TUM

Technische Universität München

Fakultät für Mathematik

Unobserved variables and applications of stochastic processes in life sciences

Gerd Bendix Koopmann

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Tim N. Hoffmann

Prüfer der Dissertation: 1. Prof. Dr. Johannes Müller

2. Prof. Dr. Markus Kirkilionis
University of Warwick, Vereinigtes Königreich

Die Dissertation wurde am 13.10.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 05.07.2022 angenommen.

Abstract

Unobserved data play a big role in applied science. The first part of the thesis deals with a model for 'Contact Tracing'. By solving and simulating this model, we can achieve insights into the effect of tracing delays on unobserved aspects of the process. E.g., the impact of Covid-19 tracing apps might not only be based on improved detection of casual contacts, but also on a drastic reduction of tracing delays.

The second part starts from the perspective of observed data of models consisting of systems of ordinary differential equations having a polynomial right-hand side. The gradient of the data is the left-hand side of the ODEs; the parameters of the right-hand side's polynomials are obtainable by regression, which is called Gradient Matching. Assume we know the ODEs, including the polynomial right-hand sides, but we did not observe all components of the data, e.g., we did not observe the current of an LC-oscillator. Then we show how to estimate the unobserved component using the model, the model parameters, and the observed voltage. Due to the local nature, this ansatz does not require initial conditions and provides robust results even if noise events at discrete time points change the trajectory of the observed system, e.g., Dirac impulses re-triggering the oscillator.

Zusammenfassung

Unbeobachtete Daten spielen in der Anwendung oft eine große Rolle. Die Arbeit beschreibt im ersten Teil ein Modell zur Kontaktverfolgung bei infektiösen Krankheiten, mit dessen Hilfe durch Lösen und Simulieren Einblicke in den Effekt der Verzögerung der Kontaktverfolgung auf unbeobachtete Aspekte des Prozesses erreicht werden können. Zum Beispiel basiert die Wirkung der Covid-19-Apps möglicherweise nicht nur auf der verbesserten Erkennung von Alltagskontakten, sondern auch auf der drastischen Reduktion der Verfolgungszeit.

In einem zweiten Teil gehen wir allgemein von den beobachteten Daten aus, die durch Modelle beschrieben werden können, welche aus Systemen gewöhnlicher Differentialgleichungen mit polynomieller rechter Seite bestehen. Der Gradient der Daten ist die linke Seite der DGLs; die Parameter der rechten Seite können durch Regression ermittelt werden. Dieses Verfahren heißt Gradient Matching. Angenommen wir kennen die DGLs und die Parameter der rechten Seite, aber wir haben nicht alle Komponenten der Daten beobachtet, zum Beispiel fehlt uns der Stromfluss eines LC-Oszillators. Wir zeigen, wie unbeobachtete Daten mithilfe der Gleichungen und Parameter des Gradient Matching Ansatzes und der beobachteten Spannung geschätzt werden können. Durch diesen Ansatz sind wir unabhängig von den Startwerten des Experiments und robust gegen einzelne Störungen, die die Lösungstrajektorie verschieben, beispielsweise gegen Dirac-Impulse die den Oszillator wiederholt anschwingen.

Acknowledgements

Contents

Introduction

When analyzing a process, there are often sub-processes, which are hard or impossible to observe. On the other hand, the processes can be modeled based on previous knowledge. In our first example, 'Contact Tracing', there is knowledge about the dynamics of the process. This enables us to build a model. In turn, the analysis and simulation of the model allows us to gain insight into the process that was not achievable before.

In another case, there are occasions when accurate and high throughput data are available without knowing the underlying process. In that case, the data can be used to derive the underlying dynamics, at least if the class of dynamical systems can be restricted [BPK16; Man+16].

A third problem is partially observed data. We assume here, that we know the dynamical system. The goal is not to reconstruct the dynamical system, although this might be possible to a certain degree. In this case, we use the information about the dynamical system to fit the unobserved or partially observed data to reconstruct the unobserved components of the process.

We start the first topic, chapter 1, dealing with contact tracing in the onset of an epidemic. In the case of infectious diseases, there is a general understanding of how people infect each other through contacts. Contact tracing means that we detect an infected individual and start a tracing event. This tracing event consists of searching the contacts of the infected individual to identify chains of infection and to stop the spread of an epidemic. Identifying and searching for the contacts of an infected individual involves time and introduces a tracing delay, the delay between collecting the information about a contact and actually stopping this contacted individual from further spreading the epidemic. Analyzing the model allows us to gain more insight into the influence of tracing delay on the effect of tracing. The influence of the tracing delay on the effect of tracing gained recent importance in the development of digital contact tracing. E.g., the effect of tracing apps as introduced to fight the Covid-19 epidemic might not only be based on the improved detection rate of casual contacts, but also on the reduced tracing delay [Kre+20]. By drastically reducing the tracing delay, the possible infected contacts have less time to further spread the disease. In the case of very long tracing delays, the infected contacts might already have recovered by

the time successful tracing has taken place. This influence of the delay hints at the critical influence of the timing. Fraser et al. [Fra+04] also state that timing is critical. Furthermore, the model enables us to classify contacts into infectors, who infected the current individual and infectees, who were infected by the current individual.

In chapter 1, we model the initial spread of the disease, the onset. The onset is characterized by the exponential spread. In the pure onset case, every contact of an infected individual involves a susceptible individual, and there is a chance of spreading the disease. When simulating a population consisting of a fixed number of individuals, we have a finite population. In a finite population, however, there is a chance of two infected individuals contacting each other, but the prevalence is low during the onset case. In that case, we can neglect the probability of two infected individuals making contact; thus, the model and the simulation match. In chapter 2, we extend the model to the endemic case, which is a steady state. This steady state might exist at a prevalence level, which enables infected individuals to contact other infected individuals at a rate that we can not neglect. We call the contacts between infected individuals cluster contacts. The more infected individuals exist, the more cluster contacts and fewer infective contacts occur. That means fewer infective contacts trace infectors, but every tracing event can include cluster contacts and the tracing by cluster contacts. Chapter 2 builds the endemic model and includes the tracing by cluster contacts. The chapter compares the simulation of the process based on a finite population of individuals to the endemic model. We analyze the change of the influence of the tracing from the onset to the endemic case. Because we have a model, we can separate the individual effect of cluster contacts when analyzing the tracing influence. When entering the endemic phase, the analysis will reveal less influence of tracing by the infectees. The steady state allows only one infection per infected individual on average and, therefore, an average of up to one infectee resulting in a tracing event. Nevertheless, the overall effect of the tracing will stay constant, because the tracing by cluster contacts will substitute the tracing by infectees in a first-order approximation.

The next approach serves as an introduction to our third perspective on unknown dynamics. It was motivated by another common situation in life science: High-quality time series of the state variables are available.

Despite the existence of the data, little might be known about the underlying processes, and the parameters are unknown. The data, of course, include information about the underlying dynamics. However, it is impossible to recover the dynamical system unambiguously from single measured trajectories. Therefore, we restrict the complexity and only allow for ODEs with a polynomial right-hand side. The derivative on the left-hand side is estimated from the data. We can estimate the structure and parameters using regression, which is called

Gradient Matching in this case. At this point, we could integrate previous knowledge about the underlying process to support the regression. Because we choose a polynomial right-hand side, which allows for a lot of collinearity, we will employ penalized regression. In many cases, that approach allows recovering a reasonable dynamical system, which is robust against shot noise within the dynamical system.

In the praxis without Gradient Matching, shot noise influencing the dynamic superimposes a problem. The noise is not part of the dynamic we want to observe, and it is not influencing the measurement apart from its localized occurrences, but results in an altered trajectory of the dynamical system. E.g., if we want to observe an LCR-Oscillator triggered by periodic Dirac-impulses, there is a new trajectory each time the system is triggered. The impulse is not part of the model, and from our point of view, its random occurrence is noise. It is convenient to record the data over multiple impulses to collect more data. However, if we use the classic approach and fit the parameters by solving the differential equations, we need initial conditions. Moreover, each new trajectory needs new initial conditions to solve a differential equation. The initial conditions can be estimated for every trajectory, and there are existing approaches to fit the parameters of a differential equation not only by starting once, but also by starting at multiple different times, to not rely on a single initial condition (see multiple shooting methods [SB96][p. 516ff.][PT07]). Additionally, the multiple shooting fixes the problem of initial values, which are very sensitive to noise, e.g., if exponential growth starts from values close to zero. However, that method requires knowledge about the location of each impulse to align the initial conditions and the start of new trajectories. The local character of Gradient Matching bears reasonable stability of the result with respect to shot noise. At the state of solving the regression, the algorithm only sees pairs of values and corresponding derivatives. Gradient Matching does not require initial values. Therefore, there are no noisy initial values from noisy data. We do not have to deal with cases like exponential growth relying on initial values starting at levels close to zero, which force us to estimate better initial values or start the evaluation of a differential equation at multiple locations. The experiment in chapter 4 illustrates that we indeed recover the appropriate parameters of the ODE in many cases despite a Dirac impulse resetting the trajectory.

This observation is also essential in our third part and serves to introduce the recovery of unobserved variables. Starting at section 3.2 this represents the central part of chapter 3. Some state variables of the dynamical system are easy to observe, and others are harder to observe or unobservable. Opposite to the parameter estimation from recorded data, we know the parameters, but we did not record all the data. In terms of the LCR-oscillator example, we know the differential equation and all components in our system, we recorded the voltage, but the current was unobservable.

Using the Gradient Matching approach, we recover unobserved state variables, vectors of samples, instead of parameters. Especially when recording long time series of state variables, the robustness of this method with respect to time-discrete events due to shot noise and the independence of initial values is an advantage. Neither the initial values nor the positions of shot noise events and corresponding new initial values have to be estimated.

We prove that superimposing a function representing noise on our observed state variable generates a continuous impact on the recovered state variable representing the unobserved state variable. This property is necessary to solve the optimization problem successfully to recover the unobserved state variable.

Finally, both ways of employing Gradient Matching, the estimation of the shape and the parameters of the dynamical system and the recovery of unobserved data, are tested in an experiment using measured data in chapter 4.

1 Contact Tracing – The Onset Case

We deal with an infectious disease spreading within a homogeneous population of finite size. The spread starts with a small number of infected individuals. The infected individuals make random contacts. Each contact of an infected individual with a susceptible individual includes the probability of infecting the susceptible individual. Contact tracing means we trace the contacts of diagnosed individuals, the index cases, to prevent the contacts from further spreading the disease.

Between the diagnosis of an index case, and that of an infected contact, there is a delay. We want to understand the influence of this delay on the process of contact tracing. Contact tracing is meant to reduce the spread of a disease by reducing every infected individual's capability to infect more contacts. Thus, we analyze the influence of the delay on the length of the infectious period.

First, we analyze the onset of an epidemic[1]. We can neglect two infected individuals making contact during the onset of an epidemic within a large random mixing homogeneous population. Thus, we can approximate the process by a branching process, which implies a population of infinite size as there are no contacts between infected individuals. The primary infected person starts a directed tree of infected individuals, where the infected individuals are the nodes, and a directed edge goes from infector to infectee (see fig. 1.1). Ball and Donnelly [BD95] prove the convergence of the branching process and the spread of the epidemic in the onset phase for large populations (further reading in [BP19, chapter 2.3, p. 353]).
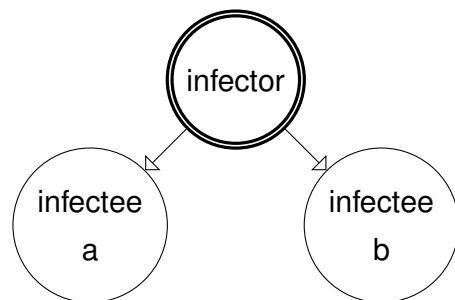
Figure 1.1: Infectors are individuals, who infected other individuals, the infectees. These infectees can be infectors themselves.

We formulate the contact tracing on top of the trees of the branching process. We distinguish one-step and recursive tracing. When we observe an infected individual, one-step contact tracing starts a tracing event for this infected individual. The tracing event follows

---

[1]This analysis of the onset case and the baseline model follow the paper [MK16] 'The effect of delay on contact tracing'.

the branches of that particular individual and analyzes the direct contacts. At this point, we can respect the tracing delay and the probability of recalling this contact successfully. Recursive tracing starts a new tracing event for every infected individual found during this process. This recursion allows for tracing whole clusters of infected individuals. However, the delays involved when following each branch add up. A small probability of recalling a contact has a similar effect compared to long delays. The probabilities are multiplied, and reaching a distanced infected individual becomes unlikely the further the distance within the tree.

When further categorizing the tracing, the directed tree separates two directions. The tree can be traced following the direction of the infection, which results in forward tracing. We call the opposite direction backward tracing. Figure 1.2 shows the influence of both directions on a highlighted individual. In the case of forward tracing, our highlighted individual, the infectee, is traced by its infector. Apart from the start of the epidemic by the index case, every infected individual has an infector. The same individual acted as an infector. As an infector, it can be traced by backward tracing by one of its infectees, if there are infectees. In the analysis, we will (artificially) only allow for backward tracing in the first step, and only for forward tracing in the second step. Combining the forward
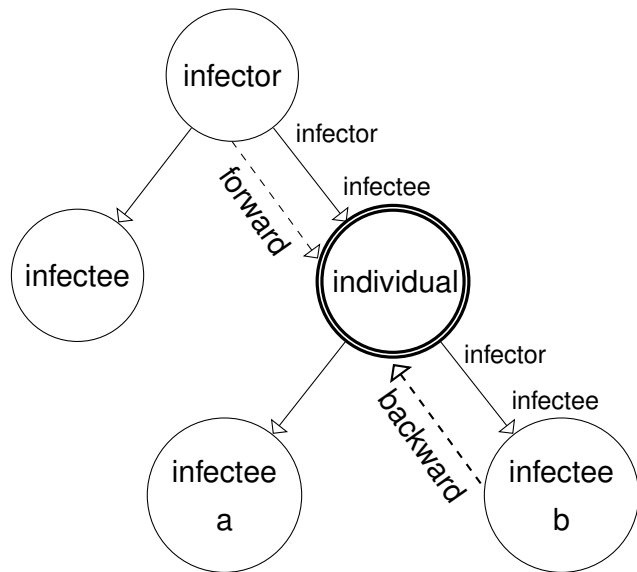
Figure 1.2: Forward and backward tracing visualized within the infection tree. The highlighted individual is an infectee when concentrating on the contact with the infector starting the tree, the index case. On the other hand, the individual is an infector and has its own infectees.

and backward direction on these trees takes all branches connected to an individual spawning a tracing event into account. We call this case full tracing.

The analysis of this branching process allows us to calculate the effect of tracing on each infected individual's probability to stay infective after a given amount of time.

We compare the analysis of the branching process with simulations of finite populations. In this individual-based model, IBM, we keep track of the status of every individual and monitor every contact. This enables us to separate forward, backward and full tracing, activate the tracing delay, vary the tracing delay, and deactivate these processes. Sampling the infective period of infected individuals results in the probabilities of being infective after

a given amount of time, which we compare to the branching process results.

In the branching process model, we stop the infected individuals from spreading the disease. The fact that the individuals are isolated and instantly recovered does not influence the branching process because there will not be a second contact. In the case of simulating an IBM, individuals are susceptible again after recovery. Due to the finite size of the population, individuals can contact infected individuals multiple times, although it is unlikely during the onset case. From the model perspective, being susceptible S or infective I is not only an attribute of the individual but a classification. Individuals transfer from the susceptible to the infective class after an infective IS contact. Observed or unobserved recovery and detection by tracing events result in a transfer from the infective to the susceptible class. We call the model SIS due to these both directions. According to Hethcote, the SIS model is one of the 'Three Basic Epidemiological Models' [Het89], and it becomes crucial when extending our initial baseline model, which bases on the branching process.

After we understand the onset, we will investigate the endemic equilibrium and the adaptions we have to implement in our branching process model in chapter 2. Leaving the onset case, we acknowledge the finite population size. The finite population introduces contacts between already infected individuals, which we call cluster contacts. These contacts influence the spread of the disease. If not every contact involves a susceptible individual, thus, not every contact involves the chance of infecting a new individual, there are fewer contacts actually spreading the disease. Furthermore, it influences the tracing. The infection trees turn into a mesh, which only resembles a tree on a very local scope. When performing one-step tracing, every infected individual still has an infector and might have infectees. In addition, the infected individual might have cluster contacts. In the endemic case, we respect cluster contact tracing beside the forward and backward tracing of the onset case.

This new aspect in tracing raises the question of how the effect of backward tracing on the probability to be infective changes when migrating from the onset to the endemic case and if it is compensated by cluster contact tracing. Furthermore, we will analyze the age structure of the endemic case compared to the onset case.

