Thesis for the Degree of Doctor of Philosophy

Methods for Stochastic Optimal Control under State Constraints

Per Rutquist

Department Electrical Engineering
Chalmers University of Technology
Göteborg, Sweden 2017

Methods for Stochastic Optimal Control under State Constraints
Per Rutquist

ISBN 978-91-7597-622-8
Doktorsavhandlingar vid Chalmers tekniska högskola,
Ny serie nr 4303
ISSN 0346-718X

Department Electrical Engineering
Chalmers University of Technology
SE-412 96 Göteborg
Sweden

Cover illustration: Contour plot of the probability density, and arrows indicating the closed-loop probability current for a stochastic system with state constraints.

Printed by Chalmers Reproservice
Göteborg, Sweden 2017

To Marielle, Alicia and Felix

Abstract

This thesis looks at a few different approaches to solving stochastic optimal control problems with state constraints. The motivating problem is optimal control of an energy buffer in a hybrid vehicle, although applications are abundant in a number of areas.

Stochastic optimal control problems can be solved via the so-called Hamilton-Jacobi-Bellman (HJB) equation. State constraints result in boundary conditions for the HJB equation causing the value function to go to infinity as the state approaches the boundary, which makes it difficult to solve this partial differential equation numerically.

Different approaches to avoiding infinite values on the boundary are investigated. First, we consider a logarithmic transformation of the value function. This results in an exact linearization, turning the HJB equation into an eigenvalue problem in the one-dimensional case, and also in higher dimensions, but then with certain restrictions on the relation between noise and control cost. Then, for a more general problem formulation, we introduce a different transform which yields a nonlinear problem. It is investigated under what conditions the boundary constraints will be well-behaved, and example problems are solved using a collocation method, demonstrating how a small number of collocation points is sufficient to yield good solutions in those cases. Finally, we consider a method starting from the Fokker-Planck equation. This yields an equivalent problem, but where the value function of the HJB equation need not be computed explicitly, but the probability density function of the closed-loop system is computed instead. This fact can be utilized to focus computational resources on the parts of the state-space that are the most relevant.

Keywords: control theory, optimal control, Hamilton-Jacobi-Bellman equation, stochastic systems, Fokker-Planck-Kolmogorov equation.

List of publications

I Rutquist, P. Breitholtz, C. and Wik, T. (2005), An Eigenvalue Approach to Infinite-Horizon Optimal Control, *Proceedings of the 16th IFAC World Congress*, Prague, Czech Republic.

II Rutquist, P. Breitholtz, C. and Wik, T. (2011), Finite-Time State-Constrained Optimal Control for Input-Affine systems with Actuator Noise, *Proceedings of the 18th IFAC World Congress*, Milano, Italy.

III Rutquist, P. Breitholtz, C. and Wik, T. (2008), On the Infinite-Time Solution to State-Constrained Stochastic Optimal Control Problems, *Automatica* 44(7):1800–1805

IV Wik, T. Rutquist P. and Breitholtz, C. (2010), State-Constrained Control Based on Linearization of the Hamilton-Jacobi-Bellman Equation, *Proceedings of the 49th IEEE Conference on Decision and Control*, Atlanta, GA, USA.

V Rutquist, P. Wik, T. and Breitholtz, C. (2014), Solving the Hamilton-Jacobi-Bellman equation for a stochastic system with state constraints *Proceedings of the 53rd IEEE Conference on Decision and Control*, Los Angeles, CA, USA.

VI Rutquist, P. Wik, T. and Breitholtz, C. (2017), State Constrained Optimal Control via the Fokker-Planck Equation, *Proceedings of the 20th IFAC World Congress*, Toulouse, France.

Acknowledgements

This book would never have come into existence if it were not for the help and encouragement from a number of people.

First, I would like to thank Claes Breitholtz and Torsten Wik, who have both done much more than could be reasonably expected from a thesis supervisor, especially as this project has dragged on for over a decade longer than expected.

I'd like to thank Per Ekdunge, for believing in me and providing me with the possibility to do a PhD thesis project while working at Volvo. Kenneth Holmström, also deserves thanks for letting me continue with my thesis work while at Tomlab.

I would also like to thank the numerous people at Volvo, at Chalmers and at Tomlab who have provided help as well as interesting comments and discussions.

The anonymous reviewers deserve thanks for their feedback on my published papers. I would also like to thank the volunteers at the Wikipedia mathematics reference desk who helped me find proofs and references when I needed them.

I am very grateful for the financial support from the *Gröna Bilen* framework, without which this project would never have come to be. Göran Johansson deserves a special thanks for making things run so smoothly for the PhD students.

Finally, I would like to thank my wife Marielle. For all the support on this project, and for all the support in life, I am eternally grateful.

Per Rutquist

Freiburg im Breisgau, August 2017

Contents

Preface

Every journey must come to an end. When I started this project, I had no idea where it would take me, or how long it would take me to finish. And as I am approaching the end, it feels more like another starting point.

But let us start from the beginning. The starting point for what would become this thesis was a buffer control problem. At Volvo, where I worked at the time, people were making numeric models of hybrid vehicles. These models all contained some kind of energy buffer – a battery, a supercapacitor, a flywheel or a pneumatic tank. The buffer can be used to provide a boost when the vehicle needs power and when braking it can absorb energy, preventing it from being wasted.

The question was: How could we maximize the usefulness of the buffer?

In this context, the word "usefulness" generally meant saving fuel, and thus money and greenhouse gas emissions. The way to save fuel was to use the combustion engine differently than before. Without a buffer, the engine must always supply exactly the amount of power that is required by the vehicle. With a buffer, there is a choice to make at every point in time: How much power should be drawn from, or put back into, the buffer? This choice determines how much power the combustion engine has to supply, and hence the fuel consumption. In theory, this choice could be left to the driver of the vehicle, but in practice hybrid vehicles always let a computer make this choice. As a result, driving a hybrid vehicle is not more difficult than driving a traditional one.

The computer code that makes this choice is called a "control algorithm", and it was part of my job to make such algorithms. When

we had made one, we would run two simulations – one of a hybrid vehicle using the control algorithm, and one of a conventional vehicle (which did not need a buffer control algorithm). We would then compare the computed fuel consumption from the two simulations to estimate the potential fuel savings. If we thought there was a possibility of improvement, we modified the control algorithm and ran simulations again. This process was repeated a number of times, trying to optimize the control policy.

