c_p = 0.0228$, the shift of the regions to the right is significantly less than that displayed in panel b).

Taking into account that in Australia there exist good harm reduction interventions and that the country has a long history of such approaches, the real value v_c will be rather large, most probably around the parameter value $v_{\max}$. Hence, we assume that the real parameter c is relatively close to zero. Consequently, the simple analysis conducted in this section underpins the assumption that control cost for providing harm reduction measures to Australia's IDUs can be neglected in the objective function, as we did in the base case.

6.3.2 Cost Parameter Scenarios for the United States

The different cost values investigated for the U.S. cocaine epidemic are $c = 2$, $c = 4$, $c = 8$, and the special case $c = 0$ that has been assessed in the base case. The results for optimal control on the (A, S)-plane for those cases are shown in Figure 6.11. Before analyzing the differences, we give a short overview over the optimal long-run steady states in the different cost scenarios.

Steady States

Substituting cost $c = 2$ into the model, the Lagrange Multiplier π_2 for the high-use steady state with full harm reduction is positive, whereas π_1 for the high-use steady state without harm reduction is negative. In the basin of attraction of the optimal high-use steady state, there is an indifference curve. At intermediate levels of use, the steady state with $\hat{v} = 0$ evaluates $\pi_1 < 0$, hence it is not a candidate for optimality. The upper boundary control steady state at intermediate numbers of users evaluates a positive Lagrange Multiplier $\pi_2 > 0$. As in the base case (see section 4.5.2), it can be shown that this steady state is dominated. Instead of applying full harm reduction $v = v_{\max}$ to remain at the steady state forever, the choice of $v^*(0) = 0$ is optimal, which leads to a path that finally approaches a no-use steady state. A new feature compared to the base case is that at and around this no-use steady state, optimal control is given by the pure use reduction policy.

That latter feature persists in the model with the higher cost parameter $c = 4$. At the intermediate-use steady states, full harm reduction leads now to a negative Lagrange Multiplier $\pi_2 < 0$, while the lower boundary control steady state has a positive Lagrange Multiplier $\pi_1 > 0$. Nevertheless, the costates are positive, whereas we expect negative costates. The backwards calculation from the intermediate-use steady state is possible, it has a two-dimensional stable manifold in the four-dimensional system. The projection of the stable paths onto the (A, S)-plane coincides with the stable manifold of the intermediate-use steady state in the two-dimensional system ($\dot{A} = 0, \dot{S} = 0$). The high-use steady states evaluate positive corresponding Lagrange Multipliers. The steady state without harm reduction is dominant. In its basin of attraction, an indifference curve is found, but the details are omitted here.

The last cost parameter is $c = 8$. The situation for the intermediate-use and no-use steady states is unchanged compared to the previous case of $c = 4$. Among the high-use steady states, the full harm reduction steady state is no longer a candidate for optimality, because we find $\pi_2 < 0$. Hence, the optimal steady state at high levels of use exhibits a pure use reduction strategy.

Optimal Control

Optimal control as a function of the state space (A, S) is shown in Figure 6.11, using the usual gray shadings for different levels of optimal control. Please note that the location and relative size of the basins of attraction of
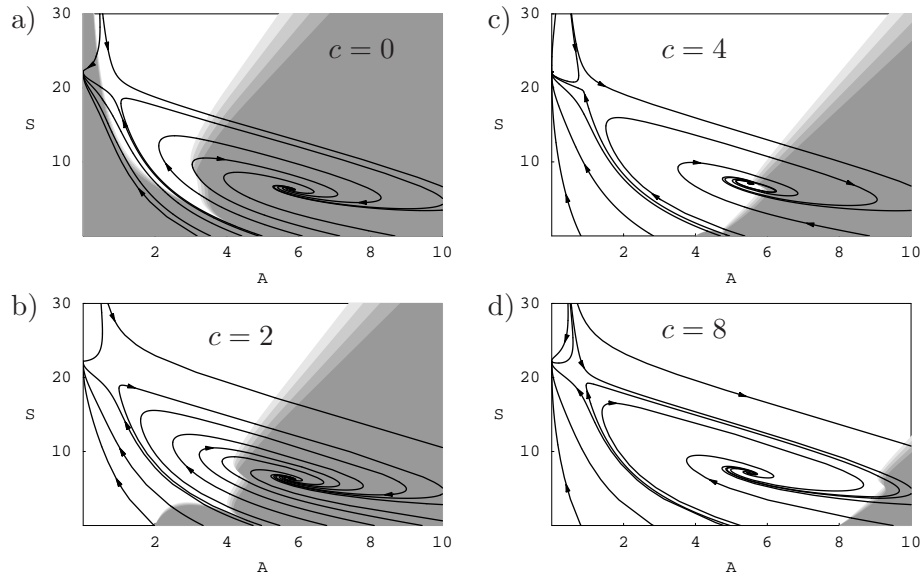
Figure 6.11: Comparison of optimal control on the (A, S)-plane for different harm reduction control cost values c in the U.S.

the two optimal steady states do not change from one panel to the other. Notwithstanding, the optimal control structure does.

Panel a) redraws optimal control for the base case $c = 0$. Harm reduction is optimal when use is widespread or when the number of users is only modest. Although $v^* = v_{\max}$ for most of the state space, a considerable part of the (A, S)-plane is white, which stands for optimal control $v^* = 0$. The boundary control regions are separated by very narrow regions of transitory harm reduction.

Switching to panel b), which shows optimal control in the case $c = 2$, there are significant changes in both basins of attraction. The first striking difference is that now optimal control at and around the no-use steady state is $v^* = 0$. Optimal control $v^* = 0$, which characterizes the white region, makes up the lion's share of the basin of attraction of the no-use steady state. When converging to the no-use steady state, interior and full harm reduction are optimal only in a tiny region with about $A = 2$ to $A = 4$ million users combined with modest numbers of susceptibles. The formerly expanded region of $v^* > 0$ is now shifted and compressed dramatically. Second, the regions where intermediate levels of harm reduction are optimal in the basin

of attraction of the high-use steady state are shifted to the right, and there is a certain distortion in the structure of those regions.

Going one step further to panel c) and the cost value $c = 4$, in the basin of attraction of the no-use steady state the regions with optimal control $v^* > 0$ are even more contracted. The pure-use reduction regime is now optimal in almost the entire basin of attraction. At the high-use steady state, optimal control is now $v^* = 0$. Compared to the base case, the white region where $v^* = 0$ has expanded considerably to the right. In the cases with lower c, the gray regions had a rather round shape with a dint on the left side of the optimal high-use steady state. Here, the optimal high-use steady state is located in the white region. The gray regions are shifted again further to the right, which results in an almost triangular form for the regions of interior and full harm reduction. There is again a dint in the gray slivers, but now it occurs on the right of the optimal high-use steady state.

When control cost is as high as $c = 8$, optimal control in the entire basin of attraction of the no-use steady state is given by $v^* = 0$. In the basin of attraction of the high-use steady state, the regions shaded in gray are shifted far to the right. This is shown in panel d) of Figure 6.11.

The value of c obviously determines how far the regions with an optimal, non-zero amount $v^* > 0$ of harm reduction are shifted to the right. On the A-axis, the switching point between $v^* = 0$ and $v^* > 0$ is located at $(A, S) = (c, 0)$. This happens because at $S = 0$, the Lagrange Multiplier π_1 from equation (6.4) is positive as long as $A < c$ holds, and the Lagrange Multiplier π_2 from equation (6.6) is positive there for $A > c$.