Before starting the analysis, the next section 1.1 will provide a brief overview of the literature on contact tracing since Hethcote and Yorke's papers on the transmission of Gonorrhea [YHN78; HY84], and we will categorize the models according to Müller and Kretzschmar [MK20]. The categories cover pair approximation models, which approximate the interaction between individuals, and phenomenological approaches, which try not to reproduce the mechanics of the spread and tracing but to map the phenomenons, the IBM Simulations and the stochastic branching process, which we use both within this thesis. Section 1.3 covers the SIS model in more detail and introduces the basic reproduction number, which

enables us to quantify the initial speed of the spread of an epidemic before we introduce the model, and analyze the influence of delays on backward, forward and full tracing.

1.1 State of the Art

We will classify the epidemiological literature dealing with contact tracing (also see the review article [MK20]) into phenomenological approaches, individual-based simulation models, pair approximation models, and models based on branching processes.

Among the first papers dealing with modeling contact tracing was Hethcote and Yorke's paper on the transmission of Gonorrhea [YHN78; HY84]. They model the number of incidences using an ODE. The model introduces contact tracing by introducing a fraction parameter f, which respects for that fraction of infected individuals to be removed. We categorize it as a phenomenological approach as they try to reproduce the outcome by the model using this fraction and do not base it on underlying mechanics.

1.1.1 Phenomenological Approaches

Phenomenological approaches are not based on first principles. They do not select a set of supposed processes, which influence the epidemic. Phenomenological approaches adapt a model, e.g., an epidemiological model consisting of a system of differential equations modeling the number of infected and susceptible individuals, to the data using correction terms and parameters. The resulting model allows for predictions regarding the development of an epidemic, but it might be hard to correlate the results with subprocesses driving or prohibiting the outbreak.

E.g., Hsieh et al. [Hsi+05] build an epidemiological model consisting of a system of ordinary differential equations modeling the number of people knowing of their HIV infection, the number of people knowing of their HIV infection because of contact tracing, the number of people not knowing of their HIV infection and the number of people actually suffering AIDS. The paper benchmarks four linear and nonlinear terms modeling the tracing, i.e., the transfer of individuals from the group of people not knowing of their infection to the group of people knowing of their infection due to contact tracing. The terms are linear to the number of people not knowing of their infection, linear to the number of people knowing of their infection, the product of people knowing and not knowing of their infection and the former term normalized by the sum of people knowing and not knowing of their infection. Fitting the model results in the best fit when using the linear model depending on the number of people not knowing of their infection. The second best fit was achieved by the term consisting of the normalized product of people knowing and not knowing their infection. The paper concludes the normalized product to be a compromise between the two linear terms.

HIV has the property to stay permanent. Infected individuals do not recover from HIV in contrast to influenza. Fraser et al. [Fra+04] uses more first principles to model influenza

and builds a system of partial differential equations, which respects the probability to stay infective as a function of time. He includes contact tracing as a fraction of newly infected individuals, which are traced and stopped from spreading the disease. This approach is applied multiple times, e.g., to model SARS-CoV-2, see Grasly et al. [Gra+20]. Feretti et al. [Fer+20] follow a similar approach, but they acknowledge the tracing delay by delaying the emerging symptoms, which lead to the detection.

1.1.2 Individual-Based Simulation Models

Simulations based on Individual-Based Models, IBMs, on the other hand, implement the mechanics on an individual level to study the outcome. They consist of the first principles and have to sample the outcome from the simulation. Because IBMs base on the individuals, they allow implementing individual-based rules, e.g., if the population is not homogeneous. Meyers et al. [Mey+05] chose a very detailed approach, even modeling cities and schools, highlighting the advantages of IBM models. Keeling et al. [KE05] compare different contact graphs from random graph, lattice, small world, spatial, and scale-free networks. The paper observes a limited growth of the epidemic in all non-random mixing networks, especially the lattices, because the epidemic results in a saturation of the environment. Imagine a growing circle. The circumference, the surface of the infected cluster, experiences linear growth. The limited number of possible non-infected contacts prohibits exponential growth, because infected individuals not living on the border area contact susceptible individuals at a much lower rate, if susceptible individuals are left. The spread within the other networks can be located between these two extremes. Small world networks include some long-range connections, which compensate for the limits set by the lattice. They conclude that it is possible to ignore the higher-order structure and only model pairs of individuals, which allow using pair approximation models, the next approach in section 1.1.3.

Keeling et al. [KHR20] model the outbreak of the coronavirus on a homogeneous population, concluding that the spread of the virus can be contained, if contact tracing detects more than 70% of the contacts. Liu et al. [Liu+15] integrate a delay to the contact tracing process, concluding that a short intervention delay and a high probability of tracing a contact can prevent large outbreaks of measles even in the case of moderate vaccination coverage between 85% to 90%. The model analyzes a very heterogeneous population generated from a synthetic population, i.a., generated from census data. These two papers are examples for two advantages of IBMs. On the one hand, they generate data on given defined contact structures to benchmark analytical models. IBMs can provide the complete

infection graph and history of contacts for further analysis. On the other hand, the capability of directly representing the contact tracing allows to integrate a very high level of detail and quick detailed results to estimate the impact of decisions regarding public health.

1.1.3 Pair Approximation Models

Mean field models approximate stochastic models by deterministic systems of ordinary differential equations. Thereby, mean field models ignore the complexity of the spatial structure, modeling the relationship between individual members of the population. Instead, they assume a relative frequency of individuals of a specified type, e.g., susceptible, infective, or recovered. This approximation simplifies the contact structure, and it derives a deterministic model like the SIR model from a stochastic process. This basic approach can be justified by its success. Hethcote [Het89] considers the SIR model among the three basic epidemiological models. Keeling [Kee99] compares the mean field model to more complex models and concludes a good capability to cover the deterministic behavior while being relatively stable, especially compared to the stochastic models. This characteristic of the mean field models highlights the essential dynamic of the model while not covering stochastic effects or, in that case, oscillations induced by the contact structure. In the process of this comparison, Keeling also introduces pair approximation models.

Pair approximation models extend mean field models by respecting the correlations of an inhomogeneous contact graph. The pair approximation approach does not only respect the relative frequency in each group like susceptible, infected or recovered. The pair approximation considers pairs of individuals and builds a differential equation modeling the pair combination like SI. Sato et al. [SMS94] use a lattice contact graph and build the model, formulate the master equations before employing a pair approximation. Eames and Keeling [EK02] show the performance of mean field and pair approximation on an inhomogeneous contact graph showing the capabilities of pair approximation to respect the contact graph. They point out that pair approximation models allow for integrating contact tracing, because they cover interactions between two individuals by separate equations.

House and Keeling [HK10] discuss the advantage of contact tracing in clustered contact graphs using the pair approximation approach.

Karrer and Newman [KN10] developed a message passing approach to model the endemic case. They do not include contact tracing, but they integrate more information about the contact graph into the system of differential equations. They do not only pass the infection along with the contact graph. They introduce a cavity state for an individual and determine the probability of being infectious for this individual by passing messages along the graph

containing the probability of staying healthy for every individual. Evaluating the differential equation for the susceptible individuals involves multiple products involving all individuals of the population. Wilkinson et al. [WBS17] take this message passing approach and prove a unique, feasible solution.

Clarke et al. [CWT12] take the pair approximation model and apply it to real-world data, even using the tracing delay as a model parameter.

Overall, we see the advantages of the underlying differential equations to get analytical results and the great effort involved, if inhomogeneous contact graphs are respected. Apart from the message passing algorithms, the pair approximation models concentrate on modeling the population level. Therefore, the model is not predestined to implement individual perspectives, but covers contact properties on the population level well.

1.1.4 Models Based on Branching Process

The pair approximation models viewed the spread of the epidemic from the perspective of the population. The branching process chooses a more individual-based level and builds an infection tree to model the influence of the neighboring individuals on the current individual. The branches are directed from the infector to the infectee. This tree is built by a birth-death process, and every birth represents a new infected individual within the tree. Every death means an individual recovered. Recovery means not being available as an infector and infectee, thus, being removed from the infection tree and breaking up the tree into a forest. During the onset of an epidemic, contacts between infected individuals are unlikely, and hence the branching process approximation is appropriate during this phase. Ball and Donelly [BD95] prove that the branching process, which assumes an infinite population size, converges to the spread of an epidemic within a finite homogeneous population in the onset case.

The infection tree allows for an implementation of contact tracing. When an infected individual triggers a tracing event, the tracing analyzes the adjacent nodes. At this point, the process can respect the probability of successfully tracing a contact and a tracing delay. If an individual is traced successfully, it recovers and is removed from the tree of infected individuals. The implementation takes the perspective of the recovered individual. Thus, we model the function of the probability to be infective after a given time since the start of the infection. The contact tracing lowers the probability of being infective.

Müller et al. [MKD00] and Ball et al. [F G11] analyze the branching process. Klinkenberg et al. [KFH06] analyze the effectiveness and the effect of delay in tracing on real epidemics like SARS. Ball et al. [BKO15] introduce a tracing delay and a probability to include unsuccessful

tracing of a contact. Though, this approach only covers forward tracing.

Recent papers adapt the branching process in order to take inhomogeneous populations into account. Okolie et al. [OM20] adapt the ideas of the pair approximation model and integrate message passing.

Our baseline model implements the tracing delay within a branching process while being able to separate the effect of backward and forward tracing. The model provides the probability of being infective as a function of the time since infection. After evaluating the effect of varying tracing delays and their variance, we adapt the branching process to the endemic case. This endemic case integrates the probability of being infective in an age-structured model. The age-structured model introduces elements of the mean field models and keeps track of the number of susceptible and infected individuals to adapt the corresponding contact probabilities.

1.2 Notation

Constants and Rates

α Spontaneous recovery rate. This recovery is not observed and therefore not traced $\alpha = (1 - p_{\text{obs}})\gamma = \left(\frac{\alpha + \sigma}{\alpha + \sigma} - \frac{\sigma}{\alpha + \sigma}\right) \cdot \underbrace{(\alpha + \sigma)}_{\gamma}$.

β Contact rate.

γ Recovery rate: $\alpha + \sigma$.

N Size of population.

p_{obs} Probability of observing a recovery: $\frac{\sigma}{\alpha + \sigma} = \frac{\sigma}{\gamma}$.

p Probability of being diagnosed, when being infected and tracing is taking place.

R_0 Basic reproduction number denoting the average number of infectees per infected individuals.

R_{ct} Reproduction number with contact tracing.

$r_{\pm}$ Update to the basic reproduction number due to forward/backward tracing.

σ Observed recovery, direct observation: $\sigma = p_{\text{obs}}\gamma$.

Functions and Variables

a Time since the start of an infection.

$\eta_{\pm}(p)$ Update to κ due to forward/backward tracing. Determining the intensity by the probability of successful tracing is a choice taken due to the approximation in p.

$i(a, t)$ Number of infected individuals of age a at time t.

$I(t)$ Number of infected individuals at time t. $I(t) = \int_0^\infty i(a, t)da$.

$\kappa(a)$ Probability of being infectious after time of infection a.

$\kappa(a, t)$ Probability of an individual of age a at time t to be infective.

$\kappa_i^+(a)$ Probability of the ith generation being infectious after time of infection a in case of forward tracing.

$\kappa_-(a)$ Probability of being infectious after time of infection a in case of backward tracing.

$\kappa^{\#}(a)$ Integral over the probability of being infectious, used for accumulated effects: $\kappa^{\#}(a) = \int_0^a \kappa(\tau)d\tau = (1 * \kappa)(a)$.

$\hat{\kappa}(a)$ Zero order approximation of $\kappa(a)$: $\hat{\kappa}(a) = e^{-(\mu+\sigma)a}$ for $a \geq 0$ and $\hat{\kappa}(a) = 0$ for $a < 0$.

$\phi(t)$ Distribution of tracing delay.

$S(t)$ Number of susceptible individuals at time t.

t Time since start of the epidemic.

T Expected tracing delay.

Function spaces

C^0 Continuous function.

C^n Continuous function having n continuous first derivatives.

Operations

$(f * g)(a)$ Convolution of f and g: $(f * g)(a) = \int_0^a f(a - \tau)g(\tau)d\tau$.

$f^{\#}(a)$ Cumulative function of f: $f^{\#}(a) = \int_0^a f(\tau)dt = (1 * f)(a)$.

Table 1.1: Notation – Contact Tracing.

1.3 Biomathematical Background

We categorize the individuals of our homogeneous population in susceptible cases S and infected I. Infective contacts allow the transition from S to I, and infected individuals can recover, modeled by a transition from I to S. This resembles the SIS-model mechanics.

To provide some mathematical background, this section introduces the SIS model and a measure for the speed of the spread of the epidemic, the basic Reproduction number R_0. In the end, we introduce the convolution and the corresponding notation, which we will use to distribute the effect of tracing and its delays in the upcoming chapters.

Before introducing the SIS model, we will start with a simpler model, the birth process, also known as the Yule process [Yul24]. The Yule process assumes each individual to give birth at a rate of β and never to die. This resembles our branching process of the onset. Every infected individual contacts other individuals at rate β spreading the disease by creating a new infected individual. The simplicity of the Yule process allows sketching the transition from a probabilistic model, which recognizes the individuals and the stochastic of individual events, to a deterministic model, which models the average by differential equations.

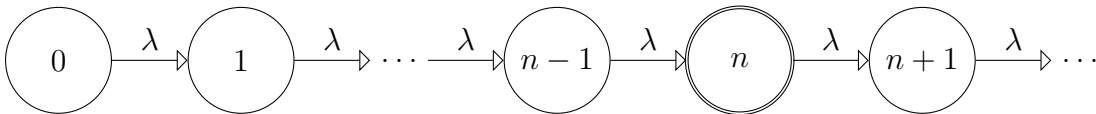The probabilistic model of this Yule process bases on a Poisson process[1].

Figure 1.3: The Poisson process. Transitions from one state to the next occur at rate λ.

Figure 1.3 shows a Poisson process starting at state 0. There are only independent transitions to the direct successor $n \to n + 1$. The transition rate is proportional to the time Δt spent in a state and the parameter λ. We choose Δt to be that short that we can neglect multiple transitions to happen within the same epoch Δt.

The probability of being in a specific state n at time t now depends on the probability to observe n positive outcomes since the start of the process at time $t = 0$. We can calculate the probability of staying in a state after time t using $P(X_{P_\lambda}(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$, the Poisson distribution.

Figure 1.4 visualizes the Yule process. We scale the birthrate β according to the size of the population. In the case of an epidemic, we have the contact rate β and scale it according to the number of infected individuals [BvW08, p. 286 ff.]. This scaling models that two infected

---

[1]Further reading about this view of the Poisson process can be found in the book of Feller [Fel68, p.446 ff.]. A more in depth and formal analysis of point processes and Poisson processes can be found in [DV98]. Bailey [Bai64, p. 84 ff.] focuses on the applications like the Yule Process.

individuals double the probability to infect a third individual compared to a single individual's probability of infecting a second one. We start the Yule process with a single individual.
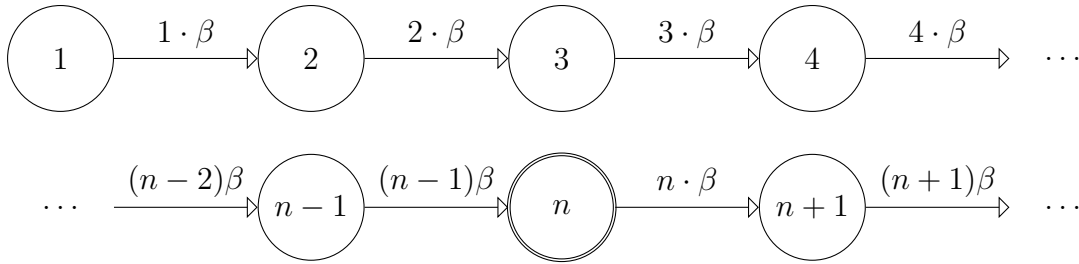

Figure 1.4: The Yule process. The transition rate is proportional to the size of the population or to the number of infected individuals.