However, rather than doing this over and over again, we wanted the computer to do it for us. The task of finding a control algorithm that gives as good a result as possible is indeed something that can be automated, and roughly speaking it belongs to a scientific field called "optimal control theory". To make matters more complicated, the way a vehicle will be driven is not pre-determined. In a model of vehicle propulsion, the "driver" of the vehicle can be considered to be a "stochastic process" – like rolling dice at every point in time do decide what the driver should do.

This brings us to stochastic optimal control. My focus, throughout this thesis work was an equation called the Hamilton-Jacobi-Bellman equation,[1] which is used to compute a "value function", the gradient of which defines the optimal control policy. In theory, this equation can always be solved and yield the optimal control policy as long as we have a good model of our system, and we have posed the optimization criteria correctly.[2] In practice, it is not always easy to solve this equation with finite computing resources, and finding a way to do so is the main theme of this work.

There were three times that I felt that I took major steps forward.[3] Each time, I was able to tackle more general problem formulations. This means that the earlier papers of this thesis are largely made obsolete by the later ones. Still, I have included them all for archeological purposes.

The final paper of this thesis goes on a different path. It focuses on

[1] The equation that I refer to here is sometimes called the "stochastic Hamilton-Jacobi-Bellman equation" to distinguish it from the special case where the noise term is zero.

[2] Technically, we also need the ability to observe the state that the system is in. What to do when the system state is not directly measurable is another very interesting problem, and could be a subject of another thesis.

[3] These are described in papers I, III, and V respectively.

finding the optimal probability distribution, rather than the optimal value function. These two problems are in fact very closely related. Looking at the problem in terms of a probability distribution instead of a value function gives a different perspective. The state constraints that caused the value function to go to infinity seemed daunting, while the same constraints cause the probability density to go to zero, which seems almost trivial.[4]

Most importantly, this approach opens up an interesting prospect for future work. Working directly with the probability distribution makes it natural to parameterize the problem in a way where the number of parameters is proportional to the probability of the system being in the subset of the state-space where these parameters apply. The total probability is always one, so the total number of parameters would be constant for a given ratio of parameters to probability. This is very different from what occurs when we use a fixed number of parameters for each "volume" of the state space, since this causes the number of parameters to increase exponentially with the number of dimensions. Hence it might be possible to work with significantly system models that utilize much larger numbers of state variables. An initial experiment where the probability density is expressed as a sum of radial basis functions seems very promising. The objective function involves integrals that would require exponential[5] time to compute, but these integrals can be estimated using Monte-Carlo methods. This results in a stochastic element in the gradients, similar to what is often the case in machine learning. Another similarity is that the algorithm is amenable to parallel computation using graphics processing units. There is very interesting work ahead, but a thesis project has to end at some point. Rather than adding another chapter to the end of this book, it will be the beginning of the next one.

[4] Although a difficulty remains in that the control actuation must go to infinity.
[5] in the number of dimensions

Chapter 1

Introduction

The subject of this thesis is stochastic optimal control under state constraints. This first chapter will give a brief introduction to stochastic optimal control and explain what state constraints are and why they are useful, by giving an example of a problem formulation where the theory applies.

1.1 Motivating problem

Consider a hybrid electric vehicle, where an electric motor/generator can work together with the combustion engine to power the vehicle at any point in time. A battery, which can store energy temporarily, is used to allow the engine to be used more efficiently than it would if it always had to match the power demand.

As a simple model, we can write

$$P_{\mathrm{D}} = P_{\mathrm{B}} + P_{\mathrm{ICE}}, \tag{1.1}$$

where P_{D} is the power the driver demands using the accelerator pedal[1], P_{B} is the power drawn from the battery (negative when the electric machine acts as a generator) and P_{ICE} is the power delivered by the internal combustion engine (which can also be negative, due to friction).

Compared to a non-hybrid, where $P_{\mathrm{D}} = P_{\mathrm{ICE}}$, the hybrid has one additional degree of freedom. A choice must be made, at each point in

[1] In a modern vehicle the power is a complicated function of pedal position and other factors, but this will be ignored here.

time, how to distribute the power between the electric machine and the combustion engine. The electric machine and the combustion engine are both controlled by computers[2], so it makes sense to let the choice of power distribution also be made by a computer. The task is then to program this computer to compute the best power distribution, given the data that is available. This data as is referred to as the state[3] of the system here, which will be discussed in more detail below.

The decided electric power will influence the battery level, which follows a differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}E_{\mathrm{B}} = -P_{\mathrm{B}} \tag{1.2}$$

where E_{B} is the amount of energy stored in the battery (often referred to as the "state of charge" when expressed as a percentage of the total capacity $E_{\max}$). We assume that the characteristics (fuel consumption, emissions, etc) of the combustion engine are known. These are measured in a laboratory setting for each type of engine. We also assume that the characteristics of the electric machine and battery, and the statistical distribution of the driver demand are known. Such statistical data can be collected by test driving the vehicle on various routes.

As a control strategy, we aim to compute a reference power $P_{\mathrm{ref}}(x)$, where x is the "state" that the system is in. This state might include the current operating point of the combustion engine and other factors that affect how the system is behaving at the moment. If we let

$$P_{\mathrm{B}} = P_{\mathrm{ref}}(x) \tag{1.3}$$

and

$$P_{\mathrm{ICE}} = P_{\mathrm{D}} - P_{\mathrm{ref}}(x), \tag{1.4}$$

then Equation (1.1) will be satisfied for any choice of $P_{\mathrm{ref}}(x)$.

It is fairly easy to compute a control policy such that the expected value of P_{B} is zero, given a probability distribution for x. This means

[2]The word "computer" is used here to refer any type of device that performs computations. It will typically be a microcontroller rather than the type of computer that sits on a desk.

[3]Some of the data should actually be considered "measured noise" rather than "state", but we do not need to make that distinction yet.

that, *on average*, the battery will neither be charged nor discharged over time. Unfortunately, this does not imply any limit on how far away from the initial value the battery level will fluctuate.