In the base case, where control costs are neglected, the policy recommendation of our model is that full harm reduction can be applied safely for low levels of use. One accepts the downside of increases in initiation, because the feedback effect from current users to susceptible non-users does not trigger so many new users such that use could explode. In the cases with control cost, harm reduction is omitted at most of those low levels. This switch in the optimal control is not because harm reduction interventions are "so bad" all of a sudden. The mechanism of initiation, essentially the effect of $g(v)$, is entirely the same as in the base case. The crucial fact is that now control interventions cost money. It is a rather simple effect, which is easy to understand from an economic point of view. In the trade-off between reduction of harm felt by users and monetary cost for the tool that helps to achieve those reductions, control is omitted. Only when prevalence grows to higher numbers of users, the tension is resolved in favor of harm reduction's application.
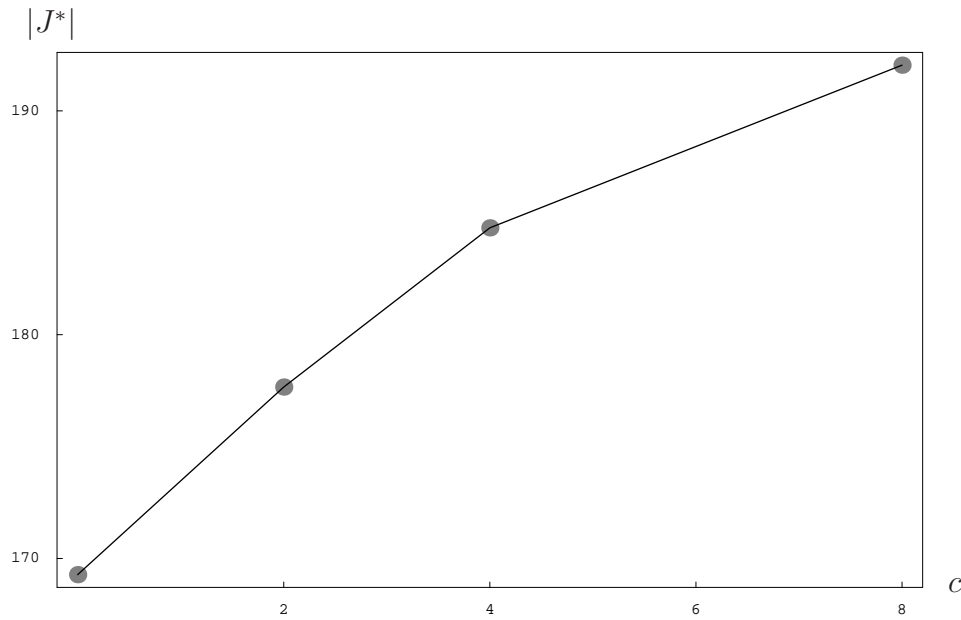
Figure 6.12: $|J^*|$ evaluated for the optimal trajectories emanating from $(A(0), S(0)) = (4, 21)$ for different values of control cost c in the U.S.

Comparison of Costs

Figure 6.12 shows the different values of control cost c investigated for the SA Harm Reduction Model parameterized for the U.S. cocaine epidemic with the corresponding values of the objective functional resulting from the optimal trajectory emanating from the initial condition at $A(0) = 4$ million cocaine users and $S(0) = 21$ million individuals being vulnerable to try cocaine. The corresponding initial costates, initial control, and optimal control over time are described in the paragraphs below. The aggregate discounted social cost $|J^*|$ along the optimal path is higher, the bigger the parameter value c for control cost is. In any of the cost scenarios, the optimal control path emanating from the chosen initial condition involves periods where interior levels of harm reduction or full force harm reduction are optimal. At those times, control is applied to exploit the benefits of the reduction in social cost. But control cost is running, and when c is higher, the accumulated control costs are higher. Furthermore, the region where no harm reduction is done, is larger when c is high. This means that the periods where there is no reduction in social cost for considerable time spans. Putting the two effects together gives an explanation why $|J^*|$ increases when c grows.

In the base case system with $c = 0$, optimal control at the initial condition suggests to do no harm reduction at all, i.e. $v^*(0) = 0$. Converging to the optimal steady state $\hat{E}_{\hat{v}=v_{\max}}$, control starts to grow later on. Once the upper bound of the feasible control values is reached, upper boundary control remains optimal while the steady state is approached. At the initial state $A(0) = 4$, $S(0) = 21$, the optimal trajectory assigns the costates $\lambda_A(0) = -9.6490$ and $\lambda_S(0) = -4.2481$. Making use of formula (4.25) yields

$$J^*|_{c=0} = \frac{1}{r} H(4.0, 21, 0, -9.6490, -4.2481) = -169.2768.$$

In the system with the cost parameter $c = 2$, the optimal trajectory emanates from $(A(0), S(0), \lambda_A(0), \lambda_S(0)) = (4.0, 21, -9.9784, -4.2400)$. The optimal initial control value is again $v^*(0) = 0$. After a period of pure use reduction, control ramps up to the maximum, and then the system converges to the steady state with full harm reduction. The resulting value of the objective functional is

$$J^*|_{c=2} = \frac{1}{r} H(4.0, 21, 0, -9.9784, -4.2400) = -177.6599.$$

The optimal trajectory in the system with $c = 4$ assigns the optimal costate values $(\lambda_A(0), \lambda_S(0)) = (-10.313, -4.26316)$ to the initial condition. Optimal initial control is again given by $v^*(0) = 0$. The optimal control strategy that leads to the long-run steady state is very complicated in this case. First, control is at the lower bound, then it increases, is at the upper bound for some time. Then it decreases again, a time span of pure use reduction follows, but then control increases again. There is another time interval for which full harm reduction is optimal, then control declines, and finally v^* remains at the lower bound until the steady state is reached. The value of the objective functional for the resulting trajectory is given by

$$J^*|_{c=4} = \frac{1}{r} H(4.0, 21.0, 0, -10.313, -4.26316) = -184.773.$$

For the high cost parameter $c = 8$, the optimal costates at the initial condition $(A(0), S(0)) = (4.0, 21)$ are $(\lambda_A(0), \lambda_S(0)) = (-10.8380, -4.4097)$. Like in the preceding cases, optimal control at the initial point is given by $v^*(0) = 0$. Optimal control over time stays at zero harm reduction first, then it increases, followed by some time interval of full harm reduction, until it declines back to the pure use reduction regime. The value of the objective functional along the resulting trajectory is

$$J^*|_{c=8} = \frac{1}{r} H(4.0, 21, 0, -10.8380, -4.4097) = -192.0424.$$

Thoughts about Plausible Values of c

Similarly to the derivation of realistic values for the control cost parameter c in the case of Australian IDU, we conduct the following rough calculations. Caulkins et al. (2002) report social costs of \$ 49 billion attributable to cocaine use in the U.S. in 2000, which implies $\tilde{\kappa} A = 49 \cdot 10^9$. In this case, data on harm reduction funding is not readily available. Hence, we depart from federal drug control spending for treatment and prevention. The ONDCP (2002) lists the federal drug control spending in the year 2001 for treatment and prevention with 3,335 and 2,578.7 million dollars, respectively, which sums up to approximately 6 billion dollars. Without a special reasoning, we assume that 10% of those interventions can be interpreted as harm reduction tactics, i.e. we arrive at $B = 6 \cdot 10^8$. The most recent World Drug Report (UNODC, 2008) lists current prevalence of cocaine use in the U.S. at the level of 7,097,000 cocaine users. With this, we arrive at $\tilde{\kappa} \approx \frac{49 \cdot 10^9}{7} = 7 \cdot 10^9$.