Probabilistic Yule Process   We can describe the probability of being in a state, i.e., experiencing a certain number of infected individuals, by describing the two possible events taking place in a period Δt relative to the prior probability. There might be no contact, which means the state did not change during the epoch Δt or a contact triggered a transition from state $n-1$. Thus, the probability $P(X_Y(t+\Delta t) = n)$ is derived from the probability $P(X_Y(t) = n)$ and not changing the state and from the preceding state $P(X_Y(t) = n-1)$ and the rate of a birth event taking place. We will use $Y_n(t) = P(X_Y(t) = n)$ for better readability:

$$Y_n(t + \Delta t) = Y_n(t)(1 - n\beta\Delta t) + Y_{n-1}(\Delta t)(n-1)\beta\Delta t + o(\Delta t).$$

Subtracting $Y_n(t)$ and dividing by Δt results in a difference quotient, which leads us to the differential equation for states n succeeding the initial state i, which fig. 1.4 set to 1:

$$Y_n'(t) = -n\beta Y_n(t) + (n-1)\beta Y_{n-1}(t) \quad \text{for } n > i.$$

We provide initial conditions, which start the system in the state i. $Y_i(0) = 1$ and $Y_n(0) = 0$ for $n \neq i$. The differential equation for the initial state reduces to $Y_i'(t) = -i\beta Y_0(t)$. We can verify by successive solving of the differential equation the formula

$$Y_n(t) = \binom{n-1}{n-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{n-i} \quad \text{for } n \geq i > 0.$$

Further reading and a method of verification using generating functions can be found in [Bai64, p. 84 ff.].

Deterministic Yule Process   When searching for a deterministic counterpart of the Yule process, we know that the rate β being scaled by the size of the population results in an exponential function. The exponential function is the solution of the following differential equation:

$$\frac{dx(t)}{dt} = \beta x(t).$$

The initial value and the growth rate, the contact rate β, define the exponential function

$$x(t) = x(0)\ e^{\beta t} \text{ with } x(0) = i, \text{ the initial population size.}$$

A deterministic form of the process is useful when modeling large populations. The larger the population, the smaller the impact of every single event from the stochastic process, and the average effect becomes more visible [Ren12, p. 19 f.]. We can hint at the convergence of the process to the differential equation using the law of large numbers as a heuristic argument. Because the process does not consist of a single scalar random variable, we would have to employ Donsker's theorem [Bil99, p. 86 ff., section 8, theorem 8.2].
Ethier and Kurtz [Kur80] [EK86, p. 452 ch. 11] concentrate on the convergence of stochastic and deterministic population models[1]. We will use the convergence results without further discussing the proof.

Stochastic SIS Process   We take the stochastic Yule process and extend it to represent a stochastic SIS process on a population of N individuals. We implement a fixed number of N states representing the infected individuals of our population. Because we know the population size, each state n represents not only a number of infected individuals, but $N-n$ also denotes the number of susceptible individuals.
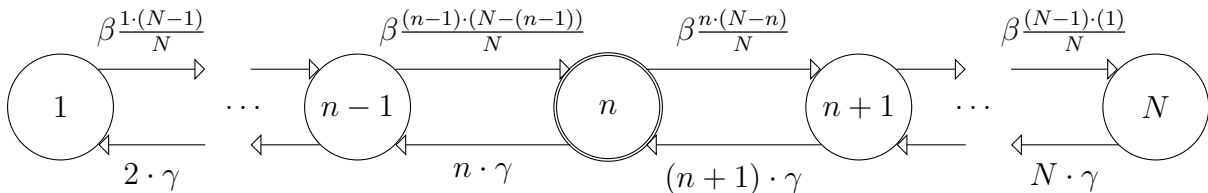

Figure 1.5: A stochastic view on the SIS process.

Figure 1.5 visualizes the transitions between the states. We introduce a recovery rate γ enabling transitions $n \to n-1$. The contact rate β, which results in infections, is not only

---

[1]Further reading and some more literature can be found in [AB00, p.39].

proportional to the number of available infected individuals as potential infectors, but also proportional to the number of available susceptible individuals to become infectees. The states 1 and N are the boundaries of our process. In the state $n = N$, there are no further susceptible individuals to contact. We block state 1 from recovering to prohibit extinction. Further reading about this model can be found in the book of Nåsell [Nås11, p. 13].
The following model demonstrates the concept and shows the stochastic process resulting in the deterministic model. We will proceed using the following deterministic version.

Model 1 (Stochastic SIS Model) [1] *A stochastic SIS model with at least one permanent infector consists of the state space:*

$$X_{SIS}(t) = \{1, \ldots, N\}, t \geq 0 \text{ representing the number of infected individuals at time t.}$$

The transition rates for the contacts β_n and for the recovery γ_n for the states

$$\beta_n = \beta \frac{n(N-n)}{N} \quad \text{for } n = 1, \ldots, N-1,$$
$$\gamma_n = \mu n, \quad \text{for } n = 2, \ldots, N$$

The infection rate in the state $n = N$ would evaluate to 0, but is left out to underline that we do not leave the state space. The state 0 and recovery rate for $n = 1$ can be removed to prohibit the probability of extinction of the endemic, as shown in fig. 1.5.
We start the process with a single individual at $t = 0$:

$$P(X_{SIS}(0) = 1) = 1,$$
$$P(X_{SIS}(0) \neq 1) = 0.$$

The probabilities at later times can be calculated using the transition rates. First, the probability of being in state n is increased by the birth rate from $n - 1$. Then there might be no transition during the epoch Δt. The third influence is the influx by the recovery rate of the succeeding state $n + 1$ [Cav78].

$$
\begin{aligned}
P(X_{SIS}(t + \Delta t) = n) = {} & P(X_{SIS}(t) = n - 1) \, \Delta t \beta \frac{(n-1)(N-(n-1))}{N} \\
& - P(X_{SIS}(t) = n) \, \left(1 - \Delta t \beta \frac{n(N-n)}{N}\right) \\
& - P(X_{SIS}(t) = n) \, (1 - \Delta t n \gamma) \\
& + P(X_{SIS}(t) = n + 1) \, \Delta t (n+1) \gamma
\end{aligned}
$$

---

[1] We use the stochastic definition of [Fel68, p.448].

Further reading about stochastic processes and turning this difference equation into differential equations, Kolmogorov equations, to achieve analytical distributions can be found in the book of Nåsell [Nås11, p.17 ff.] and in [Ros96].

We will proceed using the deterministic SIS model, which describes the average of the epidemic spread using two states, susceptible and infected. The transitions between the states are modeled using differential equations using the defined rates of recovery and contact. Figure 1.6 visualizes the deterministic SIS-Model.

Hethcote [Het89] considers the SIS model as one of the three basic epidemiological models. It originates from a more complex model from Kermack and McKendrick [KM27][1].
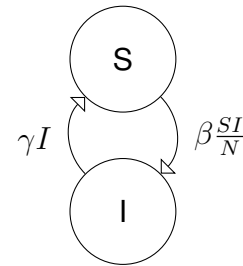

Figure 1.6: SIS Model.

Model 2 (SIS Model – Deterministic)

$$N = \textit{Size of population.}$$

$$S(t) = \textit{Number of susceptible individuals at time t.}$$

$$I(t) = \textit{Size of population.}$$

$$\gamma = \textit{Rate of recovery.}$$

$$\beta = \textit{Rate of contact.}$$

$$\dot{S}(t) = -\beta\frac{S(t)I(t)}{N} + \gamma I(t).$$

$$\dot{I}(t) = +\beta\frac{S(t)I(t)}{N} - \gamma I(t).$$

The rates of the SIS model describe the epidemic, but it is usually not easy to observe if an epidemic is going to spread and how fast this is going to happen. We introduce the basic reproduction number to characterize this important property.

Basic Reproduction Number   The basic reproduction number R_0 can be used to measure the spread of an epidemic. It denotes the number of infections of a typical infected individual during his lifetime, assuming that all contacts are susceptible. Regarding the spread, the basic reproduction number defines a threshold. If R_0 is greater than one, the disease can invade. If R_0 is smaller than one, the disease will not invade. Diekman et al. [DHM90] [DHB13, p. 161 ff.] define R_0 as the dominant eigenvalue of a positive linear operator. This underlines the complexity involved, if nonhomogenous populations are analyzed.

---

[1]We only use a much simpler version of the Karmack-McKendrick model, e.g., it includes a latency period. The paper was reprinted together with its following papers in 1991, [KM91a], [KM91b], [KM91c]. A historical summary of the modeling of infectious diseases can be found in [And82, p.1 ff.].

Due to interactions between the groups, the R_0 value might be significantly lower than the value of the most infective individual or lower than the average of the groups.

We analyze a homogenous and random mixing population, which simplifies the definition. Every infected individual is a typical individual, and we know the corresponding parameters. In the case of our SIS model, we know the contact rate and the expected lifetime of an infected individual due to the known recovery rate γ. The probability of surviving, i.e., staying infective, of a single individual can be modeled using only the recovery part of the differential equation

$$\dot{I}(t) = -\gamma I(t).$$

The probability of being infective right at the time of infection is 1:

$$I(0) = 1.$$

The solution of the differential equation is an exponential decay:

$$I(t) = e^{-\gamma t}.$$

The probability of staying infective follows the exponential distribution, and we know the expected value of an exponentially distributed random variable X with parameter γ to be $\frac{1}{\gamma}$,

$$E[X] = \int_0^\infty \gamma x e^{-\gamma x} dx = \frac{1}{\gamma}.$$

Proposition 3 (Basic Reproduction Number R_0) *The basic reproduction number R_0 denotes the unrestricted expected number of infections of a single, typical infected individual. In the SIS model from model 2, we know the contact rate β and the expected lifetime $\frac{1}{\gamma}$, which results in*

$$R_0 = \frac{\beta}{\gamma}.$$

Our model in the upcoming sections will describe the probability of being infective. We will include the impact of tracing and tracing delays. These influences are concatenated. The distribution of the delay operates on the distributed impact of the tracing. Therefore, we introduce some notation regarding the convolution.

Convolution   In our upcoming model, we concatenate the distributed effect of tracing with the tracing delay distribution. Therefore, we introduce the convolution.

Notation 4 (Convolution of f and g)

$$(f * g)(a) = \int_0^a f(a - \tau)g(\tau)d\tau.$$

We need to accumulate functions, which equals a convolution by 1. To shorten the notation, we introduce the cumulative function.

Notation 5 (Cumulative Function of f:) *The cumulative function of $f(a)$ is $f^{\#}(a)$:*

$$f^{\#}(a) = \int_0^a f(\tau)dt = (1 * f)(a).$$

1.4 Model

This baseline model consists of a branching process formulation and covers the onset case of an epidemic. This chapter follows closely the publication [MK16] 'The effect of delay on contact tracing'.

The model describes a random mixing homogeneous population. There is a primary case introducing the disease. Because we concentrate on the onset, infected individuals only have contact with susceptible individuals. These contacts occur at a rate β and transmit the disease. Thus, the rate β is the infectious contact rate in this model. As explained above, this process builds a directed graph of infected individuals, a directed infection tree. Infected individuals can recover by spontaneous unobserved recovery at the rate of α or recover by an observed recovery at the rate of σ. We call the quotient $\frac{\sigma}{\alpha+\sigma}$ the fraction of observed recoveries p_{obs}. Observed recoveries trigger a tracing event.

Within the tracing event, every branch contacting that individual is followed at the probability of recalling contacts p. Following a branch involves the tracing delay drawn from the delay distribution ϕ. When evaluating the delay, we will use the expected delay T. If the named contactee is still infective, it also is diagnosed, and removed from the class of infected individuals If we perform recursive tracing, this observed recovery triggers a new tracing event.

We analyze the tracing effect by concentrating on the probability of an infected individual to stay infective. Without tracing, we denote by $\hat{\kappa}(a)$ the probability to still be infective after time of infection a. Contact tracing lowers this probability. The directed tree of infections enables us to concentrate on forward tracing, backward tracing, and full tracing separately. Forward tracing only respects the tracing in the direction of the contact graph. Forward tracing traces the infectee by its infector. Thus, every individual, except the initially infected individual, has a predecessor, which can trigger tracing events. A predecessor, which has a probability of being traced himself, has a different probability of triggering a tracing event. We respect this history of the infection by introducing generations i. We denote the resulting probability of being infective by $\kappa_i^+(a)$. I.e., the probability of the index case to stay infective $\kappa_0^+(a)$ equals $\hat{\kappa}(a)$, because there is no infector, which can trigger a tracing event. All the following generations will experience a lowered probability of being infective due to the tracing by their infectors.

Backward tracing respects the opposite direction. The infectors are traced by their infectees. The contacts create infectees over the lifespan of the infected individual. It is a continuous process starting with the existence of the infected individual. Thus, the backward tracing does not involve generations. We denote the probability to still be infectious in the case of

backward tracing by $\kappa_-(a)$.

For mathematical reasons, we first investigate backward tracing only and then forward tracing only. It turns out that these artificial restrictions allow for a convenient analysis of the simplified process. Afterwards we use the results obtained to investigate full tracing, the combination of forward and backward tracing, which we denote with $\kappa_i(a)$, where the index denotes the generation introduced by the forward tracing.

These $\kappa(a)$ allow us to calculate the basic reproduction number R_0 according to proposition 3 and approximate reproduction numbers R_{ct} respecting the contact tracing, which we can compare.

Corollary 6 (Basic Reproduction Number R_0) *In this model, we can calculate the basic reproduction number by*

$$R_0 = \int_0^\infty \beta \hat{\kappa}(a) da.$$

Corollary 7 (The Effective Reproduction Number R_{ct}) *The reproduction number respecting forward, backward, or full contact tracing can be calculated using the corresponding probability to stay infective $\kappa(a)$ within the formula of the basic reproduction number from corollary 6.*

To compare the $\kappa(a)$ to data from a process directly implementing the population, we use a simulation based on an individual-based model[1]. It simulates each individual within the homogeneous finite population N. We start the disease by introducing several infected individuals into the population. We prohibit stochastic effects from stopping the epidemic at this point by using more than one infected individual. Each individual of the population can become infective or susceptible and has a contact history, which we use to execute contact tracing. We simulate the rates of unobserved recovery α, observed recovery σ, and the contact rate β per individual. An observed recovery triggers a tracing event. At first, in the tracing event we consider all previous contacts since the infection of the individual or since a defined period. Depending on the probability p, we decide if we recall the particular individual and include it into the tracing. Furthermore, we check the infection history. Therefore, we can separate forward tracing, backward tracing with and without cluster contacts, and full tracing with or without cluster contacts. After the tracing delay defined by the distribution ϕ, which we often set to the Dirac-Delta-Distribution $\phi = \delta_T$, we check the status of the particular individual. If the individual is infected, we initialize an observed recovery. If we use recursive tracing, we initialize a new tracing event.

---

[1] We describe the simulation and its implementation in more detail in section 6.1.1.

1.5 Backward Tracing

Our analysis of the tracing concentrates on the probability of being infective after time a after infection. Without tracing, this probability depends on the unobserved recovery rate α and observed recovery rate σ. Concentrating on an individual, the differential equation $\dot{\kappa}(a) = -(\alpha + \sigma)\kappa(a)$ describes the decreasing probability of being infective. At the time of infection $a = 0$ the individual is infected, hence $\kappa(0) = 1$. Using this boundary value, the differential equation evaluates to $\kappa(a) = e^{-(\alpha+\sigma)a}$.

We will use the probability of being infective without contact tracing repeatedly as a reference for the probabilities of being infected while performing contact tracing. Therefore, we introduce the notation of $\hat{\kappa}(a)$.

Notation 8 (The Probability of Being Infective without Contact Tracing $\hat{\kappa}(a)$) *The probability of being infective after time a after infection without contact tracing is denoted by:*

$$\hat{\kappa}(a) = e^{-(\alpha+\sigma)a} \quad \text{for } a \geq 0\text{, and } \hat{\kappa}(a) = 0 \quad \text{for } a < 0.$$

With contact tracing, this probability will be lower. A lower probability of being infective will result in a smaller reproduction number, which indicates a reduced spread of the disease or might result in the disease not invading the population.

In this section, we analyze the impact of backward tracing. We add a subscript minus to the probability of being infective. This $\kappa_-(a)$ describes the probability of being infective under backward tracing. I.e., an infected individual contacts and infects other individuals at a rate β during its lifespan. Each of these infectees can trigger a tracing event tracing its infector. In the case of one-step tracing, the infector is only traced, if the infectee experiences a supervised recovery at its rate σ. In the case of recursive tracing, an infectee further down in the infection tree can trigger a tracing event, and the traced infector will trigger a new event recursively until the tracing reaches the individual we focus on. In the progress of this section, we will derive a first-order approximation for the recursive tracing and for the one-step tracing. The analytical approximation consists of an additive first-order effect, which allows further calculations and general interpretation in opposite to the numerical solution of the integro-differential equations, which describe the recursive and one-step tracing. We will exemplify the tracing approximation for the fixed and exponential delay. The exemplification results in an even more compact expression, which we use to calculate reproduction numbers.

In the end, we discuss the influence of the different delays and variance on the contact tracing before we progress to analyze the forward tracing in the succeeding section.

The tracing incorporates a tracing delay. Therefore, we introduce the tracing distribution ϕ.

Definition 9 ϕ denotes the distribution of the tracing delay D. In the case of a fixed delay distribution $\phi = \delta_T$, we will define the fixed delay by the constant T.