The control strategy must, of course, also ensure that the battery is never charged or discharged beyond its limits. A simple fix is to limit the electric power P_{B} to zero or less whenever the battery is empty, and conversely, zero or more whenever it is full. In between those extremes, we could use a control strategy that does not look at the battery level. Let $\widetilde{P}_{\mathrm{ref}}$ denote the reference power for such a strategy, which is independent of the battery state of charge, but which satisfies $\mathrm{d}E_{\mathrm{B}}/\mathrm{d}t = 0$ on average, under our statistical model.

However, suddenly switching from one policy to another when the battery level reaches its limits is not optimal. We can do much better by adding a feedback loop, letting

$$P_{\mathrm{ref}} = \widetilde{P}_{\mathrm{ref}}(x) + u(E_{\mathrm{B}}) \tag{1.5}$$

where u is an additional power draw that is added for the purpose of keeping the battery state of charge within its limits. The equation for the energy content of the battery can now be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}E_{\mathrm{B}} = -u - \widetilde{P}_{\mathrm{ref}}(x). \tag{1.6}$$

A number of simplifying assumptions will now be made. We assume:

- that $\widetilde{P}_{\mathrm{ref}}(x)$ behaves like a white noise due to the randomness generated by the driver and the environment,
- that the "extra" cost associated with staying within limits does not depend on anything except the battery level E_{B} and the power draw u, and
- that this additional cost is a quadratic function in u and does not depend on anything else.

The second assumption is the most important one. We disregard all the complexity of the evolution of the state x, and replace it with a random noise, so that we are minimizing a cost that only depends on u and the buffer level, under the constraint

$$\frac{\mathrm{d}}{\mathrm{d}t}E_{\mathrm{B}} = -u + \text{noise}, \tag{1.7}$$

which fits into the problem description[4] given in Paper I.

1.2 Buffer control

While we used a hybrid electric vehicle in the above example, the mathematical description of many other buffers can look the same, if we make the same simplifying assumptions. Buffers are found in many different applications. Examples include energy buffers (see e.g. Sciarretta et al., 2004; Li et al., 2013), buffers for information (e.g. Chen and Wong, 1993; Choi and Park, 1994), buffers for economic assets and goods (e.g. Johnson and Montgomery, 1974; Federgruen and Heching, 1999), and of course buffers for fluids (e.g. Faanes and Skogestad, 2003; Falugi et al., 2012). In Paper IV we look at storage tanks in a sewage system and find a model that is similar to that of the hybrid electric vehicle.

Many buffers, like the one in our example, exist explicitly to deal with random disturbances. The characteristics of the disturbance are important, both in selecting the size of the buffer and the control policy. In Paper I, we show that for the simple buffer in our example, the control policy

$$u = \frac{\pi S^2}{E_{\max}} \tan\left(\frac{\pi E_{\mathrm{B}}}{E_{\max}} - \frac{\pi}{2}\right) \tag{1.8}$$

is optimal. Here, $E_{\max}$ denotes the maximum level that E_{B} is allowed to reach, i. e. the buffer capacity. (The lower level is set to zero in order to simplify the notation.) The noise variance is denoted S^2, and we can see that the control actuation will be directly proportional to the ratio of noise variance to buffer size.

Intuitively it makes sense that the control should depend on the noise variance. The larger the randomness we can expect in the future, the greater the advantage of the buffer being approximately half full, and consequently a larger cost can be accepted for moving toward that state.

[4]The example in the paper refers to a supercapacitor instead of a battery, and an average power instead of a reference power, but the mathematics are the same. Our E_{B} corresponds to x in the paper, and $\widetilde{P}_{\mathrm{ref}}(x)$ corresponds to $s\mathrm{d}w$ in the paper.

Because the noise variance has such a large influence on the optimal control policy, it is important to have a good estimate of it when computing the control. If the noise estimate is too low, then the control action will be too weak, and the buffer will frequently end up near its limits where it cannot provide any more buffering effect. On the other hand, if the noise estimate is too high, then the control will be too strong, always keeping the buffer close to the half-full state. This means paying for buffer capacity that will almost never be used.

An important conclusion that can be drawn from this is that it is impossible to compute an optimal control policy for a buffer system without taking the noise characteristics into account. This may sound obvious, but many optimal control techniques disregard the noise as a simplifying assumption, and such techniques are thus not applicable to this type of buffer systems.

1.3 State constraints

State constraints can be found in all kinds of dynamic systems. A tank must not overflow or run dry, a chemical reactor must not exceed a critical temperature, an engine must not run too fast, a robot arm must not hit a wall, and so on.

In our example, the battery level was the only state variable in the model, and this variable was constrained to remain between "empty" and "full". This is a basic example of state constraints. More generally, the state might consist of n different variables, and the state constraints imply that the state of the system cannot leave a given region in the n-dimensional state space.

Mathematically, state constraints are equivalent to imposing an infinite "cost" of visiting the states outside the permitted region, but with numeric methods it is typically better to treat them as boundary constraints.[5] The reason to consider state constraints is often that they represent actual constraints on the physical system that is being controlled. In our hybrid vehicle example, over- or undercharging the

[5]How to formulate these boundary constraints is an important question, which is treated in detail in Paper III for the case where there is a special relationship between noise and control cost, and in Paper V for a more general case.

battery might damage it, which is much worse[6] than using a bit more fuel.

From the perspective of the engineer who wants a control policy for a buffer system, a stochastic model with state constraints is useful. He or she can, for example, impose an infinite penalty for violating state constraints, and no penalty at all for states that do not violate the constraints. The optimal control will then prefer the "half-full" buffer because of the expected future cost of avoiding constraints. This is a very natural choice of penalty function, in contrast to e.g. linear-quadratic control where state constraints are simply not considered and a special form is imposed on the penalty function. It might not be possible to find a penalty function on this form that simultaneously gives good performance and low risk of violating state constraints.

[6]It may not be *infinitely* worse, as our mathematical model implies, but for the purpose of control system design, this is a good enough approximation.

Chapter 2

Multidimensional systems

Dynamic systems are often characterized by how many different variables it takes to completely describe the "state" of the system in the model. The buffer system considered in the previous chapter was reduced to a first-order system, which means that there was exactly one state variable – the buffer level.

In general, a dynamic system can contain an arbitrarily large number of states variables,[1] but when designing mathematical models for the purpose of control, this number is often reduced using simplifying assumptions. If the control policy can take a more complex model into account, then the number of simplifying assumptions can be reduced, and this should[2] result in better performance.