We assume a very pessimistic scenario of $v_c = 0.02$, which presumes that only 2% percent of the harms counted in the objective function are wiped out due to harm reduction measures. Remember that in the best case, when all harms felt by the users are cut down, this accounts for the share of $v_{\max} = 0.17$. Like in the case of Australia, we use the simple equation $\tilde{c} \cdot v_c = B$. This yields $\tilde{c} = \frac{B}{v_c} = \frac{6 \cdot 10^8}{0.02} = 3 \cdot 10^{10}$. Normalizing by $\tilde{\kappa}$, we arrive at $c = \frac{\tilde{c}}{\tilde{\kappa}} = \frac{3 \cdot 10^{10}}{7 \cdot 10^9} = 4.28571$. This value is larger than the critical value $c = 4.0649$, for which high-use steady state with upper boundary control is no longer a candidate for the optimal solution.

The assumptions we took to arrive at this parameter value bear much uncertainty. It seems plausible that a realistic value for c might be even higher. Obviously, the value of c provided by the present rough calculations depends very much on the assumption how much of prevention and treatment measures in the U.S. can be seen as harm reduction targeted on cocaine users.

6.4 Logistic Approach for Initiation into IDU in Australia

In section 2.4 we touched on the subject that power functions of the form A^α are not the only possibility to make initiation a function of the number of current users. Assuming that adverse consequences of drug use manifest when there is a certain (rather high) number of users and that the awareness of the negative consequences of drug use acts as a brake on initiation, a

logistic function is appropriate. Equation (2.4) introduces the functional form

$$I_{\log}(A, S, v) = b\, A(\bar{A} - A)\, S\, g(v) \tag{6.8}$$

for the logistic initiation function. The parameter $\bar{A}$ denotes the carrying capacity of the model. As long as $A < \frac{\bar{A}}{2}$, the term $A(\bar{A} - A)$ is increasing. At $A = \frac{\bar{A}}{2}$, the contagion effect peaks. For levels of use higher than half of the carrying capacity, initiation is decreasing in the number of users. Please note that when the number of users grows beyond the carrying capacity $\bar{A}$, initiation is negative, which is not sensible. Hence, the logistic approach is limited to the domain $A \leq \bar{A}$.

A general logistic function is displayed in Figure 2.2. It is a concave function. Hence, the logistic approach suits modeling initiation into injecting drug use in Australia, for which we assumed that the contagion effect is concave in the number of users. The U.S. cocaine epidemic's parameterization involves an exponent $\alpha > 1$ in the approach with the power function. The contagious effect is stronger, the more users exist. The logistic approach does not meet this property. Consequently, the logistic function is explored for Australian IDU, but not for cocaine use in the U.S.

The most basic approach to derive the parameter values for the new logistic approach for Australia is to evaluate the original term $b\, A^\alpha$ for the base case parameterization and for values of A between 0.01 and 0.4 in steps of 0.01. The resulting values are then used to estimate a least squares fit for the logistic form. This results in a carrying capacity of $\bar{A} = 2.005775$ and a proportionality constant $b = 0.35656$. The power function $b\, A^\alpha = 0.5112\, A^{0.8622}$ from the base case and the new logistic term $b\, A(\bar{A} - A) = 0.35656\, A\, (2.005775 - A)$ are shown in Figure 6.13. The curves are very close to each other as long as there are less than half a million of IDUs. The gray curve which stems from the functional form of the base case denoted in equation (2.2) then increases stronger, whereas the logistic function shown as a black dashed curve has a less steep incline. It peaks at $A = \frac{\bar{A}}{2} = 1.00289$. Figure 6.13 shows that the logistic function decreases for numbers of users that exceed that critical stage $A = \frac{\bar{A}}{2}$.

For the sake of conciseness, the canonical system stemming from application of Pontryagin's Maximum Principle is not presented. Conducting the usual steps (search for steady states with interior control, analysis of boundary control steady states and the corresponding Lagrange Multiplier, investigation of stability properties) we arrive at a unique optimal solution.
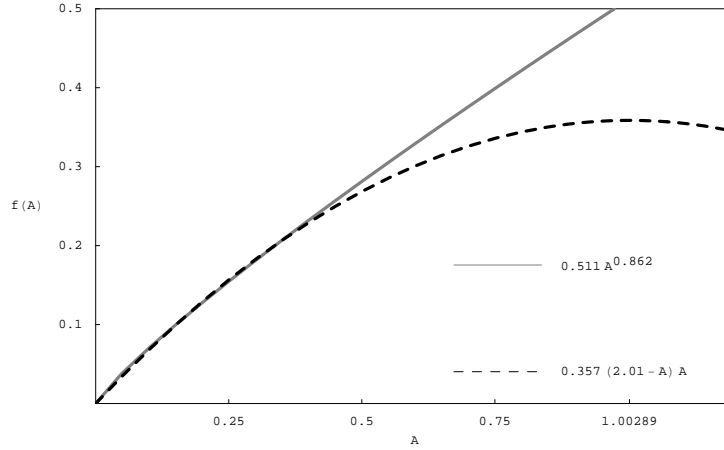
Figure 6.13: The power function $b\,A^\alpha = 0.5112\,A^{0.8622}$ from the original initiation function for Australian IDU is shown in gray. The black dashed curve depicts the new logistic approach $b\,A(\bar{A} - A) = 0.35656\,A\,(2.005775 - A)$.

It is a steady state with boundary control $\hat{v} = v_{\max}$. It is located at

$$
\begin{pmatrix} \hat{A} \\ \hat{S} \\ \hat{\lambda}_A \\ \hat{\lambda}_S \end{pmatrix} = \begin{pmatrix} 0.333769 \\ 0.154242 \\ -3.8735 \\ -2.49905 \end{pmatrix}.
$$

The fixed point evaluates a positive Lagrange Multiplier $\pi_2 = 0.305758$. The Jacobian Matrix of the new canonical system evaluated at the equilibrium exhibits Eigenvalues

$$
\begin{aligned}
\xi_1 &= 0.276486, \\
\xi_2 &= -0.236486, \\
\xi_3 &= 0.167214, \\
\xi_4 &= -0.127214.
\end{aligned}
$$

The Eigenvalues are real, with one pair of them being negative whereas the other pair is positive. Hence, the property of saddle point stability is given.