Recursive Backward Tracing   Using notation 5 to describe cumulative effects the following proposition 10 will characterize $\kappa_-(a)$, the probability of being infective after the age of infection a.

Proposition 10 (Backward Tracing - Recursive) $\kappa_-(a)$ *is given by the following integro-differential equation:*

$$\kappa'_-(a) = -\kappa(a)\{\alpha + \sigma + p\beta[(\phi * (1 - \kappa_-))(a) - \alpha(\phi * \kappa_-^{\#}(a))]\}, \quad \kappa_-(0) = 1.$$

Proof: We rewrite the equation to denote the total hazard rate, and show the abstract structure. The right-hand side of the resulting differential equation consists of all contributors to the recovery of an infected individual. An infected individual recovers by spontaneous recovery, observed recovery, and in this case by being involved in a backward tracing event, which we summarize by 'rate of tracing'(a),

$$\frac{-\kappa'_-(a)}{\kappa_-(a)} = \{\alpha + \sigma + \text{rate of tracing}(a)\}.$$

The total hazard rate consists of spontaneous recovery α, the observed recovery $\sigma = p_{\text{obs}}\gamma$, which sum up to the total recovery rate γ, and the rate of tracing. In the case of recursive backward tracing, we trigger a tracing event by observed recoveries and by tracing events. At this stage, we do not have a direct definition for the 'rate of tracing', but we can define it in an indirect manner. To characterize the rate of tracing, we calculate the rate of direct and indirect detection, which means observed recovery and recovery by being traced. We calculate the rate by subtracting the known spontaneous recovery rate from the global removal rate:

$$\frac{-\kappa'_-(a)}{\kappa_-(a)} - \alpha = \{\sigma + \text{rate of tracing}(a)\}.$$

Using this rate of observing infected individuals, we derive a formula describing the recursive process of being traced from detected descendants. This formula bases on the descendants of an individual. These descendants are created at rate β, the contact rate. The descendant is observed at the previously calculated rate at c time units after being infected. $\kappa_-(c)$ is the probability of the descendant still being infectious at time c. The integral calculates the rate of observing descendants within the interval $c \in [0, \tau)$ of being infected per infector.

$$\text{descendants observed}(\tau) = \int_0^\tau \beta \left(\frac{-\kappa'_-(c)}{\kappa_-(c)} - \alpha \right) \kappa_-(c) \, dc.$$

In other words, this term indicates the rate of observing descendants, which have been infected τ timesteps before. We now include the probability p of successfully being traced and the tracing delay. The tracing delay is denoted by the distribution $\phi(t)$. Hence, the tracing rate reads:

$$\text{rate of tracing}(a) = p(\phi * \text{descendants observed})(a)$$

$$= p \int_0^a \phi(a - \tau) \int_0^\tau \beta \left(\frac{-\kappa'_-(c)}{\kappa_-(c)} - \alpha \right) \kappa_-(c) \, dc \, d\tau$$

$$= p\beta \int_0^a \phi(a - \tau) \int_0^\tau -\kappa'_-(c) dc - \alpha \int_0^\tau \kappa_-(c) \, dc \, d\tau$$

Since $\kappa_-(0) = 1$ we obtain.

$$= p\beta \int_0^a \phi(a - \tau) \left(1 - \kappa_-(\tau) - \alpha \kappa_-^\#(\tau) \right) d\tau$$

$$= p\beta \left[(\phi * 1)(a) - (\phi * \kappa_-)(a) - \alpha(\phi * \kappa_-^\#)(a) \right].$$

Inserting the rate of tracing into the formula for the total removal rate results in the integro-differential equation stated above.   □

We can solve the integro-differential equation numerically and plot the results. The results shown in fig. 1.7 visualize the exponential decay of $\hat{\kappa}(a)$ and the graphs for $\kappa_-(a)$ corresponding to the backward tracing incorporating various values of fixed delays T. The longer the delay T, the lighter the color. At the corresponding time T, the $\kappa_-(a)$ branch away from $\hat{\kappa}$ and the backward tracing reduces the probability of being infective. The longer the delay, the smaller the absolute impact on the overall probability to be infective over time, because the probability of being infective already decayed.
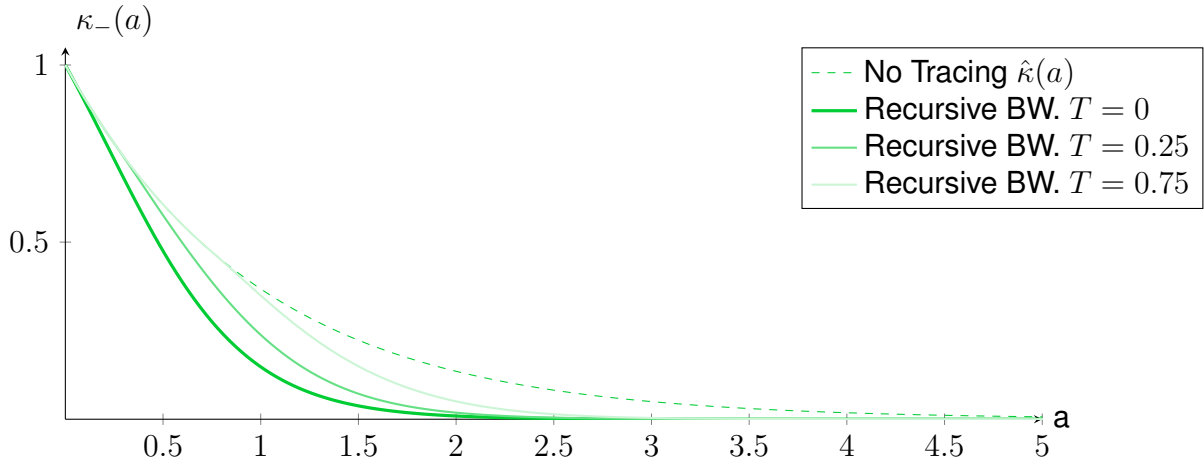
Figure 1.7: The probability $\kappa_-(a)$ of being infective including recursive backward contact tracing. The lighter colors indicate a longer tracing delay, and they show a more delayed impact of the tracing. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

Figure 1.8 shows the probability of being infective relative to the non-tracing case , $\frac{\kappa_-(a)}{\hat{\kappa}(a)}$. We will call $\frac{\kappa_-(a)}{\hat{\kappa}(a)}$ the relative representation. Dividing $\kappa_-(a)$ by $\hat{\kappa}(a)$ and therefore showing the relative gain of $\kappa_-(a)$ compared to $\hat{\kappa}(a)$ allows for an undistracted comparison of the tracing effects prohibiting the exponential decay from distorting the effects and shrinking the effects until they become indistinguishable.
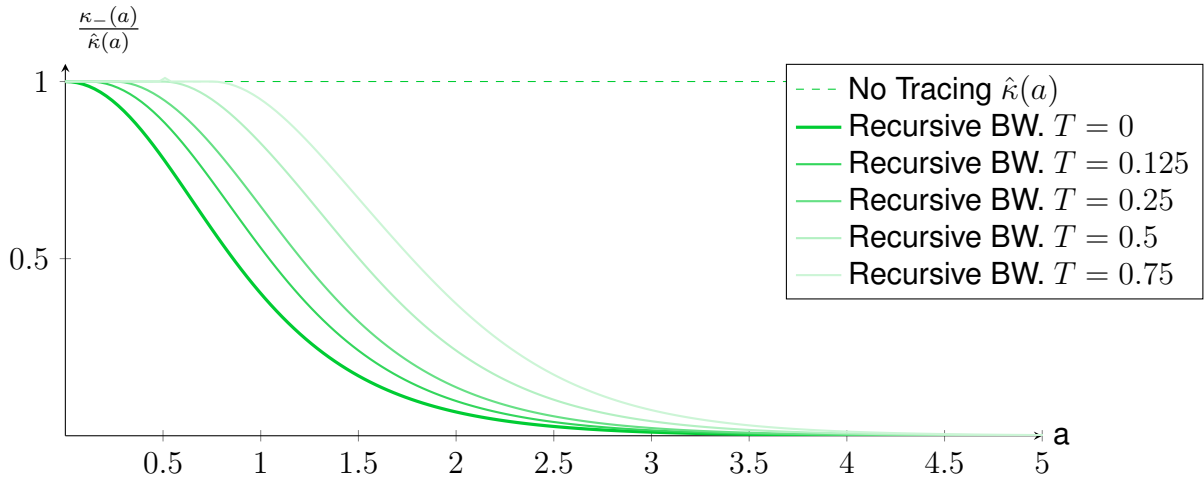

Figure 1.8: The relative probability $\frac{\kappa_-(a)}{\hat{\kappa}(a)}$ of being infective including recursive backward contact tracing compared to the no-tracing case. The lighter colors indicate a longer tracing delay, and they show a more delayed impact of the tracing. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

The relative representation in fig. 1.8 shows the similar-looking appearance of the different

$\kappa_-(a)$ apart from the shift induced by the different delays T. We just keep in mind, their relative impact acts on the exponential function of $\hat{\kappa}(a)$, which shrinks exponentially as time progresses. Thus, the absolute impact of a relative divergence shrinks exponentially.

One-Step Backward Tracing   We follow the previous thoughts about backward tracing again, but this time we will limit the tracing to one-step backward tracing.

One-step backward tracing does not start a tracing recursion to trace the infector of the infector. The absence of the recursion allows for a more direct definition of the tracing, which observed recoveries of infectees trigger. Without the recursion, the tracing algorithm does not see the full infection tree, but only the infector and infectees. They represent the tree within the local scope.

Proposition 11 (One-Step Backward Tracing) *Let $\kappa_-(a)$ denote the probability of being infectious after the time of infection a in case of one-step backward tracing. Then it can be described by the differential equation:*

$$\kappa'_-(a) = -\kappa_-(a)\left\{\alpha + \sigma + p\beta\sigma(\phi * \kappa_-^{\#})(a)\right\}, \quad \kappa_-(0) = 1.$$

Proof: Three mechanics establish the total recovery rate: Spontaneous (unobserved) recovery with rate α, the observed recovery σ, and it consists of indirect observed recovery induced by a contact tracing event.

$$\kappa'_-(a) = -\kappa_-(a)\{\alpha + \sigma + \text{rate of tracing}(a)\}.$$

In one-step tracing, only the direct observation of infectees at rate σ can result in a contact tracing event for the infector. I.e., there is no further recursion involved. The infectees are created at rate β. The contact is recalled at probability p, and there is a tracing delay ϕ.

$$\text{rate of tracing}(a) = p\int_0^a \phi(a-\tau)\int_0^\tau \beta\sigma\kappa_-(c)dc\,d\tau$$
$$= p\beta\sigma\int_0^a \phi(a-\tau)\kappa_-^{\#}(\tau)d\tau$$

Inserting the 'rate of tracing(a)' into $\kappa_-(a)$ proves the assumption.   □

It is apparent that one-step tracing is a lower bound for the impact of recursive tracing because the latter includes the direct tracing by infectees plus the recursive tracing. Figure 1.9 shows the difference in a linear plot. However, the difference is hard to observe. The backward tracing experiences an inherent delay, because the infectees have to be infected to

exist and to be able to start a tracing event. For recursive backward tracing, the infectee of an infectee has to experience an observed recovery. When the recursive tracing impacts the probability of being infective, the attenuation by $\hat{\kappa}(a)$ obstructs the effect.
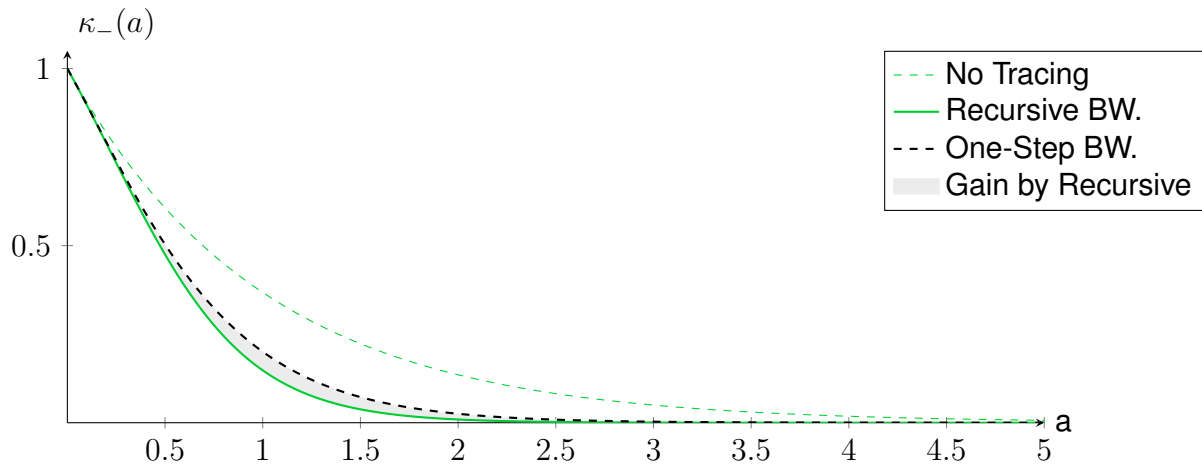

Figure 1.9: This figure compares the tracing impact $\kappa_-(a)$ of recursive and one-step tracing for $T = 0$, $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

If we investigate the gain by recursive tracing, taking a look at the relative effect at a tracing delay $T = 0$ and a probability to recall contacts of $p = 1$ shown in fig. 1.10, we observe respectable gains spanning significant periods. However, the period is located at times a after the start of the infection already experiencing significant attenuation.
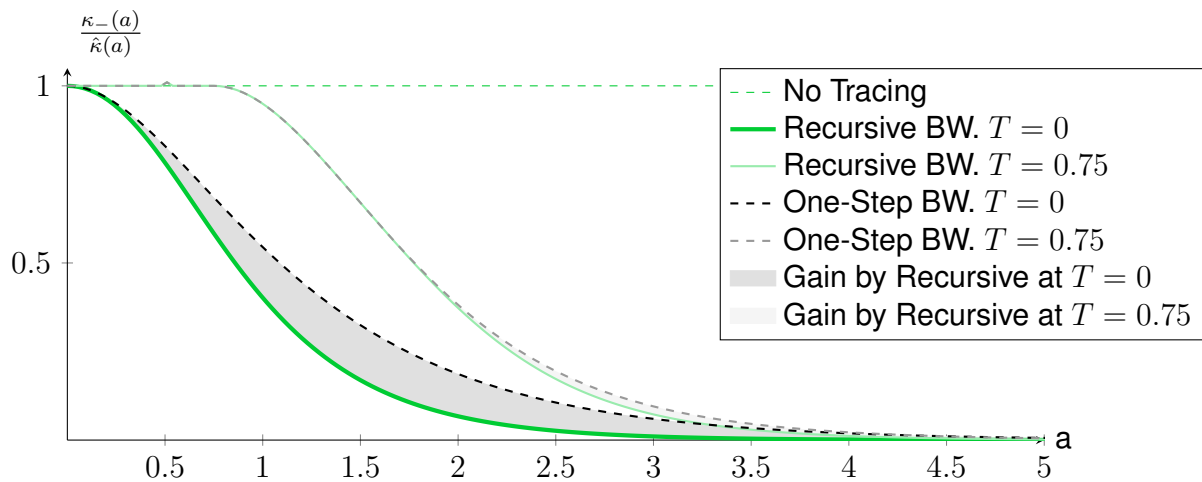

Figure 1.10: This figure compares the relative tracing impact $\frac{\kappa_-(a)}{\hat{\kappa}(a)}$ of recursive and one-step tracing for $T = 0$ and $T = 0.75$. The parameters used are $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

After observing the relative gain of recursive tracing, which results in an existing but not

huge absolute gain at perfect conditions of $T = 0$ and $p = 1$, we take a look at the difference between one-step tracing and recursive tracing at a delay of $T = 0.75$, shown in the same fig. 1.10. This time, we observe only a minor difference even in the relative plot. Recursive tracing experiences the delay multiple times.
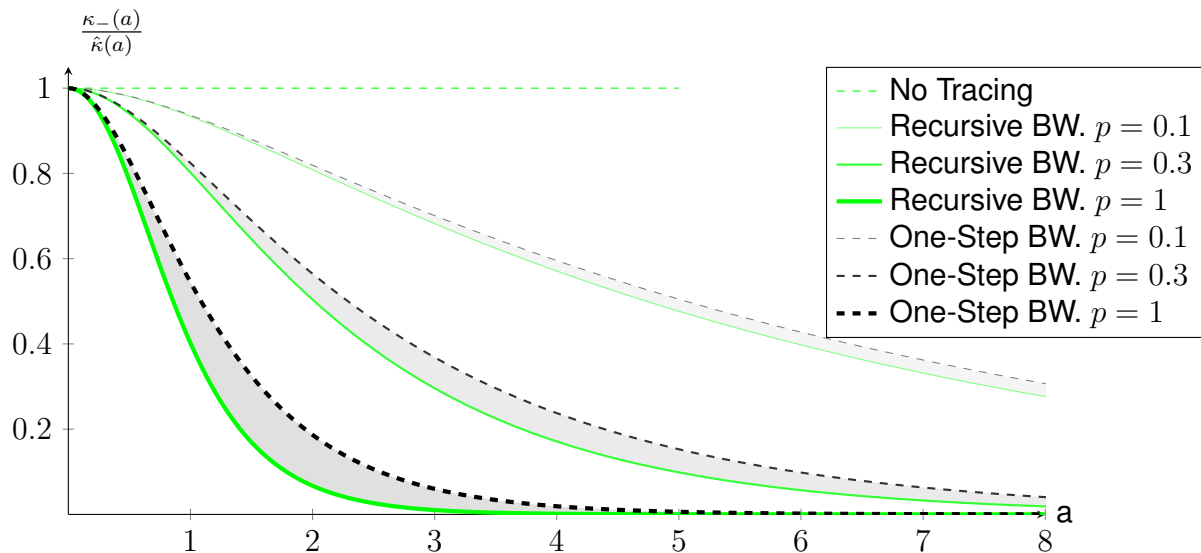

Figure 1.11: This figure compares the relative tracing impact $\frac{\kappa_-(a)}{\hat{\kappa}(a)}$ of recursive and one-step tracing for $p = 0.1$, $p = 0.3$ and $p = 1$. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $T = 0$.