In a hybrid vehicle model, there are several variables that might be available to the controller and be important in controlling the energy buffer. For example:

- The speed of the vehicle. A vehicle that is going fast is more likely to apply regenerative braking in the near future than one that is standing still.

- The slope of the road. On modern roads, the slope is often constant on long sections. The current slope gives a good prediction of the slope in the near future.

[1] Even an infinite number in certain types of models.

[2] It is also important that a model is simple enough to let all parameters be estimated from available data. Models that include everything and the kitchen sink frequently perform *worse* than simple models because of bad parameter estimates.

- The position in the drive cycle.
- The rate of acceleration (gas/break pedal position).
- The speed of the combustion engine. In a series hybrid, and with certain kind of gear configurations, it is possible to vary the speed of the combustion engine independently of the speed of the vehicle.
- The speed of the turbo charger. When the power output of the engine changes, the turbo charger can lag behind, and this limits the available power from the combustion engine.
- A second energy buffer level. Combining, e. g., a battery and a supercapacitor has many theoretical advantages, although it also makes the system more complex and expensive.

The criteria for being included is basically that a state variable should be something that changes on approximately the same time-scale as the buffer level, and influences or is influenced by the energy buffer, directly or indirectly. Different buffers have different time scales so, for example, a battery that takes tens of minutes to discharge can be controlled fairly well by an algorithm that does not take turbo boost pressure into account, because that process reaches a steady-state in seconds. In contrast, a supercapacitor, which discharges much more quickly, needs a controller that accounts for such fast processes.

In general, it is not necessary to include all of the above, but often there is a need to include four or more states variables in order to create a sufficiently complete model.

State variables like vehicle speed, road slope and accelerator/brake demand form a stochastic model of the "disturbances" that the buffer has to deal with, while engine speed and turbo pressure are influenced by the control algorithm and the disturbances.

As mentioned, we denote the number of state variables by n and the state space is said to be n-dimensional. In the case of a simple buffer, we needed to solve a set of equations on every point on a line (a one-dimensional space). If we add another state variable (say the speed of the vehicle), then we need to solve the same equations on every point on a square (a two-dimensional space). Adding yet another dimension, we get a cube, and going further we get things that

are hard to visualize, but that mathematically extend the concept of "space".

So when more states are included in the model, the practical consequence for optimal control is that equations must be solved in a higher-dimensional space.

2.1 The Hamilton-Jacobi-Bellman equation

The equation that is normally used to compute an optimal control policy is called the Hamilton-Jacobi-Bellman equation, which is based on Bellman's principle of optimality (Bellman, 1957). Simply stated: If the optimal path from A to C goes through B, then the optimal path from B to C is a portion of the path from A to C. This makes it possible to work backwards in time, at each point finding the optimum of the instantaneous cost and a future cost, which has already been computed. This future cost is referred to as the "value function" (see e. g. Bertsekas, 1996; Dreyfus, 2002). The HJB equation is a differential equation, which means that the solution is not a simple number, but a function. For our buffer example it has one value for each possible buffer level and each point in time.

In general, the solutions to partial differential equations are impossible to compute exactly. The best one can do then is to compute a large table of numbers that describe an approximation to the solution.[3] This leads us to what is known as Bellman's curse of dimensionality.

2.2 The Curse of Dimensionality

The curse of dimensionality is a difficulty often encountered whenever nonlinear optimal control problems are considered. It states that as the dimension (number of state variables) grows, it becomes increasingly hard to solve the problem using numerical methods. The reason for this is the following:

When solving a partial differential equation, the state-space is

[3]The analytic value function in Paper I is an exception to this.

divided into smaller pieces,[4] so that the computer can work with a finite number of equations even though the equation theoretically applies in an infinite number of points. If the pieces are small enough, then the solution can be a good approximation of the true solution.

For a one-dimensional problem, it might be enough to divide the state-space (a line) into ten pieces. But if we consider a two-dimensional system, then we need to divide the state-space (a square) into a hundred pieces. For a five-dimensional system, we might need 100 000 pieces. In other words, we get the same number of zeroes at the end of the number of pieces as the number of state variables in our system. The number is then said to grow exponentially. Even with very fast computers, we will never be able to consider very large systems.

Note that the curse of dimensionality is a problem for *any* numeric method that computes a control policy as a look-up table over the entire state-space. The number of values increases exponentially, independently of what method is used to compute the control policy. Two ways to avoid the curse completely are if the control policy can be given as a closed-form expression using a polynomial number of parameters, as in the case of the linear quadratic regulator, or if the control policy is never stored explicitly, as in model predictive control.

A third way would be not to compute the policy on the entire state space. We'll come back to this in Chapter 4

2.3 Linear Quadratic Gaussian Control and Certainty Equivalence

In a linear quadratic (LQ) control problem (see e. g. Anderson and Moore, 1971; Dorato et al., 1995) the system is assumed to be linear, and the cost per unit of time quadratic in the state and control variables. There are no state constraints or constraints on control actuation. While this is rarely true for a realistic system, it can be a very good approximation for small fluctuations around a given point.

[4]Not necessarily literal "pieces". Collocation methods may use global polynomials or other functions that interpolate through a set of points, and then the reasoning then applies to the number of interpolation points.

This problem formulation is popular for two reasons. First, the optimal control problem can be solved analytically for these systems. It is sufficient to solve an associated Riccati equation to obtain a linear feedback law which is optimal. An LQ regulator only needs on the order of n^2 parameters in order to control a system in n dimensions. The complexity of computing these parameters is polynomial, rather than exponential, in n, and using a computer it can be done even for very large systems. Second, so called "certainty equivalence" holds for this kind of problem when we introduce additive noise.

Certainty equivalence means that the control law that is optimal for a deterministic system with perfect state information is also optimal for either the corresponding *stochastic* system or the same system but with *imperfect* state information. In the first case, the control policy need not take future disturbances into account. It is optimal to solve a trajectory planning problem from every point in state-space. In the second case the present uncertainty in the state of the system need not be accounted for. A control policy that assumes the system to be in the "most likely" state will still be optimal.