Optimal control as a function of users A and susceptibles S is shown in Figure 6.14. The shading in gray and white is analogous to kindred gray level Figures in this thesis, e.g. Figure 4.1 or Figure 4.6. Some of the trajectories from the optimized phase portrait are shown. The arrows indicate
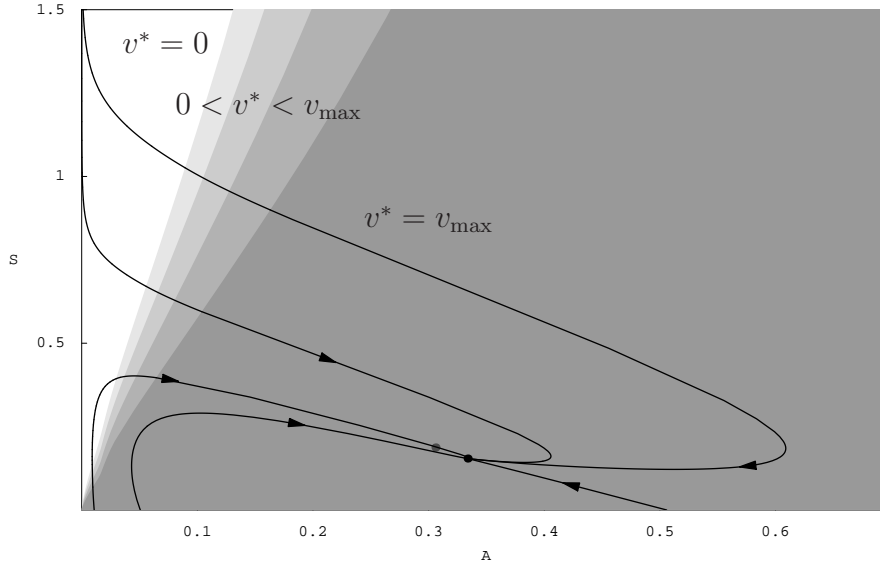
Figure 6.14: Optimal control on the (A, S)-plane for the logistic approach modeling initiation into IDU in Australia.

the direction of convergence towards the steady state. The unique optimal long-run steady state is shown as a black dot. Any trajectory that starts in the feasible domain (remember that here it is limited to $A \leq \bar{A} = 2.005775$) eventually approaches this fixed point. The gray dot close to it shows the steady state of the boundary control system with $\hat{v} = 0$, which evaluated a negative Lagrange Multiplier $\pi_1 = -0.266979$ and is therefore not a candidate for optimality.

It is evident from Figure 6.14 that both the optimal control structure and the resulting optimized phase portrait are very similar to the results derived for the base case. Those results are shown in Figure 4.1. Hence, we conclude with the finding that the results achieved for Australia's population of IDUs are not induced by the special functional form A^α. When switching to another concave function to model the contagion effect of current users to non-using susceptibles, the results are stable.

Chapter 7

Conclusions and Possible Extensions

This thesis was devoted to a two-state one-control model of drug epidemics. The main purpose was to derive optimal drug control strategies. There were two innovative aspects of the model. First, it explicitly considered a population of non-using susceptibles who are the ones to feed the drug use state. Second, a fairly novel control mechanism was modeled, which aims at the reduction of the harms felt by drug users. The parallel investigation of harm reduction for one parameter set based on data on the U.S. cocaine epidemic and another parameter set derived from data on injecting drug use in Australia showed that appropriate functional forms and the corresponding parameters depend on the specific drug and country.

Unlike classic drug control interventions like law enforcement acting like a tax driving up prices, treatment stimulating the outflow from drug use, or prevention that aims at avoiding that young people ever try drugs, the new approach does not primarily aim at reductions in use and may even increase initiation. This makes harm reduction controversial. The model introduced in this thesis tried to provide advices to the participants in the discussion whether use reduction or harm reduction is the right policy.

From the modeling perspective, boundary conditions for the control had to be considered. In a first step a static-control analysis of the performance of full harm reduction relative to the performance of pure use reduction was conducted. The most important conclusion from this static comparative analysis was that harm reduction seems to have great potentials to ameliorate social costs and harms associated with injecting drug use in Australia. For the base case scenario of the U.S. cocaine epidemic, it seems to have great

potential at the early stages of the epidemic - when drug use is rare and there is no danger that initiation will explode - or when the number of drug users is high at the later stages of the epidemic, because then the reduction in social cost counted in the objective function is beneficial. Nevertheless, the investigations of the static system revealed that if harm reduction was timed incorrectly, it may tip the epidemic and/or induce increased aggregated harm.

Applying optimal control theory to the base case model which neglects control costs and assumes that initiation into drug use is only driven by feedback effects from current use, it turned out that for the Australian parameterization of the model, harm reduction tactics were optimal for almost all possible initial conditions, but if numbers of users are very small, it was better to omit harm reduction. For the U.S. cocaine epidemic parameterization, the policy recommendation was less straightforward. Multiple steady states occurred, of which a high-use steady state and a no-use steady state were the optimal long-run solutions. For both of them, full harm reduction was optimal at the steady state and in a large region around the fixed point. An indifference curve was found for which the high-use steady state could be reached in two different ways yielding the same value of the objective function. With respect to optimal control the simple model suggested to omit harm reduction at intermediate levels of cocaine use, but for high levels of use, and when use was rather low, the optimal strategy was full application of harm reduction.

A sensitivity and bifurcation analysis was conducted to deal with the problem that there is of course some uncertainty about the base case parameterizations. We found blue sky bifurcations, neutral saddles, and Hopf bifurcation points. The Hopf bifurcation points indicate the existence of limit cycles, but given the many other interesting results in this thesis there were no further investigations in that direction. Of course, these Hopf bifurcation points provide an important starting point for future investigations of the SA Harm Reduction Model, in which we may expect to find optimal limit cycles.

The previous Chapter assessed several variations of the base case model. First, the case of decreased virulence of the cocaine epidemic in the U.S. was investigated. The structure of optimal control depending on the stage of the epidemic was not so different from the base case results, but nevertheless optimality of the multiple candidate solutions changed relative to the base case. Given the new parameter value used in this variation, eradication of the epidemic was optimal for most initial conditions, whereas only for a narrow region of initial conditions a high-use steady should be approached. The

basins of attraction of those two optimal steady states were separated by a DNSS curve.

Second, innovators were included in the initiation function. At high levels of use there was barely a change relative to the base case scenario of the U.S. cocaine epidemic, but at low levels of use, the steady state was pushed away from zero use towards a moderate number of users. Furthermore, this steady state did not exhibit optimal control at its boundary level, i.e., control took an intermediate value there. Given the parameter value modeling the group of innovators in this variation, we found that the full harm reduction strategy was no longer optimal at low levels of drug use. The innovators give a drive to initiation that is not governed by the size of current use. Even at low levels of use, harm reduction would attract new innovators which triggers undesired increases in use. On the other hand, innovators into drug use only marginally changed the results for Australian IDU.

Third, a linear cost function was introduced. The important conclusion from the investigation of the model using different cost parameter values was that harm reduction cost can indeed be neglected in the objective function when modeling Australian IDU. For the U.S., additional investigations on budgets and harm reduction interventions seem necessary to arrive at a recommendation whether control costs can be omitted or not.

The last variation of the model showed that the results derived for Australia are fairly robust with respect to the functional form used to model initiation into drug use. Using a logistic approach for the initiation function, steady state values and the structure of optimal control turned out to be structurally stable.

Considering the evolution of optimal control as computed for the U.S. and Australian parameterization and for any of the variations, the results showed once more that for a problem that varies over time, the optimal control interventions can be quite distinct over time, too. Furthermore, we could deduce that control interventions that pursue different aims (i.e., use reduction versus harm reduction) may both have merits over the course of a drug epidemic, but depending on the numbers of current users and non-using susceptible individuals not necessarily at the same time.