We observe a similar effect when varying the probability of recalling a contact p, which we show in fig. 1.11. Every recursion step involves a new probability p of recalling a contact, and multiplying realistic values of around $p = 0.3$ reduces the effect of the recursion fast. Although, we do not experience the interaction between the delay and attenuation of $\hat{\kappa}$, which seems to increase the effect of the delay.
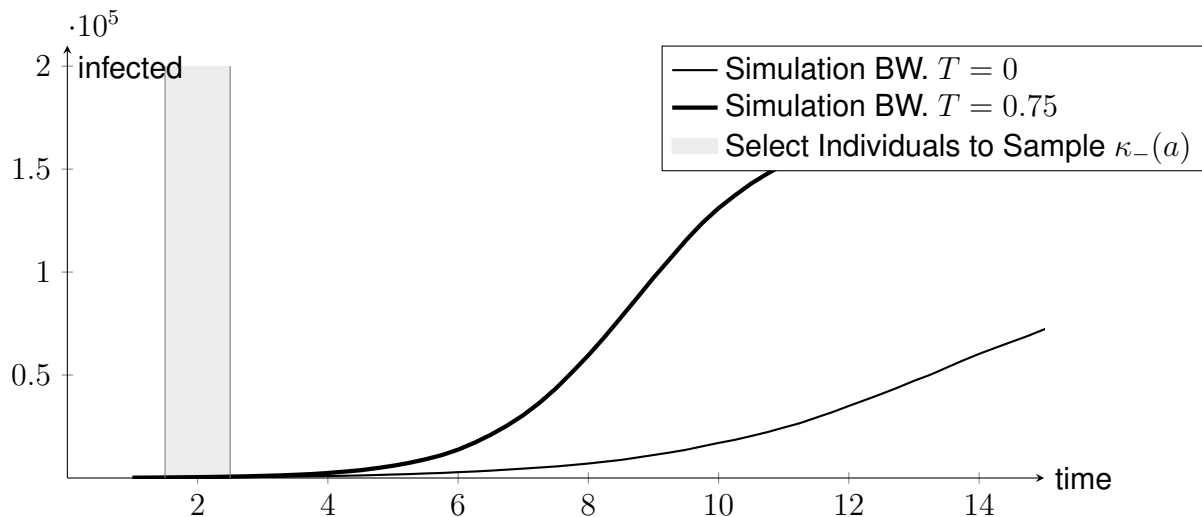
Figure 1.12: This figure shows the count of infected individuals among two populations within an IBM simulation with one-step backward tracing using the tracing delay of $T = 0$ and $T = 0.75$. The gray area visualizes the period used to sample individuals to determine the corresponding $\kappa_-(a)$. The simulations use a population of size $4 \cdot 10^5$ with one-step backward tracing. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

Simulation IBM simulations[1] support the results of the model. We choose one-step tracing using $T = 0$ and $T = 0.75$ as two examples. Figure 1.12 shows these two populations of $4 \cdot 10^5$ individuals and highlights the epoch we use to sample infected individuals to estimate the corresponding $\kappa_-(a)$.

Figure 1.13 shows the matching shape of the stochastic branching model and the IBM simulation apart from stochastic variance.
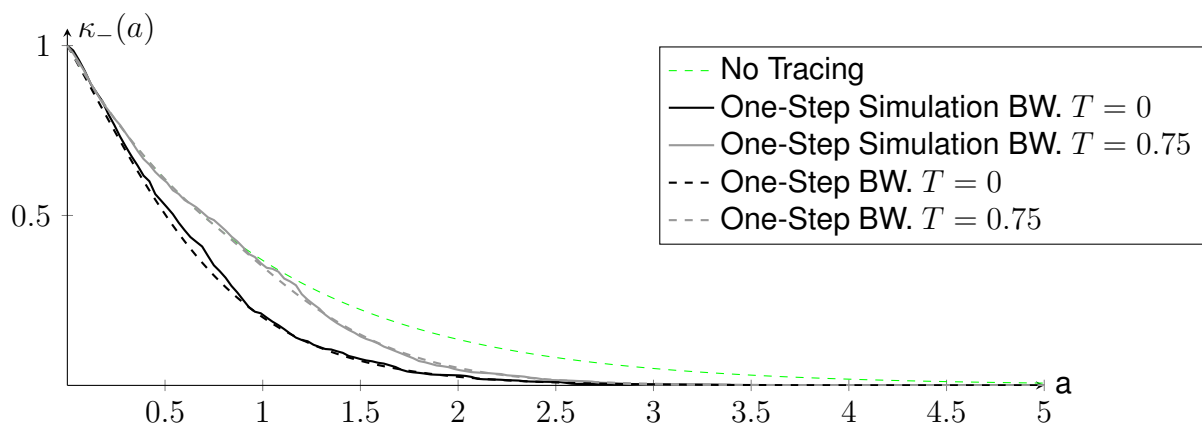

Figure 1.13: This figure compares the tracing impact $\kappa_-(a)$ of the simulation shown in fig. 1.12 and the one-step tracing model for $T = 0$ and $T = 0.75$. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

---

[1]Section 6.1.1 describes the implementation of the IBM simulation in more detail.

1.5.1 Approximations

As the integro-differential equation of recursive backward tracing from proposition 10 can only be solved numerically, we will provide an analytical approximation in proposition 12. The zeroth order of this approximation consists of the removal rate without contact tracing $\hat{\kappa}(a)$ from notation 8.

We assume the probability p, which states the probability of recalling a specific contact, to be small. This small probability allows us to perform a power series ansatz to approximate proposition 10.

Proposition 12 (Backward Tracing – First Order) *The first-order approximation of $\kappa_-(a)$ in p:*

$$\kappa_-(a) = \hat{\kappa}(a) - p\, p_{obs}\beta\hat{\kappa}(a)(1 * \phi * (1 - \hat{\kappa}))(a) + O(p^2).$$

Proof: Assuming p to be small, we start using a power series ansatz with respect to p.

$$\kappa_-(a) = \sum_{i=0}^{\infty} p^i \kappa_{-,i}(a).$$

where $\kappa_i(a)$ are independent of p. Then we insert the ansatz into proposition 10:

$$\kappa_-(0) = 1,$$

$$\sum_{i=0}^{\infty} p^i \kappa'_{-,i}(a) = -\left(\sum_{i=0}^{\infty} p^i \kappa_{-,i}(a) \right) \cdot$$
$$\left\{ \alpha + \sigma + p\beta \left[\left(\phi * \left(1 - \sum_{i=0}^{\infty} p^i \kappa_{-,i} \right) \right)(a) - \alpha \left(\sum_{i=0}^{\infty} p^i \phi * \kappa^{\#}_{-,i} \right)(a) \right] \right\}.$$

Equating powers of p results for p^0 and p^1 in:

$$\kappa'_{-,0}(a) = -\kappa_{-,0}(a)(\alpha + \sigma),$$
$$\kappa_{-,0}(0) = 1,$$
$$\kappa'_{-,1}(a) = -\kappa_{-,1}(a)(\alpha + \sigma) - \beta\kappa_{-,0}(a) \left[(\phi * 1)(a) - (\phi * \kappa_{-,0}(a) - \alpha(\phi * \kappa^{\#}_{-,0})(a) \right],$$
$$\kappa_{-,1}(0) = 0.$$

Solving the zeroth order results in an exponentially decaying function:

$$\hat{\kappa}(a) := \kappa_{-,0}(a) = e^{-(\alpha+\sigma)a}.$$

With $\kappa^{\#}_{-,0}(a) = \frac{1}{\alpha+\sigma}(1 - \hat{\kappa}(a))$ we reformulate the brackets of the first order:

$$\left[(\phi * 1)(a) - (\phi * \kappa_{-,0})(a) - \alpha(\phi * \kappa^{\#}_{-,0})(a)\right]$$
$$= (\phi * (1 - \hat{\kappa}))(a) - \frac{\alpha}{\alpha + \sigma}(\phi * (1 - \hat{\kappa}))(a)$$
$$= \frac{\sigma}{\alpha + \sigma}(\phi * (1 - \hat{\kappa}))(a).$$

This results in the simplified equation for the first order:

$$\kappa'_{-,1}(a) = -(\alpha + \sigma)\kappa_{-,1}(a) - \beta\kappa_{-,0}(a)\left[\frac{\sigma}{\alpha + \sigma}(\phi * (1 - \hat{\kappa}))(a)\right].$$

Recall $p_{\text{obs}} = \frac{\sigma}{\alpha+\sigma}$. Therewith, we obtain:

$$\kappa_{-,1}(a) = -e^{-(\alpha+\sigma)a}\int_0^a \left(e^{(\alpha+\sigma)\tau}\right)\frac{\beta\sigma}{\alpha+\sigma}e^{-(\alpha+\sigma)\tau}(\phi * (1 - \hat{\kappa}))(\tau)d\tau$$
$$= -\int_0^a \beta\frac{\sigma}{\alpha+\sigma}e^{-(\alpha+\sigma)a}(\phi * (1 - \hat{\kappa}))(\tau)d\tau$$
$$= -\beta\, p_{\text{obs}}\, \hat{\kappa}(a)(1 * \phi * (1 - \hat{\kappa}))(a).$$

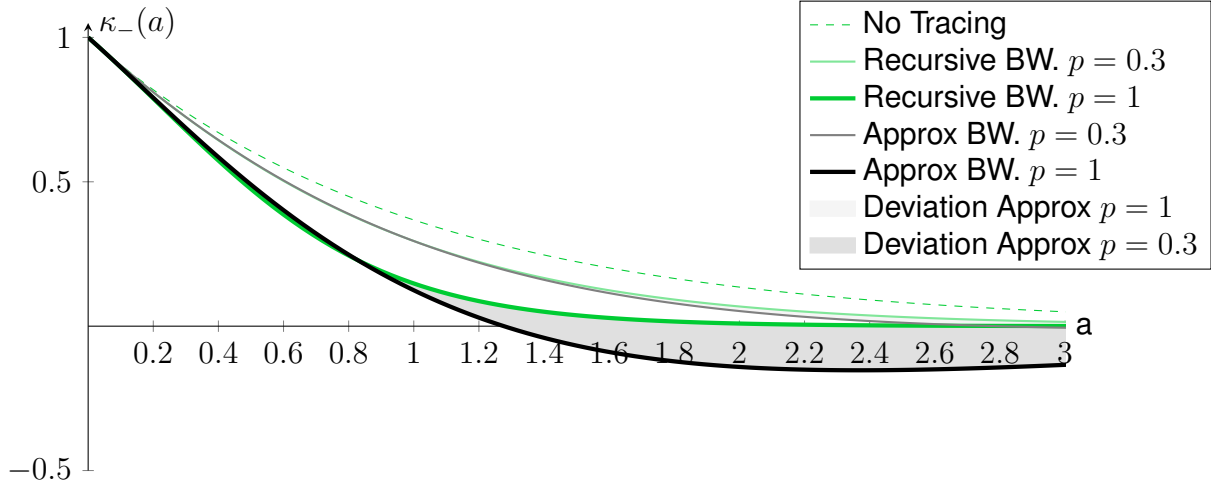Evaluating the first two orders of the ansatz proves the assumption.   □

Figure 1.14: This figure compares the tracing impact $\kappa_-(a)$ of recursive tracing and its approximation for $p = 0.3$ and $p = 1$. The linear plot shows that die relative difference in the case of small values of p happens at a very small scale. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $T = 0$.

Figure 1.14 shows the approximation of recursive backward tracing and the recursive backward tracing. In the case of $p = 1$, there is a visible deviation, but for smaller values of p

like $p = 0.3$, the approximation around p achieves a good match.

We now calculate the first-order approximation of one-step backward tracing, similar to the first-order approximation of backward tracing.

Proposition 13 (First Order Approximation of One-Step Backward Tracing)

$$\kappa_-(a) = \hat{\kappa}(a) - p\, p_{obs}\beta\hat{\kappa}(a)(1 * \phi * (1 - \hat{\kappa})(a) + \mathcal{O}(p^2).$$

Proof: The proof follows the ansatz of proposition 12. We start with a power series.

$$\kappa_-(a) = \sum_{i=0}^{\infty} p^i \kappa_{-,i}(a).$$

We insert this ansatz into proposition 11.

$$\kappa'_-(a) = -\kappa_-(a)\left\{\alpha + \sigma + p\beta\sigma(\phi * \kappa_-^{\#})(a)\right\}, \quad \kappa_-(0) = 1$$

$$\Rightarrow \kappa_-(0) = 1, \quad \sum_{i=0}^{\infty} p^i \kappa_{-,i}(0) = 1,$$

$$\sum_{i=0}^{\infty} p^i \kappa'_{-,i}(a) = -\left(\sum_{i=0}^{\infty} p^i \kappa_{-,i}(a)\right)\left\{\alpha + \sigma + p\beta\sigma\left(\sum_{i=0}^{\infty} p^i \phi * \kappa_{-,i}^{\#}\right)(a)\right\}.$$

Separating the powers 0 and 1 results in the following equations:

$$\kappa'_{-,0}(a) = -\kappa_{-,0}(a)(\alpha + \sigma),$$
$$\kappa_{-,0}(0) = 1,$$
$$\kappa'_{-,1}(a) = -(\alpha + \sigma)\kappa_{-,1}(a) - \beta\kappa_{-,0}(a)\left[\sigma(\phi * \kappa_{-,0}^{\#})(a)\right],$$
$$\kappa_{-,1}(0) = 0.$$

Again we achieve $\kappa_{-,0}(a) = \hat{\kappa}(a) = e^{-(\alpha+\sigma)a}$ and simplify the square brackets:

$$\left[\sigma(\phi * \kappa_{-,0}^{\#})(a)\right] = \frac{\sigma}{\sigma + \alpha}(\phi * (1 - \hat{\kappa})(a)).$$

The following calculations are identical to proposition 12, and variation of parameters will solve the ordinary differential equation proving the assumption. □
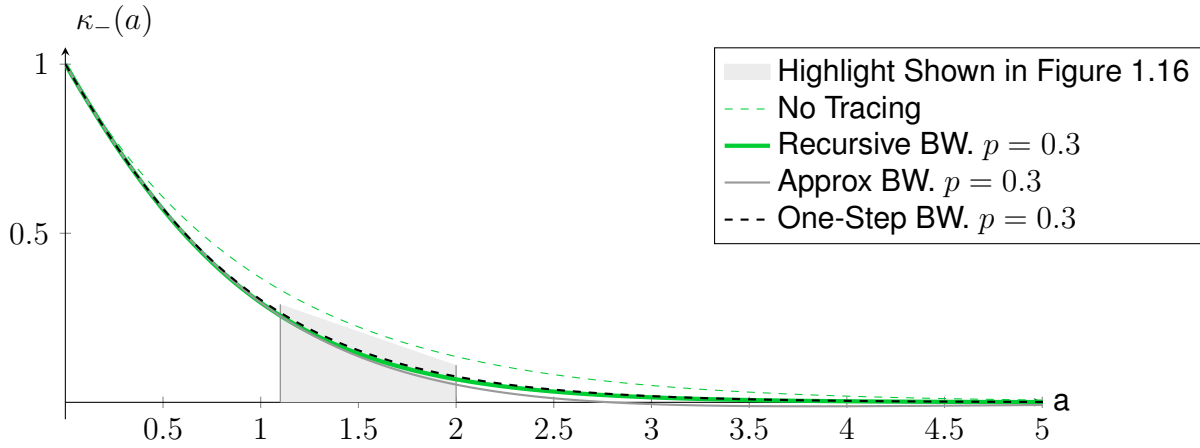
Figure 1.15: This figure compares the tracing impact $\kappa_-(a)$ of one-step tracing and its approximation for $p = 0.3$. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $T = 0$.