For an LQ system with Gaussian disturbances, both forms of certainty equivalence hold, and the combination of a state estimator (Kalman and Bucy, 1961) and optimal feedback control is called a linear quadratic Gaussian (LQG) regulator. However, for stochastic systems where state constraints are important, LQG theory cannot be applied. It is simply not possible to approximate such constraints by an LQ system formulation in a meaningful way.

2.4 Model Predictive Control

A method that is becoming increasingly popular is Model Predictive Control (MPC) (see, e. g. Morari and Lee, 1999; Camacho and Alba, 2013). As the name indicates, this method consists of using a computer model to predict the future behavior of the system, and then compute the optimal control *on-line*. The advantage of this method is that it can account for fairly complicated models, including models with state constraints and constrains on control inputs, while sidestepping the curse of dimensionality by not attempting to store a complete control policy, but instead computing optimal trajectories as needed in real time.

Originally limited to relatively small and/or slow LQ systems by the requirement that the duration of each optimization run cannot exceed the sample time of the system, advances in both algorithms and raw computing power mean that modern nonlinear MPC can handle fast and complex systems. Still, the need to solve an optimization problem at each time step limits how complex the model is allowed to be, as well as how long the prediction horizon can be, and the computed trajectories will only be locally optimal. For nonlinear nonconvex problems, there is no guarantee of finding the global optimum in general.

MPC that accounts for disturbances is known as robust MPC. Accounting for stochastic disturbances will drastically increase the difficulty, as this means optimizing an infinite number of possible trajectories instead of just one. For this reason, MPC is often implemented under the certainty-equivalence assumption. When disturbances are taken into account, they are often modeled in a simplified way, such as only considering the worst-case scenario, or only considering disturbances in a small subset of the time-steps that are simulated. (For examples, see (Bemporad and Morari, 1999).)

For buffer problems, we have seen that the optimal control policy is often determined almost entirely by the properties of the disturbance. Hence, for the purpose of optimal control, simplifying the model by removing the disturbance altoghether is probably not a good idea.

2.5 Collocation Methods

Collocation is a way of solving differential equations approximately. Instead of trying to find a function that satisfies the equation in every point, the idea is to find a function that satisfies the equation in a given set of points. This set of points is typically chosen as the node points of a quadrature rule (see, e. g. Boyd, 2001).

Collocation methods suffer from the curse of dimensionality, just as any numerical method that attempts to represent the value function over the entire n-dimensional state-space. However, collocation methods are relatively efficient in comparison to other methods of solving partial differential equations, and are often used for high-

dimensional systems, for example in quantum mechanics.[5]

In contrast to finite element methods, which typically work with polynomial basis functions of a fixed degree, collocation methods typically work with global polynomials where the degree of the polynomial increases with the number of collocation points – so-called pseudo-spectral methods. If the solution is smooth, in the sense that high-order derivatives are bounded, then this can result in exponentially imporoved accuracy as the number of points increases.

Unfortunately, the value function of the HJB equation does not satisfy this requirement, as it goes to infinity near boundary constraints. In order to use collocation methods, we need to transform the function into something that is as smooth as possible.

2.6 The logarithmic transformation

For a dynamic system with dynamics that are affine in the control input, and an associated cost function that is quadratic in the control input and where the coefficient matrix of this quadratic function is the inverse of the noise covariance matrix, a logarithmic transformation can be used to show that the *nonlinear* HJB equation is equivalent to a *linear* partial differential equation describing a diffusion with drift. This remarkable result was pointed out by Fleming and Mitter (1982).[6]

This transformation also makes it easy to handle state constraints. If the value function V of the HJB equation is expressed as the negative logarithm of another function Z

$$V = -2\kappa \log Z, \tag{2.1}$$

where κ is a scalar constant that is given by the relationship between noise covariance and cost of control actuation, then the gradient of

[5]The Schrödinger equation is frequently solved for high-dimensional systems. Quantum state variables add up just like in control theory. If two particles – say the two electrons in a Helium atom – interact in a three-dimensional universe, then the Schrödinger equation must be solved in a six-dimensional state-space.

[6]Fleming and Mitter (1982) focus on the diffusion problem and point to the equivalence with an optimal control problem as a way of obtaining results about the former. Later work, such as (Rutquist, 2005) and (Kappen, 2005b) focus on the opposite direction of the equivalence arrow.

the value function becomes

$$\nabla V = -\frac{2\kappa}{Z}\nabla Z, \tag{2.2}$$

and the homogenous boundary constraint

$$Z = 0 \tag{2.3}$$

makes the gradient of V, and hence the control actuation, go to infinity in such a way that the HJB equation is satisfied. (See Paper III for details.)

With further assumptions on the system model, and a specific choice of coefficient in the log-transformation, the diffusion process is governed by the Schrödinger equation (Rosenbrock, 2000). The Schrödinger equation is not only linear, but often also self-adjoint, allowing for very efficient techniques in obtaining numeric solutions.

The solution to the linear partial differential equation of certain diffusion problems can be expressed as the expected value of a path integral, and thus approximated by simulating random paths (Feynman, 1948; Kac et al., 1951). This method is applicable to the diffusion problems that result from the log-transformed optimal control problem (Fleming, 1982; Kappen, 2005a). In a sense, this circumvents the curse of dimensionality, because the accuracy that results from averaging random trials depends on the number of trials and the statistical variance in the outcomes, and not explicitly on the number of dimensions of the state-space.

2.7 An alternative transformation

We would still like to find a transformation that is similar to the logarithmic transformation, but which does not require a relationship between control cost and noise covariance.

The HJB equation depends on the gradient of the value function, ∇V, but not on the function value, V, itself. In Paper V, an alternative transformation is defined in the gradients:

$$\nabla^T V = -\frac{2}{Z}K\nabla^T Z, \tag{2.4}$$

where K is a matrix-valued function of the state. Compared to the logarithmic transformation, (2.2), which has one scalar parameter κ

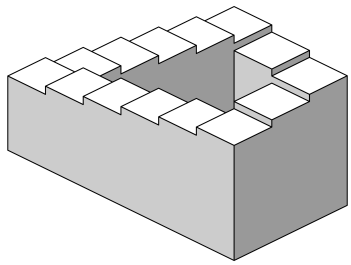

Figure 2.1: The "impossible stairs" (Penrose and Penrose, 1958) is a well-known optical illusion. Intuitively, we know that a path cannot climb continuously and yet return to the same point. (Image credit: Sakurambo / Wikipedia.)