To the extent that one can generalize the results from these stylized models, the answer to the opponents in the discussion may be that neither the one nor the other strategy is unequivocally the best. Perhaps both are good, important, and advantageous, but possibly not at all times. The main message is that policy makers should recognize that use reduction is not the unique mechanism that can ameliorate drug related problems and reduce so-

cial costs. A decision maker aiming for a comprehensive cost-minimizing drug control strategy should take into consideration that harm reduction may have benefits. Use reduction tactics are traditional and most have been evaluated to be effective. Nevertheless, at certain stages of the dynamic evolution of a drug problem, it may be fruitful with respect to minimizing social costs and harms borne to society to apply harm reduction mechanisms. Proponents of use reduction, and of harm reduction, should not demonize the ideas and downplay the arguments of the other. Rather, they ought to accept that the ideas of the other party may be fruitful and beneficial in at least certain circumstances, and try to reconcile their perceptions in order to build models that will yield better drug control strategies for the future.

By no means it can be claimed that this thesis was the definitive answer how periods with and without harm reduction should vary over time. Much more, the present investigations are only the first steps into the treatise of an important novel area in drug policy. There are various possible extensions that would allow going a step further than the work presented in this thesis. The extensions that should be taken into consideration for future work include the following ideas and approaches:

- People who claim that harm reduction sends the wrong message would probably like to find a more pronounced increase in initiation than modeled in the current work. Additionally, they might criticize that the possibility of an increase of the quantity consumed by prevalent users is ignored in the model. On the other hand, harm reduction proponents do not cease to argue that there is evidence that harm reduction tactics do not have negative effects, in particular that they do not increase initiation. Hence, it might be of interest to investigate other functional forms to model several possibly negative effects triggered by harm reduction.

- Considering the different variations conducted for the U.S. cocaine epidemic and the corresponding changes in the optimal solution, it seems interesting to investigate a model that recognizes a certain amount of innovators into drug use, that assumes a decreased virulence, and that considers an appropriate amount of control costs.

- A bifurcation analysis should be conducted with respect to other parameters, and the Hopf bifurcation occurring when the parameter μ is varied for the U.S. cocaine epidemic (see Chapter 5) should be analyzed in more detail, in order to answer the question whether optimal limit cycles exist.

- Additional control strategies and their impact on the drug epidemic could be included in the harm reduction model.

- The drug use state A could be split up into states modeling different degrees of consumption. Similar to the LH-model (cf. Behrens et al., 1999; Knoll and Zuba, 2004), there could be a division into occasional use, where people are not dependent on the drug (light users), and a state that models those who are dependent on the drug and thus use it with a much higher frequency (heavy users). Auch a SLH-model could be of particular importance, because it would track the career of a drug-using individual from the moment entering the population of susceptible non-users, through the state of relatively unproblematic light use to the escalation to the more problematic state of dependent heavy use. The distinction of different states of drug use might go along with a distinction of the effects of harm reduction measures in the different groups. Adverse consequences for personal health might be more of a concern for the H state, whereas the multiplier effect of the function $g(v)$ could be more pronounced for the flow from the S-state to the L-state.

- Bultmann (2008) extends a (single-state) A-model presented by Grass et al. (2008) by assuming that the drug price is a random variable. Even in that rather simple model, complex optimal solutions occur such as stochastic Skiba (DNSS) points. Given the rich structure of the two-state SA-model presented in this thesis, we may expect that a stochastic version should give rise to even more interesting behavior such as stochastic DNSS curves.

Appendix A

Appendix

A.1 Local Stability Behavior

The local stability behavior of a steady state $(\hat{A}, \hat{S})$ of a two-dimensional system of differential equations $\dot{A}$, $\dot{S}$ can be determined by linearization around the steady state. The Jacobian Matrix J evaluated at a steady state $(\hat{A}, \hat{S})$ is

$$J|_{(\hat{A}, \hat{S})} = \begin{pmatrix} \frac{\partial \dot{A}}{\partial A}|_{(\hat{A}, \hat{S})} & \frac{\partial \dot{A}}{\partial S}|_{(\hat{A}, \hat{S})} \\ \frac{\partial \dot{S}}{\partial A}|_{(\hat{A}, \hat{S})} & \frac{\partial \dot{S}}{\partial S}|_{(\hat{A}, \hat{S})} \end{pmatrix}.$$

The steady state's stability properties are determined with the help of the Eigenvalues e_1 and e_2 of this matrix, or equivalently, evaluating the trace $\tau := \mathrm{tr}(J|_{(\hat{A}, \hat{S})})$ and the determinant $\Delta := \det(J|_{(\hat{A}, \hat{S})})$ of $J|_{(\hat{A}, \hat{S})}$. Note, that the symbol τ used here is different from the "innovators term" introduced to the initiation function in Chapter 2 and used in section 6.2 of Chapter 6. As mentioned in Strogatz (1994), the type and stability of all different sorts of fixed points can be presented on a single diagram (see Figure A.1).

For simplicity we denote the entries of the Jacobian Matrix at the steady state as

$$J|_{(\hat{A}, \hat{S})} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The characteristic equation is given by

$$(a - \lambda)(d - \lambda) - b\,c = \lambda^2 - (a + d)\,\lambda + a\,d - b\,c = 0. \qquad (A.1)$$
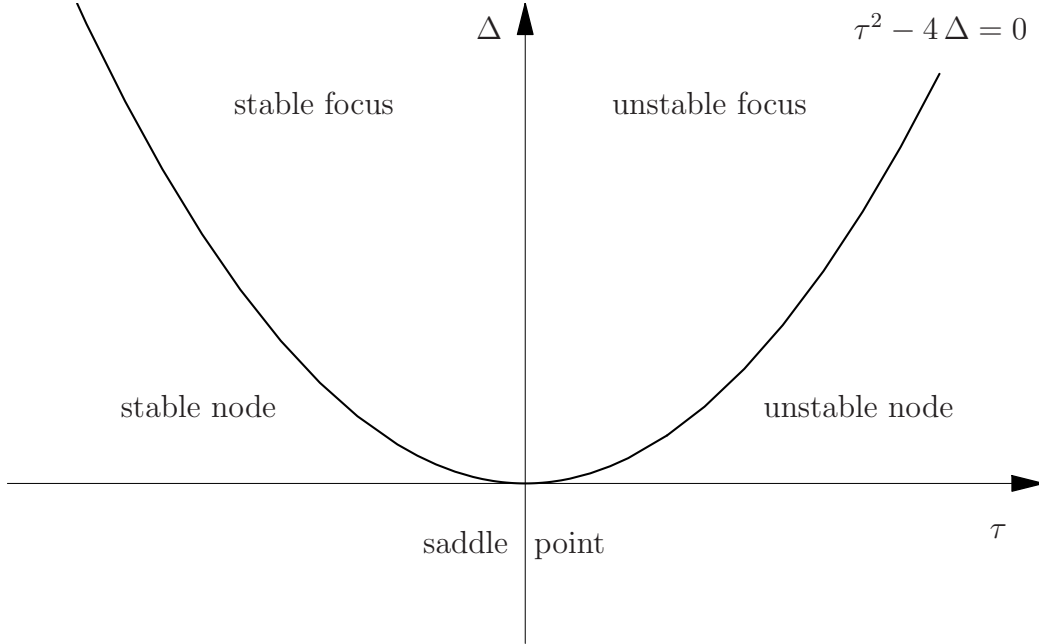
Figure A.1: Type and stability of all different fixed points of a two-dimensional system can be determined with the help of the trace τ and the determinant Δ of the Jacobian Matrix $J|_{(\hat{A},\hat{S})}$.