In recursive tracing, the probability of tracing a path of length 2 or longer has the order $\mathcal{O}(p^2)$. Thus, the first-order approximation of the backward tracing and the first-order approximation of one-step backward tracing coincide. Figure 1.15 shows the match of recursive and one-step backward tracing compared to their approximation. Figure 1.16 shows that the approximation error and the impact of recursive tracing have the same sign. Thus, the approximation provides even slightly better results in the case of recursive tracing, a coincidence of no further importance.

The fact that we can approximate recursive and one-step backward tracing using the same approximation becomes more comprehensible when recalling figs. 1.10 and 1.11. Figure 1.11 shows recursive and one-step backward tracing using different values of p, visualizing that having to recall contacts multiple times multiplies the probability and results in small differences for probabilities like $p = 0.3$ or $p = 0.1$. Figure 1.10 shows an even stronger effect when comparing the fixed delay $T = 0$ and $T = 0.75$. In the case of a tracing delay, the recovery rates α and σ amplify the effect. The impact of the tracing is shifted to epochs that cover probabilities converging to zero.
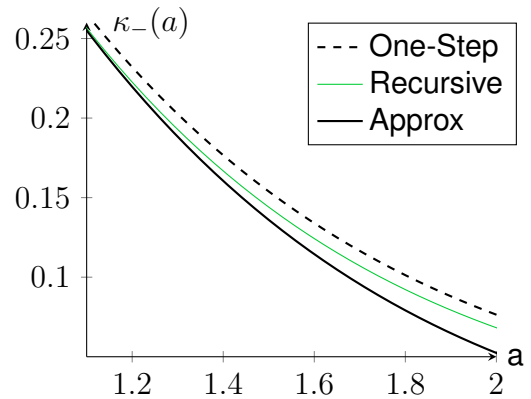

Figure 1.16: This figure magnifies the differences between the one-step and recursive backward tracing and the corresponding approximation, which are shown in fig. 1.15.

1.5.2 Specific Delays

We exemplify the equation by introducing specific delays. The less general equations are easier to evaluate and reveal more structure when analyzing the resulting reproduction numbers in the next subsection.

We choose the fixed delay due to its simple nature and the exponential delay because of its use in modeling survival functions.

Fixed Delay

Assume we replace the distribution ϕ modeling the tracing delay by a fixed delay T:

$$\phi(a) = \delta_T(a).$$

This distribution allows replacing the convolution by a simple shift, simplifying the calculations. The following remark introduces this exemplification into the first-order approximation of proposition 12.

Proposition 14 (First-Order Approximation for a Fixed Delay)

$$\kappa_-(a) = \begin{cases} \hat{\kappa}(a) - p\, p_{\textit{obs}}\beta\hat{\kappa}(a)\left\{(a-T) - \frac{1-\hat{\kappa}(a-T)}{\alpha+\sigma}\right\} + \mathcal{O}(p^2) & \textit{for } a \geq T \\ e^{-(\alpha+\sigma)a} = \hat{\kappa}(a) & \textit{for } a < T \end{cases}$$

Proof: We recall the first-order approximation of backward tracing stated in proposition 12:

$$\kappa_-(a) = \hat{\kappa}(a) - p\, p_{\textsf{obs}}\beta\hat{\kappa}(a)(1 * \phi * (1 - \hat{\kappa}))(a) + O(p^2).$$

First, we evaluate the convolution integral introducing the delay.

$$\phi * (1 - \hat{\kappa})(a) = \begin{cases} (1 - \hat{\kappa}(a - T)) & \text{for } a > T \\ 0 & \text{for } a \leq T \end{cases}$$

The delay only takes place in the first-order approximation. As the zeroth-order approximation is independent of the delay, we conclude

$$\kappa_-(a) = \hat{\kappa}(a) = e^{-(\alpha+\sigma)a}, \quad \text{for } a < T.$$

After introducing this shift into the tracing, we can compute the first convolution, which calculates the cumulative effect. Therefore, we recall that $\int_0^a f(\tau)dt = (1 * f)(a)$ from notation 5. Then we replace the convolution introducing the delay by the shift we precalculated

before and integrate.

$$(1 * \phi * (1 - \hat{\kappa}))(a) = \int_0^a \phi * (1 - \hat{\kappa})(\tau)d\tau = \int_T^a (1 - \hat{\kappa}(a - T))d\tau$$

$$= (a - T) - \frac{1}{\alpha + \sigma}\left(1 - e^{-(\alpha+\sigma)(a-T)}\right).$$

Using this term in the first-order approximation of $\kappa_-(a)$ for $a \geq T$ results in the assumption:
$\kappa_-(a) = \hat{\kappa}(a) - p\, p_{\text{obs}}\beta\hat{\kappa}(a)\left\{(a - T) - \frac{1}{\alpha+\sigma}(1 - \hat{\kappa}(a - T))\right\} + \mathcal{O}(p^2).$ □

Exponential Delay

Of course, a fixed delay does not take place in reality, but it represents a rough approximation. Other distributions can represent more realistic assumptions of the delay. If it can be assumed that the probability to survive, which means to stay undetected by a search for an identified contact, remains constant over time, the exponential distribution is a direct result. We do not analyze more complex delays like the Weibull distribution, which could model a decreasing probability to find a recalled contact or distributions modeling a non-monotonous shape of the probability to be found, which could respect the time to set up the search and a decreasing interest over time.

The exponential distribution does not model the fixed time to set up the search for an identified contact. However, it models the constant probability of an individual to be found by the contact search of a previously started contact tracing event.

Proposition 15 *Choosing $\phi(a) = \frac{1}{T}e^{-\frac{a}{T}}$ results in:*

$$\kappa_-(a) = \hat{\kappa}(a) - p\, p_{\text{obs}}\beta\hat{\kappa}(a)\left\{a + T\frac{T(\alpha + \sigma)}{1 - T(\sigma + \alpha)}\left(\left(1 - e^{-\frac{a}{T}}\right) - \frac{1 - \hat{\kappa}(a)}{(T(\sigma + \alpha))^2}\right)\right\} + \mathcal{O}(p^2).$$

Proof: To prove the claim we insert the exponential delay into the first order approximation. We start with the evaluation of $1 * \phi * (1 - \hat{\kappa}))(a)$,

$$(1 * \phi * (1 - \hat{\kappa}))(a) = \int_0^a \phi * (1 - \hat{\kappa})(\tau)d\tau =$$

$$= \frac{1}{T}\int_0^a \int_0^\tau e^{-\frac{\tau-\tau'}{T}} \cdot (1 - e^{-(\sigma+\alpha)\tau'})d\tau'd\tau$$

$$= \frac{1}{T}\int_0^a [e^{-\frac{\tau}{T}}Te^{\frac{\tau'}{T}}]_0^\tau - e^{-\frac{\tau}{T}}\int_0^\tau e^{(\frac{1}{T}-(\sigma+\alpha))\tau'}d\tau'd\tau$$

$$= \frac{1}{T}\int_0^a T - e^{-\frac{\tau}{T}}T - e^{-\frac{\tau}{T}}\left[\frac{T}{1 - T(\sigma + \alpha)}e^{(\frac{1}{T}-(\sigma+\alpha))\tau'}\right]_0^\tau d\tau$$

$$= a + e^{-\frac{a}{T}}T - T - \frac{1}{1 - T(\sigma + \alpha)} \int_0^a e^{-\frac{\tau}{T}} \left[e^{(\frac{1}{T} - (\sigma + \alpha))\tau'} \right]_0^\tau d\tau$$

$$= a + e^{-\frac{a}{T}}T - T-$$

$$\frac{1}{1 - T(\sigma + \alpha)} \left(\left[\frac{1}{-(\sigma + \alpha)} e^{(\frac{1}{T} - \frac{1}{T} - (\sigma + \alpha))\tau} \right]_0^a + \left[T e^{\frac{-\tau}{T}} \right]_0^a \right)$$

$$= a + e^{-\frac{a}{T}}T - T + \frac{1}{1 - T(\sigma + \alpha)} \left(T \left(1 - e^{\frac{-a}{T}} \right) - \frac{1 - e^{-(\sigma + \alpha)a}}{(\sigma + \alpha)} \right)$$

$$= a + \frac{T^2(\sigma + \alpha)}{1 - T(\sigma + \alpha)}$$

$$\left(-\frac{1 - T(\sigma + \alpha)}{T(\sigma + \alpha)} \left(1 - e^{\frac{-a}{T}} \right) + \frac{1}{T(\sigma + \alpha)} \left(1 - e^{\frac{-a}{T}} \right) - \frac{1 - e^{-(\sigma + \alpha)a}}{T^2(\sigma + \alpha)^2} \right)$$

$$= a + T \frac{T(\alpha + \sigma)}{1 - T(\sigma + \alpha)} \left((1 - e^{-\frac{a}{T}}) - \frac{1 - \hat{\kappa}(a)}{(T(\sigma + \alpha))^2} \right).$$

Using this term we achieve the representation of the assumption.   □

1.5.3 Reproduction Number

The reproduction number enables us to compare the effect of different tracing strategies and different delays. In this section, we will compare the reproduction number of backward tracing with fixed delay and with an exponential delay.

First, we recall the basic reproduction number from corollary 6, $R_0 = \int_0^\infty \beta\hat{\kappa}(a)da = \left[\frac{-\beta\hat{\kappa}(a)}{(\alpha+\sigma)}\right]_0^\infty = \frac{\beta}{(\alpha+\sigma)}$. Then, we instantiate the effective reproduction number from corollary 7 $R_{\text{ct}} = \int_0^\infty \beta\kappa_-(a)da$ using the first-order approximation of backward tracing from proposition 12.

Corollary 16 *The effective reproduction number in case of backward tracing using the first-order approximation from proposition 12 reads:*

$$R_- = \int_0^\infty \beta\left(\hat{\kappa}(a) - p\,p_{\text{obs}}\beta\hat{\kappa}(a)(1 * \phi * (1 - \hat{\kappa}))(a)\right)da + \mathcal{O}(p^2).$$

Reproduction Number in Case of Fixed Delay Backward Tracing

When using the backward tracing approximation, we can further exemplify the statement in case of a fixed delay $\phi(a) = \delta_T(a)$.

Proposition 17 *The reproduction number in case of fixed delay backward tracing is stated by:*

$$R_{-,f} = \frac{\beta}{\alpha+\sigma} - p\,\beta^2\,p_{obs}\frac{\hat{\kappa}(T)}{2(\alpha+\sigma)^2} + \mathcal{O}(p^2).$$

Proof: First, we recall the first-order approximation of one-step backward tracing from proposition 13 and introduce it into the formula to calculate the reproduction number:

$$R_{-,f}(a) = \int_0^\infty \beta\hat{\kappa}(a)da - \int_0^\infty \beta p\,p_{\text{obs}}\beta\hat{\kappa}(a)(1 * \phi * (1 - \hat{\kappa})(a)da + \mathcal{O}(p^2).$$

$$= \int_0^\infty \beta\hat{\kappa}(a)da - p\,\beta^2\,p_{\text{obs}}\int_0^\infty \hat{\kappa}(a)\left((a - T) - \frac{1 - \hat{\kappa}(a - T)}{\alpha+\sigma}\right)da + O(p^2).$$

Then we concentrate on the backward tracing integral to evaluate:

$$
\int_T^\infty \hat{\kappa}(a)\left((a-T) - \frac{1-\hat{\kappa}(a-T)}{\alpha+\sigma}\right) da
$$

$$
= \int_T^\infty \hat{\kappa}(a) a\, da - \int_T^\infty \kappa(a) T\, da - \left(\int_T^\infty \frac{\hat{\kappa}(a)}{\alpha+\sigma} da - \int_T^\infty \frac{\hat{\kappa}(a)\hat{\kappa}(a-T)}{\alpha+\sigma} da\right)
$$

$$
= \left[\frac{-\hat{\kappa}(a)a}{\alpha+\sigma}\right]_T^\infty - \int_T^\infty \frac{-\hat{\kappa}(a)}{\alpha+\sigma} da - \left[\frac{-\kappa(a)T}{\alpha+\sigma}\right]_T^\infty
$$

$$
\quad - \left(\left[\frac{-\hat{\kappa}(a)}{(\alpha+\sigma)^2}\right]_T^\infty - \left[\frac{-\hat{\kappa}(2\cdot a)}{2(\alpha+\sigma)^2\hat{\kappa}(T)}\right]_T^\infty\right)
$$

$$
= \frac{\hat{\kappa}(T)T}{\alpha+\sigma} - \left[\frac{\hat{\kappa}(a)}{(\alpha+\sigma)^2}\right]_T^\infty - \frac{\kappa(T)T}{\alpha+\sigma} - \left(\left[\frac{-\hat{\kappa}(a)}{(\alpha+\sigma)^2}\right]_T^\infty - \left[\frac{-\hat{\kappa}(2\cdot a)}{2(\alpha+\sigma)^2\hat{\kappa}(T)}\right]_T^\infty\right)
$$

$$
= \left[\frac{-\hat{\kappa}(2\cdot a)}{2(\alpha+\sigma)^2\hat{\kappa}(T)}\right]_T^\infty = \frac{\hat{\kappa}(2\cdot T)}{2(\alpha+\sigma)^2\hat{\kappa}(T)} = \frac{\hat{\kappa}(T)}{2(\alpha+\sigma)^2}.
$$

Using this auxiliary calculation to solve the whole integral results in the statement.   □

Corollary 18 *We can express the reproduction number in the case of fixed delay backward tracing using the basic reproduction number:*

$$
R_{-,f} = \frac{\beta}{\alpha+\sigma} - p\,\beta^2\,p_{obs}\frac{\hat{\kappa}(T)}{2(\alpha+\sigma)^2} + \mathcal{O}(p^2)
$$

$$
= R_0 - p\,p_{obs}R_0^2\frac{\hat{\kappa}(T)}{2} + \mathcal{O}(p^2).
$$

Reproduction Number in Case of Exponential Delay and Backward Tracing

Proposition 19 *The reproduction number in the case of exponential delay backward tracing is stated by:*

$$
R_{-,e} = \frac{\beta}{\alpha+\sigma} - \frac{p\,p_{obs}\beta^2}{(\sigma+\alpha)^2 2(1+T(\alpha+\sigma))}.
$$

Proof: First, we recall the approximation of backward tracing in the case of exponential delay from proposition 15. We introduce this probability into the formula for the effective

reproduction number from corollary 7.

$$R_{-,\mathrm{e}} = \int_0^\infty \beta\hat{\kappa}(a) - \beta p \, p_{\mathrm{obs}}\beta\cdot$$
$$\hat{\kappa}(a)\left\{ a + T\frac{T(\alpha+\sigma)}{1-T(\sigma+\alpha)}\left((1 - e^{-\frac{a}{T}}) - \frac{1-\hat{\kappa}(a)}{(T(\sigma+\alpha))^2} \right) \right\} da + \mathcal{O}(p^2).$$

Then, we expand the term to solve the resulting smaller integrals:

$$R_{-,\mathrm{e}} = \int_0^\infty \beta\hat{\kappa}(a) - \beta p \, p_{\mathrm{obs}}\beta\left\{ \int_0^\infty \hat{\kappa}(a)a\,da \right.$$
$$+ T\frac{T(\alpha+\sigma)}{1-T(\sigma+\alpha)}\left(\left(\int_0^\infty \hat{\kappa}(a)da - \int_0^\infty \hat{\kappa}(a)e^{-\frac{a}{T}}da \right) \right.$$
$$\left. \left. - \int_0^\infty \frac{\hat{\kappa}(a)}{(T(\sigma+\alpha))^2}da + \int_0^\infty \frac{\hat{\kappa}(a)^2}{(T(\sigma+\alpha))^2}da \right) \right\} + \mathcal{O}(p^2).$$