(that is constant over the state-space) the alternative formulation introduces many more degrees of freedom.

This can be used to formulate natural boundary conditions and obtain well-behaved (albeit not linear) partial differential equations for a wider class of problems than that where the logarithmic transformation is useful.

Care must be taken when defining a function via its gradient. If we allow the "gradient" to be an arbitrary function of the state, then we might end up with something analogous to the "impossible stairs" in Figure 2.1. To ensure that there exists a function with the specified gradient, additional constraints are introduced. These must be satisfied together with the partial differential equation that results from the HJB equation.

2.8 Fokker, Planck and Kolmogorov

The Fokker-Planck (FP) equation (Fokker, 1914; Planck, 1917) describes the evolution in time of the probability distribution associated with a dynamic process. Kolmogorov (1931) also deduces this equation for the evolution *forwards* in time, but in addition considers the evolution *backwards* in time of the probability of hitting a "target set". The backwards Kolmogorov equation is identical to the Hamilton-Jacobi-Bellman equation, except that it considers the dy-

namics to be a pre-determined function of state and time, whereas the HJB equation allows for choosing the optimal control actuation at each point in time.

The HJB equation can be directly derived from the adjoint of the Fokker-Planck equation when minimizing the expected cost (Annunziato et al., 2014). However, the adjoint FP equation also contains the probability density ρ in every term. To obtain the traditional HJB equation, the adjoint FP equation is scaled by $1/\rho$, making it independent of the probability density.[7]

Solving the FP equation as it is usually formulated would yield KKT-multipliers that behave as the value function of the HJB equation, i. e. that are unbounded near the state constraints. In order to obtain multipliers that are suitable for a numeric solver, the equation should be re-scaled. In Paper VI, inspired by the Schrödinger equation, the probability density ρ is expressed as the square of another function ϕ, and the FP equation is then scaled by $1/\phi$. This yields a problem formulation that seems to work well with the numeric solver for the sample problems that the method has been tested on.

2.9 PROPT

The software package PROPT (Rutquist and Edvall, 2010) was used to implement Gauss-Legendre collocation in each dimension of the partial differential equations that are solved in Paper V and Paper VI.

Although PROPT is written to work along one dimension (typically time), it is possible to use it recursively to perform collocation in arbitrary dimension, as long as the geometry is a simple hypercube, i. e. upper and lower bounds on each state variable.

Collocation turns the partial differential equation into a nonlinear programming, i. e. an objective function and a set of equations. These are then passed on to a numeric solver.

[7] In a sense, this is the root cause of the numeric difficulties in solving the HJB equation near boundary constraints, where ρ goes to zero.

2.10 Numeric Optimization

There are various techniques for solving a nonlinear programming (see, e. g. Bertsekas, 1999; Bazaraa et al., 2013).

In our case, the commercially available solver SNOPT (Gill et al., 1997) was used. Like many solvers, SNOPT finds the optimum by searching for solutions to the Karush–Kuhn–Tucker conditions (Kuhn, 2014). It employs a method called sequential quadratic programming, successively solving simpler problems where the objective is approximated by a quadratic function and the constraints by linear constraints.

Chapter 3

Summary of the included papers

This thesis is based on a total of six published papers. Because the problem statement has largely remained the same throughout the work, there is some repetition in the initial sections of each paper.

The odd-numbered papers provide three different methods for solving Hamilton-Jacobi-Bellman equations that result from state-constrained stochastic optimal control problems. They follow a logical sequence where each paper generalizes the previous one. The tangent function control policy of Paper I can easily be deduced using the method presented in Paper III, which in turn can be seen as a special case of the method in Paper V.

Paper II applies the method from Paper I to finite-horizon systems. (This paper was published after paper IV, but the papers have been arranged logically by subject matter rather than chronologically.)

Paper IV provides additional analysis and numeric examples for the method described in Paper III.

Finally, Paper VI presents a way of solving the state-constrained stochastic optimal control problem of Paper V, without using the Hamilton-Jacobi-Bellman equation. Even though the size of the numeric problem (number of equations and unknowns) is comparable to those of the method in Paper V, this provides a new avenue for attacking the problem, and numeric convergence seems to be better.

3.1 Paper I

> Rutquist, P. Breitholtz, C. and Wik, T. (2005), An Eigenvalue Approach to Infinite-Horizon Optimal Control, *Proceedings of the 16th IFAC World Congress*, Prague, Czech Republic.

This paper looks at first order systems, where the state constraints simply take the form of a lower and an upper bound on the one state.

The idea of reformulating the Hamilton-Jacobi-Bellman equation as an eigenvalue problem is introduced. The infinite-horizon optimal control problem for a nonlinear, first order, positive recurrent stochastic process is addressed.

It is first shown how an eigenvalue problem is obtained starting from the HJB equation, and how the solution to the eigenvalue problem defines the optimal control. The method is then demonstrated on a fuel cell / supercapacitor system, where it is used to find the optimal way to divide a load between the two power sources as a function of the state of charge of the supercapacitor.

For a simple stochastic buffer problem it is shown that the optimal feedback control policy can be derived analytically and takes the shape of the trigonometric tangent function: $u = -\tan(x)$

It is also exemplified how numeric methods can be used, implying that nonlinear models can be treated as well.

3.2 Paper II

> Rutquist, P. Breitholtz, C. and Wik, T. (2011), Finite-Time State-Constrained Optimal Control for Input-Affine systems with Actuator Noise, *Proceedings of the 18th IFAC World Congress*, Milano, Italy

This is the only paper in the series that does not focus on the infinite-horizon case, but instead considers what happens when the system is operated for a finite time.

For the system in consideration, it is possible to separate the time- and state-dependent parts of the partial differential equations from Paper I. This makes it possible to express the optimal cost to go as a sum of eigenfunctions.

This might be interesting from a theoretical perspective although in practice it is probably better to integrate the PDE backwards in time instead, but we did not pursue this approach any further.

3.3 Paper III

> Rutquist, P. Breitholtz, C. and Wik, T. (2008), On the Infinite-Time Solution to State-Constrained Stochastic Optimal Control Problems, *Automatica* 44(7):1800–1805

In this paper, the eigenvalue approach is extended to systems with more than one state variable. It is shown how a logarithmic transformation leads to an exact linearization of the Hamilton-Jacobi-Bellman equation. This linear partial differential equation and boundary conditions are discussed using the analogy of a convection-diffusion problem, and it is demonstrated how existing, commercially available, finite element software can be used to solve these optimal control problems. The types of boundary conditions that result from state constraints in different contexts are investigated.