Determining the trace $\tau = a + b$ and the determinant $\Delta = a\,d - b\,c$, we express the Eigenvalues of the Jacobian Matrix as

$$e_{1,2} = \frac{1}{2}(\tau \pm \sqrt{\tau^2 - 4\Delta}).$$

Furthermore, we can write the characteristic equation in the form $(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - \lambda_1\lambda - \lambda_2\lambda + \lambda_1\lambda_2 = 0$, whereas in equation (A.1) the characteristic equation is written in the form $\lambda^2 - \tau\,\lambda + \Delta = 0$. Comparison of coefficients in the two expressions immediately yields the identities $\Delta = \lambda_1\lambda_2$ and $\tau = \lambda_1 + \lambda_2$.

To arrive at the diagram in Figure A.1, we make the following observations:

- If $\Delta < 0$, the expression under the root is positive, so both Eigenvalues are real and they have opposite signs. Hence, the fixed point is a saddle point.

- If $\Delta > 0$, we have to distinguish the cases $\tau^2 - 4\Delta > 0$ and $\tau^2 - 4\Delta < 0$. In the first case, the Eigenvalues are real, which characterizes

nodes. Foci satisfy $\tau^2 - 4\Delta < 0$, which means that the Eigenvalues are conjugate complex. The parabola $\tau^2 - 4\Delta = 0$ is the borderline between spirals and nodes. On this curve we encounter star nodes and degenerate nodes.

The stability of the nodes and spirals is determined by τ. For $\tau < 0$, both Eigenvalues are negative or have negative real parts, which means the fixed point is stable. Unstable spirals and nodes have $\tau > 0$. The borderline $\tau = 0$ between stability and instability is where neutrally stable centers live. There, the Eigenvalues are purely imaginary.

- If $\Delta = 0$, at least one of the Eigenvalues is zero. Then the fixed point is not an isolated fixed point. There is either a whole line of fixed points, or a plane of fixed points, if $J = 0$.

From Figure A.1 one can easily infer that saddle points, nodes and spirals (foci) are the most relevant types of fixed points, because they occur in large open regions of the (τ, Δ)-plane on the diagram. As analyzed above, centers, stars, degenerate nodes and non-isolated fixed points live only on hairline cases in this plane.

A.2 Proof that $\lambda_0 \neq 0$

We consider a dynamic optimal control problem with infinite time horizon. In contrast to problems with a finite horizon, a constant λ_0 multiplies the integrand of the objective function (see equation (4.6)). We cannot set $\lambda_0 = 1$ a-priori, because the degenerate case $\lambda_0 = 0$ can appear. This degenerate case has to be excluded. There is the general assumption of non-degenerate costates $(\lambda_0, \lambda_S, \lambda_A) \neq 0$. If we can show that $\lambda_0 \neq 0$, then the constant can be set to $\lambda_0 = 1$ without loss of generality.

Let us assume that $\lambda_0 = 0$. The Hamiltonian H (with general initiation function including τ) is then reduced to

$$H = k\,\lambda_S - S\,\delta\,\lambda_S - A\,\lambda_A\,\mu + (\lambda_A - \lambda_S)\,(\tau + b\,A^\alpha)\,S\,g(v).$$

The necessary optimality condition for an interior control is $H_v = 0$. We get

$$H_v = (\lambda_A - \lambda_S)\,(\tau + bA^\alpha)\,S\,g'(v) = 0. \tag{A.2}$$

For the factor $g'(v)$, there holds $g'(v) > 0$ for all admissible values of v. Thus, the above expression can only be equal to zero, if one of the other 3 factors in equation (A.2) is equal to zero.

- At $S = 0$, the system dynamics takes the value $\dot{S} = k > 0$. Thus, the pool of susceptibles would immediately grow to a positive number of users. In section 3.1 we have shown that the A-axis, $S = 0$, is a repeller. Consequently, the number of susceptibles can be zero at the initial condition $S(0) = 0$, but $S(t) = 0$ for a $t \neq 0$ cannot occur.

- For the costates we expect that $\lambda_A < \lambda_S$, because in their interpretation as shadow prices of the states this relation means that an increment in use is always validated to be worse than an increment in susceptibles. Please note that we expect that the costates are both negative in our model because users cause costs and susceptibles are potential future users.

- For the case with innovators, $\tau > 0$, the expression $\tau + bA^\alpha$ cannot be zero for $A \geq 0$, because $b > 0$.

  In the case without innovators, $\tau = 0$, $A = 0$ would have to hold. If once there is no use, $\dot{A} = 0$ holds. Then we encounter the ideal case that the drug epidemic never starts. Of course, this is the most desirable case for the decision maker, but the underlying intention is to investigate optimal control of an illicit drug problem with $A(t) \neq 0$ for the initial time $t = 0$.

Concluding, the degenerate case $\lambda_0 = 0$ is not appropriate here, and we can set $\lambda_0 = 1$ without loss of generality for our calculations.

A.3 Mangasarian Sufficiency Conditions

The crucial point about the Mangasarian sufficiency conditions is that the Hamiltonian must be jointly concave in states and control for a solution that fulfills the necessary conditions to be indeed optimal. The concavity is investigated with the help of the Hessian Matrix of the Hamiltonian H with respect to A, S, and v.

$$
M = \begin{pmatrix} H_{AA} & H_{AS} & H_{Av} \\ H_{SA} & H_{SS} & H_{Sv} \\ H_{vA} & H_{vS} & H_{vv} \end{pmatrix}.
$$

Concavity follows from this matrix M being negative semi-definite. To this end, we must show that the first leading principal minor of M, M_1, is negative, the second leading principal minor M_2 positive and that the third M_3 is negative again.

The first leading principal minor is given by

$$M_1 = H_{AA} = -A^{\alpha-2} \, b \, S \, (\alpha - 1)\alpha(\lambda_S - \lambda_A) \, g(v).$$

The second leading principal minor is

$$M_2 = H_{AA}H_{SS} - H_{SA}H_{AS} = -A^{-2+2\alpha} \, b^2 \, \alpha^2 \, (\lambda_A - \lambda_S)^2 \, g(v)^2.$$

M_3 is the determinant of M. Looking at the expression for M_2, we find a row of squares which are all positive, the term $A^{-2+2\alpha}$ is positive, too. Thus, with the leading minus, $M_2 < 0$ holds. Consequently, the Hamiltonian is not concave and the Mangasarian sufficiency conditions do not hold.

Nevertheless, we can be quite sure that the solutions we encountered are optimal. First of all, the state dynamics induce the nice property, that the steady states are located on a single line in the (A, S)-plane, see equation (3.3) in section 3.1. The search for steady states was conducted with two software packages independently of each other, thus most probably all possible steady states were found. Some steady states were not feasible (Lagrange Multiplier negative). The backward calculation from the remaining candidates for optimality is able to cover the entire first quadrant including the axes. Some candidates were dominated by better solutions. Hence, we can be quite sure that the determined trajectories of the system are indeed the optimal ones.

A.4 Numerics, Software, and Further Technicalities

The analyses presented in this thesis are conducted with Wolfram's Mathematica 5.2 (1988-2005), with Matlab 7.1 (1984-2005), and with the help of MatCont CL (2007) and the toolbox OCMat (2008).

General information on the programs used can be accessed at

- http://www.wolfram.com (Mathematica),

- http://www.mathworks.com (Matlab),

- http://www.matcont.ugent.be (MatCont and MatCont CL).