Now, we solve the individual integrals, starting with the known basic reproduction number:

$$\int_0^\infty \beta\hat{\kappa}(a)da = \int_0^\infty e^{-(\alpha+\sigma)a}da = \frac{\beta}{\alpha+\sigma} = R_0, \tag{1.1}$$

$$\int_0^\infty \hat{\kappa}(a)a\,da = \frac{1}{(\alpha+\sigma)^2}, \tag{1.2}$$

$$\int_0^\infty \hat{\kappa}(a)da = \frac{1}{(\alpha+\sigma)}, \tag{1.3}$$

$$\int_0^\infty \hat{\kappa}(a)e^{-\frac{a}{T}}da = \frac{T}{1+(\alpha+\sigma)T}, \tag{1.4}$$

$$\int_0^\infty \frac{\hat{\kappa}(a)}{(T(\sigma+\alpha))^2}da = \frac{1}{(\alpha+\sigma)T^2(\sigma+\alpha)^2}, \tag{1.5}$$

$$\int_0^\infty \frac{\hat{\kappa}(a)^2}{(T(\sigma+\alpha))^2}da = \frac{1}{(2\alpha+\sigma)T^2(\sigma+\alpha)^2}. \tag{1.6}$$

Now, we subtract the solutions of eqs. (1.5) and (1.6), and add the result to eqs. (1.2) to (1.4) while respecting the factor $T\frac{T(\alpha+\sigma)}{1-T(\sigma+\alpha)}$ to evaluate the curly brackets of the reproduction number.

$$\frac{1}{(\alpha+\sigma)^2} + T\frac{T(\alpha+\sigma)}{1-T(\sigma+\alpha)}\left(\frac{1}{\alpha+\sigma} - \frac{T}{1+(\alpha+\sigma)T} - \frac{1}{2(\alpha+\sigma)T^2(\sigma+\alpha)^2} \right)$$
$$= \frac{1}{(\sigma+\alpha)^2 2(1+T(\alpha+\sigma))}$$

We achieve the backward tracing update by adding the factor $p\,p_{\mathsf{obs}}\beta^2$.

$$= \frac{p\,p_{\mathsf{obs}}\beta^2}{(\sigma+\alpha)^2 2(1+T(\alpha+\sigma))}.$$

$\square$

Corollary 20 *We can express the reproduction number in the case of exponential delay backward tracing using the basic reproduction number:*

$$\begin{aligned}
R_{-,e} =& \frac{\beta}{\alpha+\sigma} - \frac{p\,p_{obs}\beta^2}{(\sigma+\alpha)^2 2(1+T(\alpha+\sigma))} + \mathcal{O}(p^2) \\
=& R_0 - \frac{p\,p_{obs}R_0^2}{2(1+T(\alpha+\sigma))} + \mathcal{O}(p^2).
\end{aligned}$$

Comparison of Fixed and Exponential Delay

We choose the representation of the reproduction numbers from corollaries 18 and 20 to highlight the differences,

$$\begin{aligned}
R_{-,\mathsf{f}} =& R_0 - p\,p_{\mathsf{obs}}R_0^2 \frac{\hat{\kappa}(T)}{2} + \mathcal{O}(p^2), \\
R_{-,\mathsf{e}} =& R_0 - p\,p_{\mathsf{obs}}R_0^2 \frac{1}{2(1+T(\alpha+\sigma))} + \mathcal{O}(p^2).
\end{aligned}$$

First, we notice that $\frac{1}{1+T(\alpha+\sigma)}$ is the $[0/1]$ Padé approximation of $\hat{\kappa}(T)$ [Pad92, p.14][BG96, p.8]:

$$\begin{aligned}
\hat{\kappa}(T) =& e^{-(\alpha+\sigma)T}, \\
e^{-(\alpha+\sigma)T} + \mathcal{O}(T^2) =& [1/0]_{\hat{\kappa}}(T) = \frac{1}{1+T(\sigma+\alpha)}.
\end{aligned}$$

Apart from the $\hat{\kappa}(T)$ and its approximation, the approximated reproduction numbers are similar. Figure 1.17 visualizes that the effect of exponential delay tracing is bigger than the effect of fixed delay tracing for the practical levels of α, σ and T. We know that $a \geq 0$, $\sigma \geq 0$, and $T \geq 0$. Furthermore, we observed large recovery rates require very small delays to not prohibit an effect of tracing to become visible. On the other hand, long delays require very
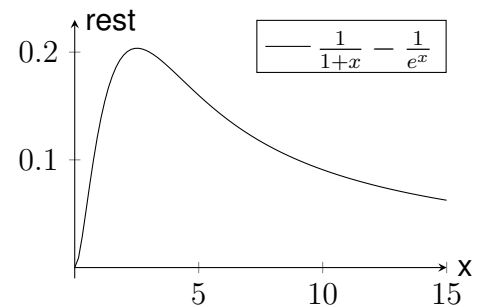

Figure 1.17: Differences of exponential delay and fixed delay compared.

small recovery rates, because otherwise, the probabilities will reach similar low levels before the contact tracing can take effect.

It is easy to see that $\frac{1}{1+x} \geq e^{-x}$ for $x \geq 0$ because the denominators show $e^x \geq 1+x \quad \forall x \geq 0$. This observation raises the assumption that the higher variance of the exponential delay might be an advantage.

Therefore, we imagine a distribution consisting of two fixed delays $\frac{\delta_{T+s}+\delta_{T-s}}{2}$ for $T - x > 0, s > 0$. We observe

$$R_0 - p\, p_{\mathsf{obs}} R_0^2 \frac{\hat{\kappa}(T)}{2} \geq R_0 - p\, p_{\mathsf{obs}} R_0^2 \frac{\hat{\kappa}(T+s) + \hat{\kappa}(T-s)}{4}.$$

This is true due to the convexity of e^{-T}

$$\frac{\hat{\kappa}(T)}{2} \leq \frac{\hat{\kappa}(T+s) + \hat{\kappa}(T-s)}{4}.$$
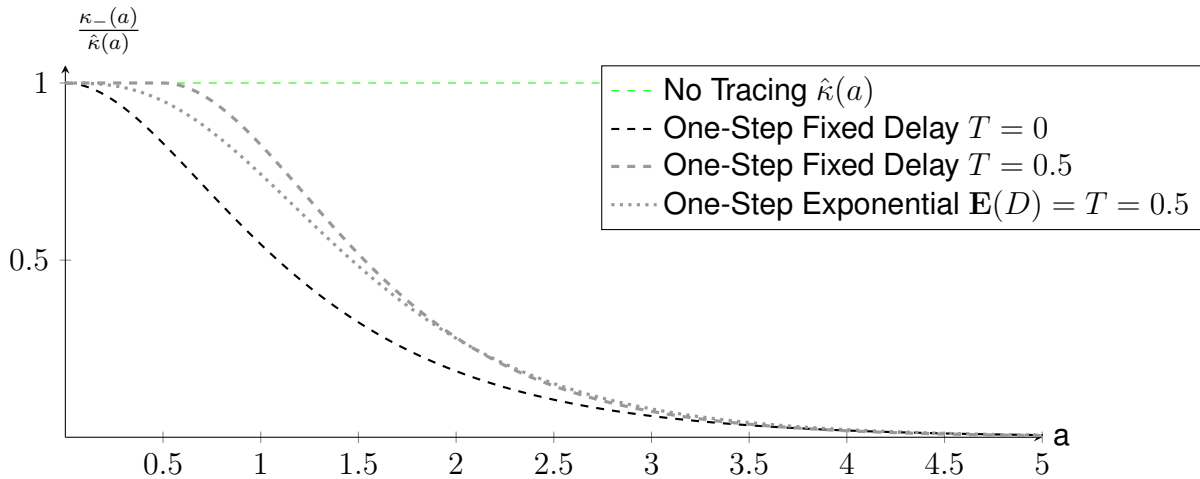

Figure 1.18: The relative probability $\frac{\kappa_-(a)}{\hat{\kappa}(a)}$ of being infective including one-step backward contact tracing without a delay, with fixed delay $T = 0.5$ and with an exponential delay with expected value $\mathbf{E}(D) = 0.5$. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

We can observe the advantage of exponential delay over the fixed delay when looking at the probabilities of being infected after the start of an infection. Figure 1.18 shows that the exponential distribution variance allows for a significant impact during the beginning of an infection when more infective contacts can be prohibited.

The IBM simulation shown in fig. 1.20 and fig. 1.19 shows the same results. We choose a larger epoch to sample the individuals for $\kappa_{-,\mathsf{e}}(a)$ compared to the fixed delay from fig. 1.12 to counter the additional variance gained by the exponential distribution.
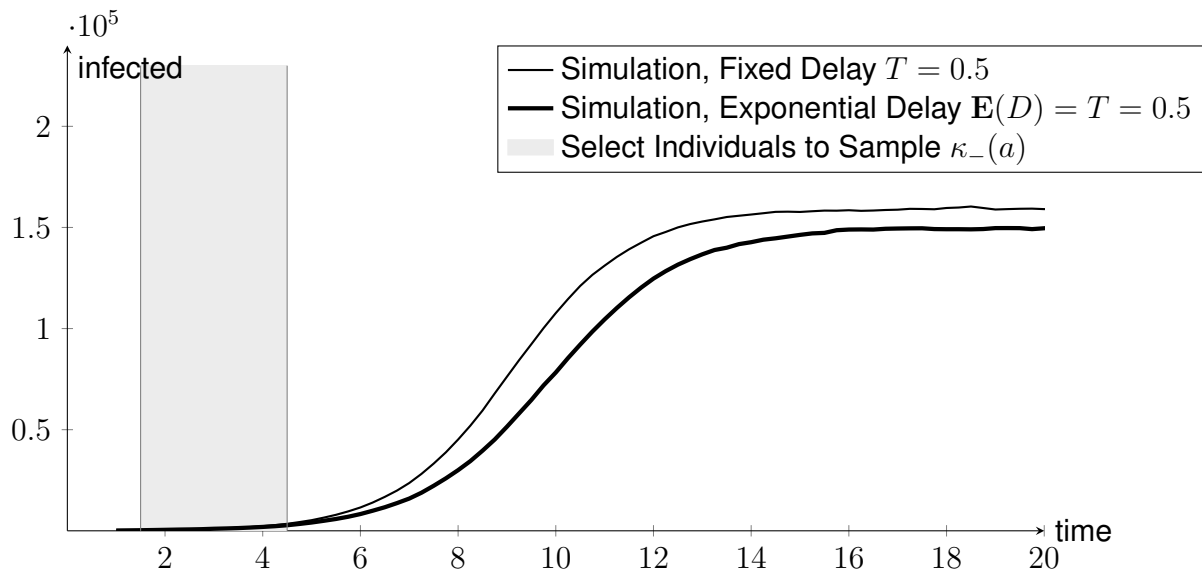
Figure 1.19: This figure shows the count of infected individuals among two populations within an IBM simulation with fixed-delay backward tracing using the fixed tracing delay of $T = 0.5$ and an exponential delay of $\mathbf{E}(D) = 0.5$. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.
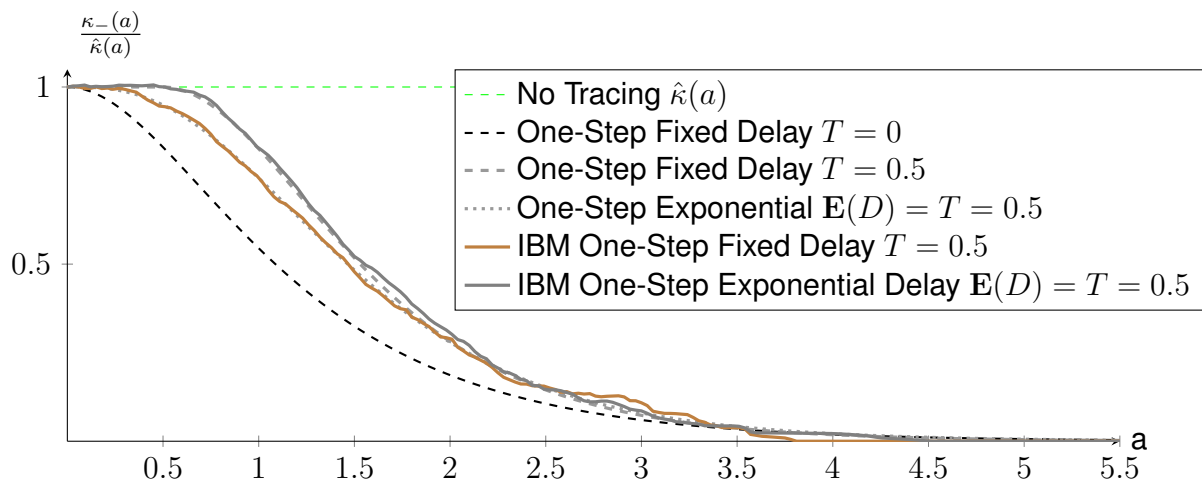

Figure 1.20: The same graphs as shown in fig. 1.18, but we added the two IBM simulations shown in fig. 1.19 including one-step backward tracing with a fixed delay $T = 0.5$ and tracing with an exponential delay $\mathbf{E}(D) = 0.5$.

Discussion – Influence of Variance on the Reproduction Number

The larger impact of exponential delay in backward tracing might be supported by the result of [WL07]. They analyze the spread of disease and summarize that a lower variation in the life expectancy of a generation corresponds to a larger R_0. The exponential delay increases the variation of the life expectance of surviving individuals.

Nevertheless, we observe that the variation of the tracing delay allows for the tracing within the early phase of infections when the probability of being infective is large, and more secondary infections can be prohibited. On the other hand, the slower detections do not balance this advantage because of the spontaneous recovery. If there is a very long tracing delay, there is a realistic chance that the individual already recovered. The convex shape of $\kappa_-(a)$ results in a bigger average infectivity of the traced individuals than a fixed delay would detect. However, the spontaneous recovery rate is a necessity for this argumentation.

1.6 Forward Tracing

In this section, we will model forward tracing by employing a recursive approach aligned by the generations of infected individuals. The recursion starts at the zeroth generation. The zeroth generation can not be traced by an event triggered by a previous generation. Therefore, the probability of being infective of the zeroth generation is $\hat{\kappa} = e^{-(\alpha+\sigma)a}$, see notation 8. To model the following generations of infected individuals, we will introduce a notation $\kappa_i^+(a|b)$.

Definition 21 ($\kappa_i^+(a|b)$) $\kappa_i^+(a|b)$ is the probability for an individual of generation i to be infectious at age since infection a, given that the infector has had the age since infection b at the time of the infectious contact.

Proposition 22 (Recursive Forward Tracing) $\kappa_i^+(a)$ *for $i > 0$ follows from the following recursive formula:*

$$\kappa_{i-1}^+(b)\kappa_i^+(a|b) = \hat{\kappa}(a)\left\{\kappa_{i-1}^+(b) - p\int_0^a\left(-{\kappa_{i-1}^+}'(b+c) - \alpha\kappa_{i-1}^+(b+c)\right)\int_c^a \phi(a'-c)da'dc\right\}$$

$$\kappa_i^+(a) = \frac{\int_0^\infty \kappa_i^+(a|b)\kappa_{i-1}^+(b)db}{\int_0^\infty \kappa_{i-1}^+(\tau)d\tau}.$$

Proof: Without contact tracing, the probability of an individual being infectious at the age a of the infection is given by $\hat{\kappa}(a)$, the probability of the zeroth generation. If the infector is still alive, the chance of tracing lowers the probability of being still infectious. We are now going to prove the probability of being infective under the condition of the status of the infector $\kappa_i(a|b) = \hat{\kappa}(a)\left\{1 - p \text{ rate of tracing}_i(a, b)\right\}$. The rate of tracing$_i(a, b)$ depends on the tracing delay and the age of the infector, given that the infector was infective at the time the infection took place. Below, we will show that

$$\kappa_i(a|b) = \hat{\kappa}(a)\left\{1 - p\int_0^a\left(\frac{-{\kappa_{i-1}^+}'(b+c)}{\kappa_{i-1}^+(b)} - \frac{\alpha\kappa_{i-1}^+(b+c)}{\kappa_{i-1}^+(b)}\right)\int_c^a \phi(a'-c)da'dc\right\}.$$

To be traced as an infectee, the infector needs to become detected. We know that the infection took place at the age of infection b of the infector. Particularly, the infector was infectious at that age. Hence, the probability that the infector is still infectious at age since infection $a + b$ reads $\kappa_{i-1}^+(a + b)/\kappa_{i-1}^+(b)$.