This paper was first published in February 2005 as a Technical Report at Chalmers (Rutquist et al., 2005), and included in a licentiate thesis (Rutquist, 2005), which was presented in June. At the time we were unaware of concurrent work by Itami (2005), which was published in June, and by Kappen (2005a,b), which was published in November.[1]

[1] We *should* have been aware of these works in 2007, when preparing the final version of the paper for inclusion in *Automatica*. Also, and in particular, we should have been aware of Fleming and Mitter (1982), who, to our current knowledge, were the first to point out the exact linearization of the HJB equation. It is an unfortunate oversight that the journal publication does not cite any of them.

3.4 Paper IV

> Wik, T. Rutquist P. and Breitholtz, C. (2010), State-Constrained Control Based on Linearization of the Hamilton-Jacobi-Bellman Equation,[2] *Proceedings of the 49th IEEE Conference on Decision and Control*, Atlanta, GA, USA

This paper provides some further analysis of the method presented in Paper III, and discusses the choice of disturbance covariance matrix in the stochastic model. If the disturbance covariance is not chosen to match measured data, it can instead be used as a tuning parameter (or, rather, matrix of tuning parameters) to adjust the control directions in an intuitive way. Sufficient and necessary conditions for when the method can be applied are derived, and their physical interpretation is discussed.

A waste-water treatment control problem with two buffers is used as an illustration.

3.5 Paper V

> Rutquist, P. Wik, T. and Breitholtz, C. (2014), Solving the Hamilton-Jacobi-Bellman equation for a stochastic system with state constraints *Proceedings of the 53rd IEEE Conference on Decision and Control*, Los Angeles, CA, USA

A new method is derived, based on the variable transformation (2.4), which turns the HJB equation into a combination of an eigenvalue problem and a set of partial differential equations. In contrast to Paper III, there is no requirement of an inverse relation between the disturbance covariance and the cost of control. As a numerical illustration, the optimal control of a Linear Quadratic Gaussian system with state constraints is computed. A reasonably accurate solution is obtained even with a small number of collocation points (three in each dimension), which indicates that the method can be used on high order systems, mitigating the curse of dimensionality.

[2]The main author of this paper is Torsten Wik. My main contribution was in the numeric computation of the control policies and simulations shown in Figures 2 – 5.

This paper has been extended since its original publication, and for this reason the notation and equation numbering does not match the original. The boundary condition that appears as $nn^TK = K^Tnn^T$ in the conference paper has been replaced by $n^TK = \kappa n^T$, and the example has been extended with a section on computing the stationary probability density function and computing the expected value of the residual.

Paper VI

> Rutquist, P. Wik, T. and Breitholtz, C. (2017), State Constrained Optimal Control via the Fokker-Planck Equation, *Proceedings of the 20th IFAC World Congress*, Toulouse, France

Rather than starting from the Hamilton-Jacobi-Bellman equation we now start from the Fokker-Planck equation when computing the optimal control policy numerically. Only one PDE needs to be solved, while the infinite boundary conditions are still avoided. Preliminary testing indicates that this method is not only faster but also more robust.

The paper has been slightly modified since the original publication. In the conference paper, the ∇W term of the Fokker-Planck equation was inadvertently omitted. This has been rectified. (The numeric example had $\nabla W = 0$, and is thus unchanged.)

Chapter 4

Conclusions and Future Work

Many important real-world problems can be expressed as stochastic optimal control problems with state constraints. Conventional methods, such as linear quadratic control and model predictive control, have difficulties handling the interaction between stochastic noise and state constraints. This thesis work investigates a series of methods for solving this type of problems, with increasingly general problem formulations.

A few of the main contributions of this thesis are:

- A closed form-solution on the form $u = -\tan(x)$ to the buffer problem in Paper I.
- A demonstration of the logarithmic transformation that linearizes the Hamilton-Jacobi-Bellman equation, and an analysis of the resulting partial differential equation and boundary constraints in Paper III.
- An alternative transformation which does not require the control cost matrix to be the inverse of the noise covariance matrix in Paper V.
- Pointing out a possible way of exploring the relation between the HJB equation and the Fokker Planck equation in solving stochastic optimal control problems with state constraints in Paper VI.

4.1 Future Work

There is much left to do where this thesis ends.

- The most glaring omission is that all of the methods presented are based on having perfect state information. In other words, we assume that all states are measured directly, and that the measurements are completely accurate and instantaneous. This might be the case if the controlled system is itself a computer simulation, but it is never true for physical systems. It would be very useful to have a rigorous way of combining these optimal control methods with estimated states.

- The partial differential equations of Papers V and VI are similar to eigenvalue problems, and it seems plausible that a specialized solver could be developed for these problems by incorporating techniques from eigenvalue solvers into a nonlinear solver.

- The methods used in this thesis all relied on solving the partial differential equations on the entire state space. Such an approach will always be subject to the curse of dimensionality, even though Moore's law and algorithmic improvements will slightly increase the set of tractable problems over time.

4.2 Towards a cure for the curse of dimensionlity

The curse of dimensionality would seem to make it impossible to compute explicit control laws for nonlinear systems in high dimension. Still, we know that such control laws are possible.

Humans routinely learn to control large nonlinear stochastic systems. So our brains clearly have a way to get around the curse, and it is not surprising that the control community has started looking towards techniques such as deep learning for stochastic optimal control.

Preliminary numerical studies, based on the approach in Paper VI, indicated that it should be possible to develop a method that, in a sense, escapes the curse of dimensionality.

When we explicitly compute the probability density function of the closed-loop system, it would make sense to use some parametrization of the state space which emphasizes regions of high probability. One way to do that would be to use radial-basis functions. That is: instead of using a fixed grid and computing the probability density associated with each grid point, we could use freely moving basis functions, where each one represents a constant amount of probability. Since the number of basis functions is constant (for a given amount of probability per basis function) calculation time is no longer exponential in the number of dimensions, and the curse is circumvented.