With respect to the solution techniques and their implementation, Grass et al. (2008) provides an excellent textbook. It also introduces the Matlab toolbox OCMat that was used to conduct significant parts of the present analysis. OCMat is available via http://www.eos.tuwien.ac.at/OR/OCMat. From an often considerable number of switches from boundary to interior control and vice versa there arose a numerical complexity of the model. This led to the challenging task of adjusting and extending the toolbox to be able to handle the several switches.

While conducting the investigations of the SA Harm Reduction Model presented herein, a modus operandi of rotational and complementary work with the different programs turned out to be convenient and advantageous. In particular, Matlab provides the advantage of continuation tools, which allow continuing an existing solution to find further solutions. With respect to calculation of indifference curves and DNSS curves, this is a great benefit. Most of the plots presented in this thesis stem from Mathematica, this shall not mislead to the assumption that Matlab played an unimportant role in the analysis.

To calculate the trajectories that converge towards the steady states, there are two approaches. One can either solve Initial Value Problems (IVPs) or Boundary Value Problems (BVPs). To achieve the results presented herein, both methods were employed. The IVP approach is used with Mathematica, whereas Matlab makes use of the BVP approach.

Using the IVP approach, the system of differential equations is solved in reversed time. The initial values for the IVP have to be chosen in an appropriate neighborhood of the steady state. To find such a neighborhood, we follow the approach introduced in Knoll and Zuba (2004), which makes use of the Eigenvectors of the Jacobian Matrix evaluated at the steady state to span an ellipsoid close around the steady state.

Different from this, the BVP approach for infinite time horizon optimal control problems uses a so-called asymptotic boundary condition. The detailed definition is given in Grass et al. (2008).

Grass et al. (2008) also give a definition of DNSS curves. Please note that in this thesis we do not use the denominations as defined in the textbook, which is for the sake of differentiation between curves that allow for different ways to approach a steady state (called indifference curves in this thesis), and curves characterized by equally optimal convergence to two different steady states (called DNSS curves here).

With respect to the location of the indifference curves and the DNSS curve found in the scenario of decreased virulence of the U.S. cocaine epidemic in

section 6.1, it has to be mentioned that the junction points between the curves might constitute so-called multiple DNSS point. Feichtinger & Steindl (2006) found two DNSS curves in a production/inventory model, whereby those two curves intersect. At the intersection point, a threefold Skiba point in the state space was detected. From this point there emanate three different trajectories with the same accumulated total cost.

Bibliography

[1] Almeder, C., Caulkins, J.P., Feichtinger, G., Tragler, G. (2001). *Age-specific multi-state initiation models: Insights from considering heterogenity.* Bulletin on Narcotics, 53, 1, 105-118

[2] Almeder, C., Caulkins, J.P., Feichtinger, G., Tragler, G. (2004). *An Age-Structured Single- State Initiation Model - Cycles of Drug Epidemics and Optimal Prevention Programs.* Socio Economic Planning Sciences, 38, 91-109

[3] Arrow, K.J., Kurz, M. (1970). *Public Investment, the Rate of Return, and Optimal Fiscal Policy.* Baltimore: The Johns Hopkins Press

[4] Azim, T., Hussein, N., Kelly, R. (2005). *Effectiveness of harm reduction programmes for injecting drug users in Dhaka city [Electronic Version].* Harm Reduction Journal 2005, 2, 22

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

[6] Behrens, D.A., Caulkins, J.P., Tragler, G., Haunschmied, J.L., Feichtinger, G. (1999). *A Dynamic Model of Drug Initiation: Implications for Treatment and Drug Control.* Mathematical Biosciences 159(1), 1-20

[7] Behrens, D.A., Caulkins, J.P., 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

[8] Bultmann, R. (2008). *Stochastic Skiba Points: An Example from Models of Illicit Drug Consumption.* Mimeo

[9] Caulkins, J.P., Kaplan, E.H., Lurie, P., O'Connor, T., Ahn S. (1998). *Can Difficult-To-Reuse Syringes Reduce the Spread of HIV Among Injection Drug Users?.* Interfaces, 28, 3, 23-33

[10] Caulkins, J.P., Everingham, S., Rydell C.P., Chiesa, J., Bushway, S. (1999) *An Ounce of Prevention, a Pound of Uncertainty. The Cost-Effectiveness of School-Based Drug Prevention Programs.* MR-923-RWJ. RAND, Santa Monica, CA

[11] Caulkins, J.P., Pacula, R., Paddock, S., Chiesa, J. (2002). *School-Based Drug Prevention: What Kind of Drug Use Does it Prevent?.* MR-1459-RWJ. RAND, Santa Monica, CA

[12] Caulkins, J.P., Behrens, D.A., Knoll, C., Tragler, G., and Zuba, D. (2004). *Markov Chain Modeling of Initiation and Demand: The Case of the US Cocaine Epidemic.* Health Care Management Science 7(4), 319-329

[13] Caulkins, J.P. (2005): *"Memo #80: SA Model of Harm Reduction".* Mimeo

[14] Caulkins, J.P., Dietze, P., Ritter, A. (2007). *Dynamic compartmental model of trends in Australian drug use.* Healthcare Management Science 2007, 10(2), 151-162. DOI: 10.1007/s10729-007-9012-0

[15] Caulkins, J.P., Tragler, G., Wallner, D. (in submission). *When In A Drug Epidemic Should the Policy Objective Switch from Use Reduction to Harm Reduction?*

[16] Coutinho, R.A. (1998). *HIV and hepatitis C among injecting drug users: success in preventing HIV has not been mirrored for hepatitis C.* British Medical Journal, 317, 424-25

[17] Dave, D. (2004). *Illicit Drug Use Among Arrestees and Drug Prices.* NBER Working Paper #10648, Cambridge, MA, http://ideas.repec.org/p/nbr/nberwo/10648.html

[18] Dave, D. (2008). *Illicit Drug Use among Arrestees, Prices and Policy.* Journal of Urban Economics 63 (2), 694-714

[19] Easingwood, C.J., Mahajan, J., Muller E. (1983) *A nonuniform influence innovation diffusion model of new product acceptance.* Marketing Science, 2 (3), 273-295

[20] Feichtinger, G., and Hartl, R.F. (1986). *Optimale Kontrolle ökonomischer Prozesse - Anwendungen des Maximumprinzips in den Wirtschaftswissenschaften.* Walter de Gruyter, Berlin

[21] Feichtinger, G., and Steindl, A. (2006). *DNS Curves in a Production/Inventory Model.* Journal of Optimization Theory and Applications, 128, 2, 295-308. DOI: 10.1007/s10957-006-9017-8

[22] Gable, R.S. (2004). *Comparison of acute lethal toxicity of commonly abused psychoactive substances.* Addiction, 99, 6, 686-696(11)

[23] 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.* Springer Verlag, Heidelberg

[24] Grossman, M. (2004). *Individual Behaviors and Substance Abuse: The Role of Price.* NBER Working Paper #10948, Cambrigde, MA, http://ideas.repec.org/p/nbr/nberwo/10948.htmlS

[25] Guckenheimer, J., and Holmes, P. (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields.* New York: Springer

[26] Health Canada (2003). *Information on Canada's Drug Strategy.* Health Canada, http://www.hc-sc.gc.ca/ahc-asc/media/nr-cp/2003/2003_34bk1_e.html given as background information to *Renewal of Canada's Drug Strategy to help reduce the supply and demand for drugs*, Health Canada News Release, May 2003, 2003-34, http://www.hc-sc.gc.ca/ahc-asc/media/nr-cp/2003/2003_34_e.html

[27] Johnson, R.A., Gerstein, D.R., Ghadialy, R., Choy, W., and Gfoerer, J. (1996). *Trends in the Incidence of Drug Use in the United States, 1919-1992.* U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, Washington, DC

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

[29] Knoll, C., and Zuba, D. (2004). *Dynamic Models of the U.S. cocaine epidemic: Modeling Initiation and Demand and Computing Optimal Controls.* Doctoral Thesis, Institute for Mathematical Methods in Economics, Vienna University of Technology

[30] Kuznetsov, Y.A. (1998). *Elements of Applied Bifurcation Theory* (2nd ed.). New York: Springer

[31] Leonard, D., and Long, N.V. (1992). *Optimal Control Theory and Static Optimization in Economics.* Cambridge University Press, Cambridge

[32] MacCoun, R.J. (1998). *Toward a Psychology of Harm Reduction.* American Psychologist, Vol. 53, No. 11, 1199-1208

[33] Matcont CL 2.4 (2007). http://www.matcont.ugent.be.

[34] Mathematica 5.2 software (1988-2005). http://www.wolfram.com

[35] Matlab 7.1 software (1984-2005). http://www.mathworks.com

[36] Moore, T.J. (2005). *Monograph No. 01: What is Australia's "drug budget"? The policy mix of illicit drug-related government spending in Australia.* DPMP Monograph Series, Fitzroy: Turning Point Alcohol and Drug Centre

[37] Musto, D. (1987). *The American Disease.* (2nd edition) New Haven, CT: Yale University Press

[38] OCMat (2008). http://www.eos.tuwien.ac.at/OR/OCMat

[39] Office of National Drug Control Policy (2002). *FY 2003 budget summary.* Information retrieved online on October 8th, 2008, see http://www.whitehousedrugpolicy.gov/policy/budget.html

[40] Rigter, H. (2006). *What Drug Policies Cost: Drug Policy Spending in the Netherlands in 2003.* Addiction, 101, 323-329

[41] Riley, D.M. (1993). *The Application of Harm Reduction Measures in a Prohibitionist Society.* Ottawa, CCSA

[42] Ritter, A. and Cameron, J. (2005). *Monograph No. 06: A systematic review of harm reduction.* DPMP Monograph Series, Fitzroy: Turning Point Alcohol and Drug Centre

[43] Scavuzzo, M. (editor) (1996). *Harm Reduction Protocol.* The Chicago Recovery Alliance, see http://www.anypositivechange.org/harmREDprot.pdf

[44] Strogatz, S.H. (1994). *Nonlinear Dynamics and Chaos.* Addison-Wesley Publishing Company

[45] Tragler, G., Caulkins, J.P., Feichtinger, G. (2001). *Optimal dynamic allocation of treatment and enforcement in illicit drug control.* Operations Research, 49(3), 352-362

[46] United Nations Office on Drugs and Crime (2005). *2005 World Drug Report.* Oxford University Press

[47] United Nations Office on Drugs and Crime (2008). *2008 World Drug Report.* Retrieved online on October 8th, 2008, see http://www.unodc.org/unodc/en/data-and-analysis/WDR-2008.html

[48] Wallner, D. (2005). *Optimal Control of the U.S. Cocaine Epidemic: Traditional Approach vs. Harm Reduction.* Master Thesis, Institute for Mathematical Methods in Economics, Vienna University of Technology

[49] Winkler, D., Caulkins, J.P., Behrens, D., Tragler, G. (2004). *Estimating the Relative Efficiency of Various Forms of Prevention at Different Stages of a Drug Epidemic.* Socio-Economic Planning Sciences, 38(1), 43-56

[50] Zeiler, I. (2007). *Optimal Dynamic Control with DNSS Curves: Multiple Equilibria in Epidemic Models of HIV/AIDS and Illicit Drug Use.* Doctoral Thesis, Institute for Mathematical Methods in Economics, Vienna University of Technology

List of Figures

173

List of Tables

Acknowledgements

With this thesis I finish my doctorate studies at the Vienna University of Technology. The thesis could not have been written without the advice and support of several people.

First of all, I want to express my gratitude to my adviser Gernot Tragler, to Gustav Feichtinger, who is heading a project this thesis stems from, and to Jonathan Caulkins, who contributed several suggestions and insightful comments.

Furthermore, I want to thank Doris Behrens, Roswitha Bultmann, Dieter Grass, Josef Haunschmied, Christoph Lanz, and Irmgard Zeiler for their support, enthusiasm, advice, patience and/or sharing of experiences.

My family and my friends supported me in many different ways while I was working on this thesis. In particular, I want to thank my mother Gerda Wallner.

Lebenslauf

Dipl.-Ing. Dagmar Wallner
Deinhardsteingasse 11/8
1160 Wien
E-Mail: dagmar.wallner@gmail.at
Geboren am: 27. Mai 1982 in 7400 Oberwart
Staatsangehörigkeit: Österreich

AUSBILDUNG

01/06 -	Projektassistentin (FWF) an der Technischen Universität Wien, Institut für Wirtschaftsmathematik, Forschungsgruppe für Operations Research und nichtlineare dynamische Systeme
11/05 -	Doktoratsstudium der technischen Wissenschaften an der Technischen Universität Wien
10/00 - 10/05	Technische Mathematik an der Technischen Universität Wien Studienzweig Wirtschaftsmathematik Abschluss mit Auszeichnung Diplomarbeit "Optimal Control of the U.S. Cocaine Epidemic: Traditional Approach vs. Harm Reduction" (Prof. Gustav Feichtinger)
08/04 - 02/05	Erasmus-Auslandssemester Universidad de Alicante, Spanien
09/92 - 06/00	Zweisprachiges Bundesgymnasium Oberwart

FERIALARBEITEN, PRAKTIKA und BERUFSERFAHRUNG

08/06 - 12/06	Praktikum an der Fakultät für Mathematik an der Universidad Pedagogica Nacional in Bogota, Kolumbien
10/00 - 12/05	Firma P.H.I. Technisches Planungsbüro und Handelsges.m.b.H. Unterstützung der Geschäftsführung in diversen Bereichen
07/00	Ferialpraxis in der Technischen Abteilung der Sozialbau AG
07/99	Ferialpraxis in der Technischen Abteilung der Sozialbau AG

WEITERE QUALIFIKATIONEN

Sprachen	Englisch, verhandlungsfähig
	Spanisch, verhandlungsfähig
	Französisch, Grundlagen
	Kroatisch, Grundlagen

INTERESSEN UND EHRENAMTLICHE TÄTIGKEITEN

11/05 -	IAESTE Vienna
	Vereinsmitglied
	Jobraising, Unterstützung bei Firmenmesse-Organisation, Betreuung internationaler Praktikaten
07/03 - 06/05	Fachschaft Technische Mathematik
	gewählte Studienrichtungsvertreterin
	Budgetverwaltung, Organisation und Leitung eines Fachschaftsseminars, Mitglied in diversen Kommissionen
02/01 - 12/05	Fachschaft Technische Mathematik
	Mitarbeit in Interessensvertretung der Studenten
10/00 - 07/04	PAIN, Pannonisches Institut, Güttenbach
	Freie Mitarbeiterin
	Übersetzungen aus dem Kroatischen ins Deutsche