The algorithm would look something like this:

Start with a set of n randomly positioned "droplets" of probability ;
while *convergece* **do**
    sample m points from each "droplet" under the probability distribution for that droplet;
    for *each sample point* **do**
        compute gradient of *this point's contribution to total cost* with respect to parameters ;
        update parameters using negative gradient ;
    end
end

This mehtod will not necessarily explore the entire state space. (Doing so would, by definition, be subject to the curse of dimensionality.) So we may not always find a global optimum. On the other hand, it is better to get something that is good enough in reasonable time, than to wait forever for something that is perfect.[1]

[1] This conclusion is not limited to the subject of control theory.

Bibliography

Anderson, B.D.O. and Moore, J.B. (1971). *Linear Optimal Control*. Prentice-Hall, Inc.

Annunziato, M., Borzi, A., Nobile, F., and Tempone, R. (2014). On the connection between the Hamilton-Jacobi-Bellman and the Fokker-Planck control frameworks. *Applied Mathematics*, 05(16), 2476–2484.

Bazaraa, M.S., Sherali, H.D., and Shetty, C.M. (2013). *Nonlinear programming: Theory and algorithms*. John Wiley & Sons.

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

Bemporad, A. and Morari, M. (1999). Robust model predictive control: A survey. *Robustness in identification and control*, 207–226.

Bertsekas, D.P. (1996). *Dynamic Programming and Optimal Control, 2nd Edition*. Athena Scientific.

Bertsekas, D.P. (1999). *Nonlinear programming*. Athena Scientific.

Boyd, J. (2001). *Chebyshev and Fourier Spectral Methods*. Dover books on mathematics. Dover Publications.

Camacho, E.F. and Alba, C.B. (2013). *Model predictive control*. Springer Science & Business Media.

Chen, C.T. and Wong, A. (1993). A self-governing rate buffer control strategy for pseudoconstant bit rate video coding. *IEEE Transactions on Image Processing*, 2(1), 50–59.

Choi, J. and Park, D. (1994). A stable feedback control of the buffer state using the controlled lagrange multiplier method. *IEEE Transactions on Image Processing*, 3(5), 546–558.

Dorato, P., Abdallalh, C., and Cerone, V. (1995). *Linear-Quadratic Control, An Introduction*. Prentice-Hall.

Dreyfus, S. (2002). Richard Bellman on the birth of dynamic programming. *Operations Research*, 50(1), 48–51.

Faanes, A. and Skogestad, S. (2003). Buffer tank design for acceptable control performance. *Industrial & engineering chemistry research*, 42(10), 2198–2208.

Falugi, P., Kountouriotis, P.A., and Vinter, R.B. (2012). Differential games controllers that confine a system to a safe region in the state space, with applications to surge tank control. *IEEE Transactions on Automatic Control*, 57(11), 2778–2788.

Federgruen, A. and Heching, A. (1999). Combined pricing and inventory control under uncertainty. *Operations research*, 47(3), 454–475.

Feynman, R.P. (1948). Space-time approach to non-relativistic quantum mechanics. *Reviews of Modern Physics*, 20(2), 367.

Fleming, W.H. (1982). Logarithmic transformations and stochastic control. In W.H. Fleming and L.G. Gorostiza (eds.), *Advances in Filtering and Optimal Stochastic Control*, volume 42 of *Lecture Notes in Control and Information Sciences*, 131–141. Springer, Berlin Heidelberg.

Fleming, W.H. and Mitter, S.K. (1982). Optimal control and nonlinear filtering for nondegenerate diffusion processes. *Stochastics: An International Journal of Probability and Stochastic Processes*, 8(1), 63–77.

Fokker, A.D. (1914). Die mittlere energie rotierender elektrischer dipole im strahlungsfeld. *Annalen der Physik*, 348(5), 810–820.

Gill, P.E., Murray, W., Michael, and Saunders, M.A. (1997). SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Journal on Optimization*, 12, 979–1006.

Itami, T. (2005). Nonlinear optimal control as quantum mechanical eigenvalue problems. *Automatica*, 41(9), 1617–1622.

Johnson, L.A. and Montgomery, D.C. (1974). *Operations research in production planning, scheduling, and inventory control*, volume 6. Wiley New York.

Kac, M. et al. (1951). On some connections between probability theory and differential and integral equations. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. The Regents of the University of California.

Kalman, R.E. and Bucy, R.S. (1961). New results in linear filtering and prediction theory. *Journal of basic engineering*, 83(1), 95–108.

Kappen, H.J. (2005a). Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11), P11011.

Kappen, H.J. (2005b). Linear theory for control of nonlinear stochastic systems. *Physical review letters*, 95(20), 200201.

Kolmogorov, A. (1931). Uber die analytischen methoden in der wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104, 415–458.

Kuhn, H.W. (2014). Nonlinear programming: a historical view. In *Traces and emergence of nonlinear programming*, 393–414. Springer.

Li, X., Hui, D., and Lai, X. (2013). Battery energy storage station (bess)-based smoothing control of photovoltaic (pv) and wind power generation fluctuations. *IEEE Transactions on Sustainable Energy*, 4(2), 464–473.

Morari, M. and Lee, J.H. (1999). Model predictive control: past, present and future. *Computers & Chemical Engineering*, 23(4), 667–682.

Penrose, L.S. and Penrose, R. (1958). Impossible objects: A special type of visual illusion. *British Journal of Psychology*, 49(1), 31–33.

Planck, M. (1917). *Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie*. Reimer.

Rosenbrock, H. (2000). Doing quantum mechanics with control theory. *IEEE Transactions on Automatic Control*, 45(1), 73–77.

Rutquist, P. (2005). On infinite-horizon state-constrained optimal control - with application to a fuel cell APU. Technical Report R020/2005, Chalmers University of Technology, Göteborg, Sweden.

Rutquist, P., Breitholtz, C., and Wik, T. (2005). On the infinite-time solution to state-constrained stochastic optimal control problems. Technical Report R006/2005, Chalmers University of Technology, Göteborg, Sweden.

Rutquist, P. and Edvall, M. (2010). PROPT – Matlab optimal control software. Technical report, Tomlab Optimization.

Sciarretta, A., Back, M., and Guzzella, L. (2004). Optimal control of parallel hybrid electric vehicles. *IEEE Transactions on control systems technology*, 12(3), 352–363.