At time $b + c$, the rate of being observed as an infector is the probability of being still infectious multiplied by the rate of being observed, i.e., the removal rate minus the unobserved

removals α.

$$\frac{\kappa_{i-1}^+(b+c)}{\kappa_{i-1}^+(b)}\left(\frac{-\kappa_{i-1}^+{}'(b+c)}{\kappa_{i-1}^+(b+c)}-\alpha\right)=\left(\frac{-\kappa_{i-1}^+{}'(b+c)}{\kappa_{i-1}^+(b)}-\frac{\alpha\kappa_{i-1}^+(b+c)}{\kappa_{i-1}^+(b)}\right).$$

Now we incorporate a delay distributed according to ϕ, and we introduce the time $a' \in [0,a)$ at which a tracing event is triggered. This results in the distributed hazard rate at time a'.

$$\int_0^{a'}\left(\frac{-\kappa_{i-1}^+{}'(b+c)}{\kappa_{i-1}^+(b)}-\frac{\alpha\kappa_{i-1}^+(b+c)}{\kappa_{i-1}^+(b)}\right)\phi(a'-c)dc.$$

The distributed effect of the rate to be traced at a' is now accumulated by integrating over the interval $[0,a)$.

$$\int_0^{a}\int_0^{a'}\left(\frac{-\kappa_{i-1}^+{}'(b+c)}{\kappa_{i-1}^+(b)}-\frac{\alpha\kappa_{i-1}^+(b+c)}{\kappa_{i-1}^+(b)}\right)\phi(a'-c)dcda'$$

$$=\int_0^{a}\left(\frac{-\kappa_{i-1}^+{}'(b+c)}{\kappa_{i-1}^+(b)}-\frac{\alpha\kappa_{i-1}^+(b+c)}{\kappa_{i-1}^+(b)}\right)\int_0^{a}\phi(a'-c)da'dc.$$

Inserting this rate of tracing results in the intermediate equation to prove

$$\kappa_i^+(a|b)=\hat{\kappa}(a)\left\{1-p\int_0^{a}\left(\frac{-\kappa_{i-1}^+{}'(b+c)}{\kappa_{i-1}^+(b)}-\frac{\alpha\kappa_{i-1}^+(b+c)}{\kappa_{i-1}^+(b)}\right)\int_c^{a}\phi(a'-c)da'dc\right\}.$$

This $\kappa_i^+(a|b)$ states the probability of being infective under the condition that the infector is of age $b+a$. To achieve $\kappa_i^+(a)$, we respect all possible ages of the infector b. We start calculating the probability to be infective and the infector, being of a particular age of the infection b, still being infective. Therefore, we multiply the conditional probability by $\kappa_{i-1}^+(b)$. Thus, we achieve the first equation to prove.

To obtain $\kappa_i^+(a)$, we note that the distribution of the age since infection of the infectors at the time of the infection, which is given by

$$\frac{\beta\kappa_{i-1}^+(b)}{\int_0^{\infty}\beta\kappa_{i-1}^+(\tau)d\tau}=\frac{\kappa_{i-1}^+(b)}{\int_0^{\infty}\kappa_{i-1}^+(\tau)d\tau}.$$

Introducing this distribution into $\kappa_{i-1}^+(b)\kappa_i(a|b)$ and integrating all possible ages of the infector b proves the claim. $\square$

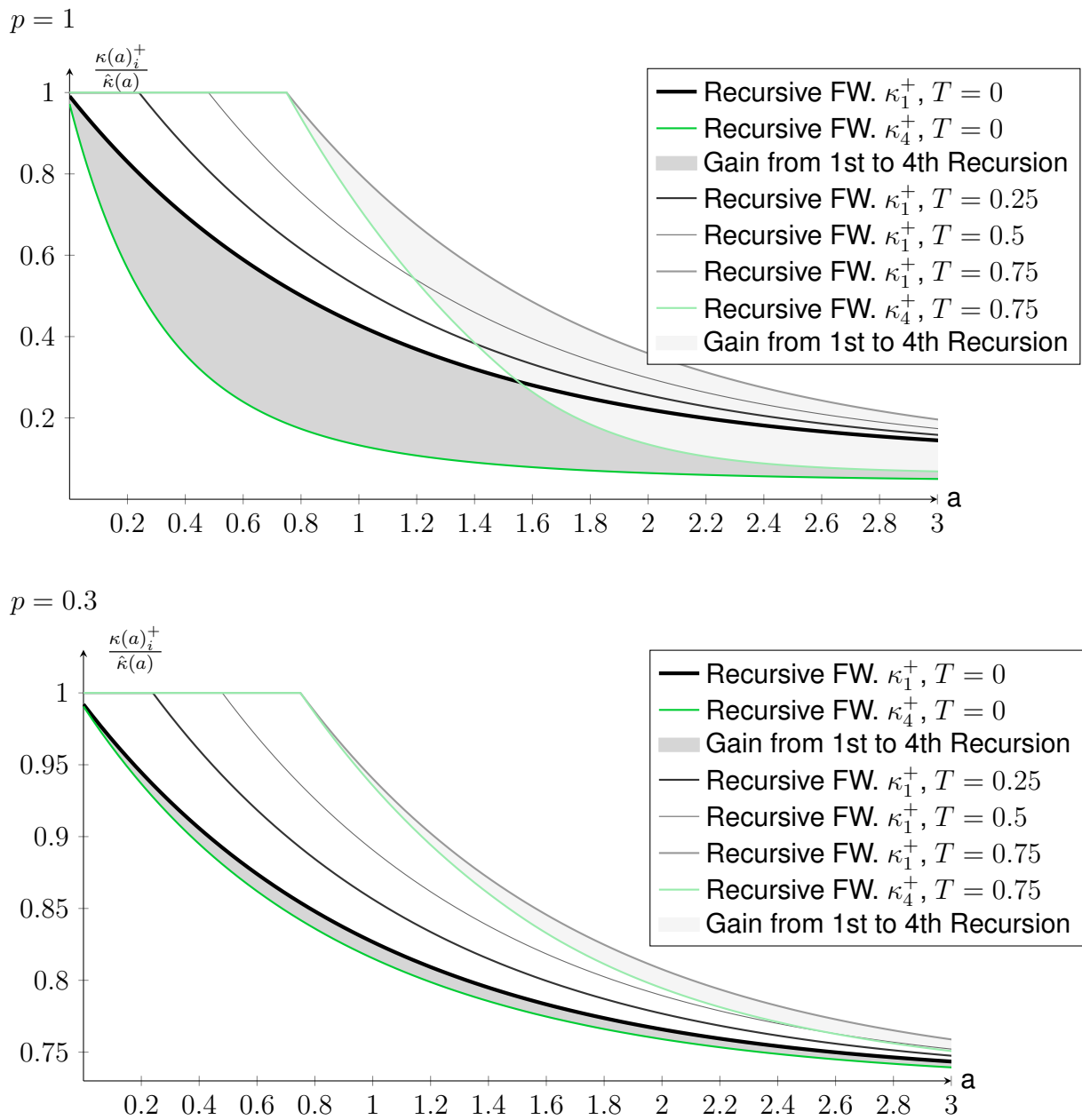Figure 1.21 visualizes the recursive forward tracing with various delays from $T=0$ to

Figure 1.21: The figures show recursive forward tracing with various fixed delays from $T = 0$ to $T = 0.75$ in the case of $p = 1$ and $p = 0.3$. The gray are shows the improvement due to the recursion of the tracing between generation one and four. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

$T = 0.75$ in the case of $p = 1$ and $p = 0.3$. In the case of $T = 0$ and $T = 0.75$, we did not only plot the first generation $\kappa_1^+(a)$, but also the fourth generation $\kappa_4^+(a)$ and highlight the improved impact of the fourth generation's recursive tracing in gray.

We choose to plot the probability of being infective respecting forward tracing $\kappa_i^+(a)$ relative to the no tracing case $\hat{\kappa}(a)$, i.e., $\frac{\kappa_i^+(a)}{\hat{\kappa}(a)}$, to highlight the differences between the generations, which become smaller for lower probabilities of recalling contacts. We show this by lowering the probability of recalling a contact to $p = 0.3$.

Both plots show the start of the tracing after the end of the tracing delay. Furthermore, we can concentrate on the recursive effect and observe that the influence of the recursive effect starts at the same time as the tracing of the first generation. This recursive effect without additional delay differs from the backward tracing case. In the recursive backward tracing, the effect of every additional individual, which was infected and then recovered, involved an additional delay until it influenced the current individual.

In the case of forward tracing, the previous generations' delay, e.g., the delay of the infector of the infector, can progress even before the current individual is infected. This is shown in fig. 1.22. Figure 1.22 shows an infection tree of three individuals of three generations on the left, the zeroth, first, and second generation. On the top, there is a timeline. Below the timeline, the fig. 1.22 shows three graphs showing the κ_i^+ corresponding to each generation. The graphs are arranged according to this timeline. On top of these general graphs, we overlay the interactions of a possible example, which demonstrates how recursive tracing can reach the individual of generation two right after the tracing delay of the infector progressed.

In the case of the first generation, an individual is traced by its infector. In the case of generations $i > 1$ and recursive contact tracing, this infector suffers a higher hazard rate and triggers more tracing events because it is traced by its infector.

One-Step Forward Tracing

In recursive tracing, the direct detection at rate σ and the indirect detection by contact tracing trigger tracing events. In one-step tracing, only direct detection triggers a contact tracing event. Similar to section 1.5, 'One-Step Backward Tracing', we replace the implicit description of observed recoveries, i.e., the total hazard rate minus spontaneous recoveries at rate α, by the direct detection at rate σ.
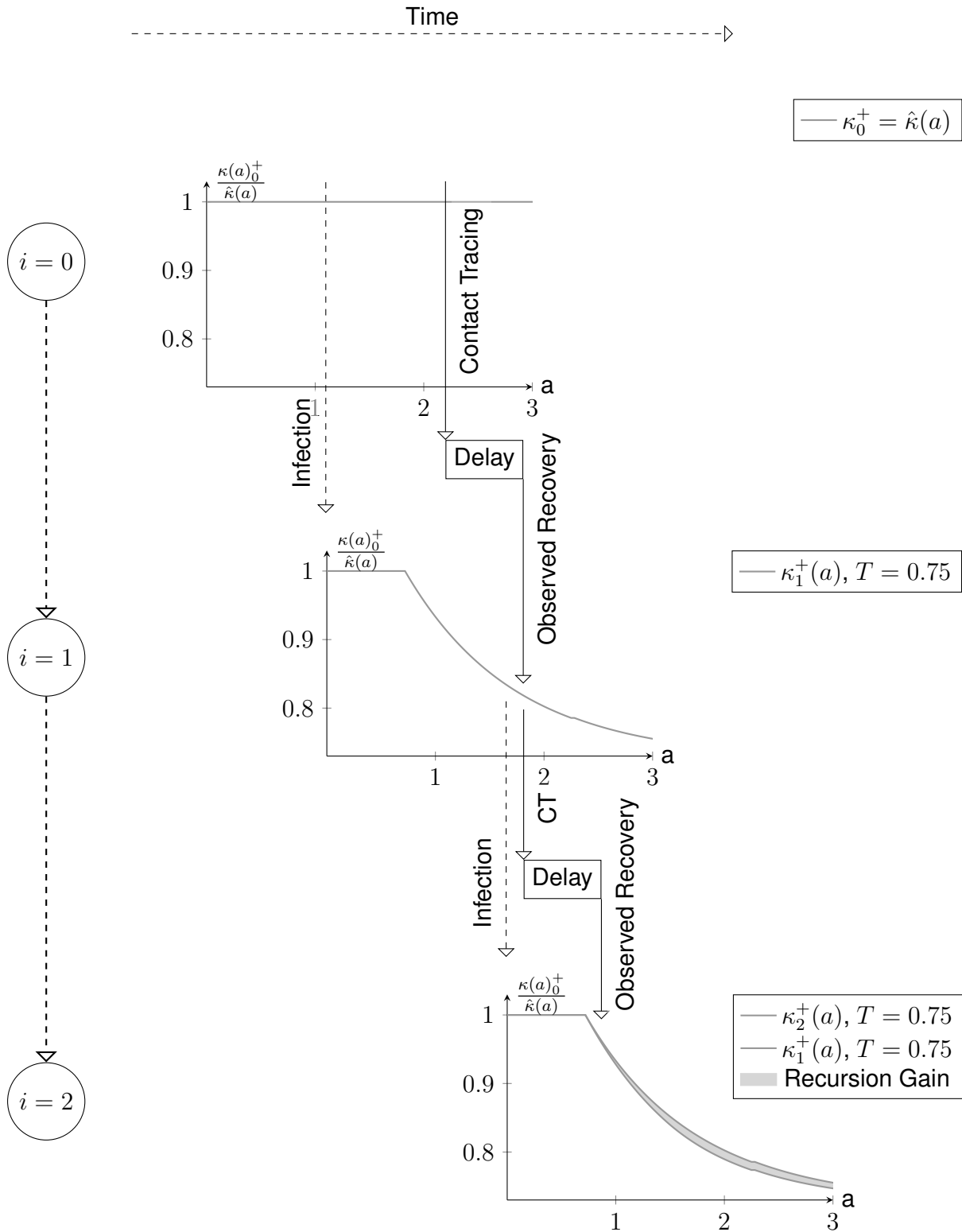
Figure 1.22: We visualize recursive tracing by describing a possible timeline of the interactions between the generations. We relate the interactions to the probabilities of being infective $\frac{\kappa_i^+(a)}{\hat{\kappa}(a)}$ for three generations $i = 0$, $i = 1$ and $i = 2$ by overlaying the interactions on top of the graphs. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 0.3$.

Proposition 23 (One-Step Forward Tracing) *The recursive formula for $\kappa_i^+(a)$ for $i > 0$ in the case of one-step forward tracing is given by:*

$$\kappa_{i-1}^+(b)\kappa_i^+(a|b) = \hat{\kappa}(a)\left\{\kappa_{i-1}^+(b) - p\sigma \int_0^a \kappa_{i-1}^+(b+c)\int_c^a \phi(a'-c)da'dc\right\}$$

$$\kappa_i^+(a) = \frac{\int_0^\infty \kappa_i^+(a|b)\kappa_{i-1}^+(b)db}{\int_0^\infty \kappa_{i-1}^+(\tau)d\tau}.$$

Figure 1.23 shows one-step tracing at $p = 1$ using similar parameters as fig. 1.21. We observe a reduced effect of the tracing for generations $i = 4$ compared to $i = 1$ in contrast to the positive effect we observed for recursive tracing in fig. 1.23.
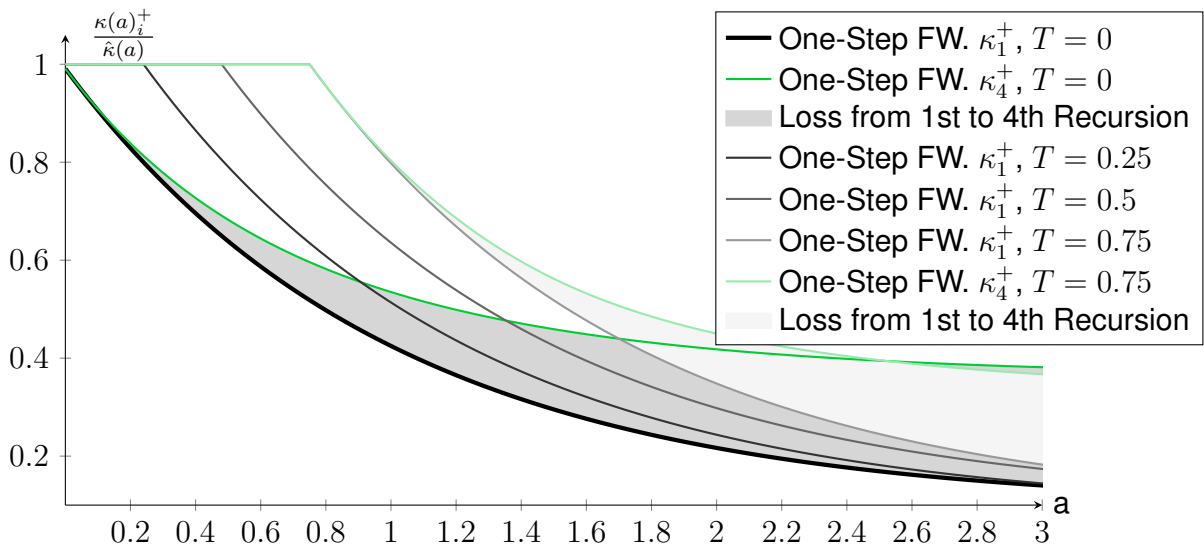

Figure 1.23: The figure shows recursive forward tracing with various fixed delays from $T = 0$ to $T = 0.75$ in the case of $p = 1$. The gray are shows the decreasing effect of the tracing between generation one and four. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

Figure 1.24 concentrates on forward tracing without delay and compares the first four generations of recursive and one-step tracing. In the case of recursive tracing, we observe an increasing effect from generation to generation. The tracing of previous generations results in more traced individuals within the generation $i - 1$. More triggered tracing events at the generation $i - 1$ result in more tracing at generation i.

In contrast to the improved recursive tracing, the detected cases in generation $i - 1$ do not start new tracing events in the case of one-step forward tracing. Furthermore, the detected individuals are removed and do not experience an observed recovery at rate σ to trace the individual of generation i. Therefore, we observe a decreased effect of the tracing. In fig. 1.24 this introduces an alternating sequence. Only unobserved recoveries α and
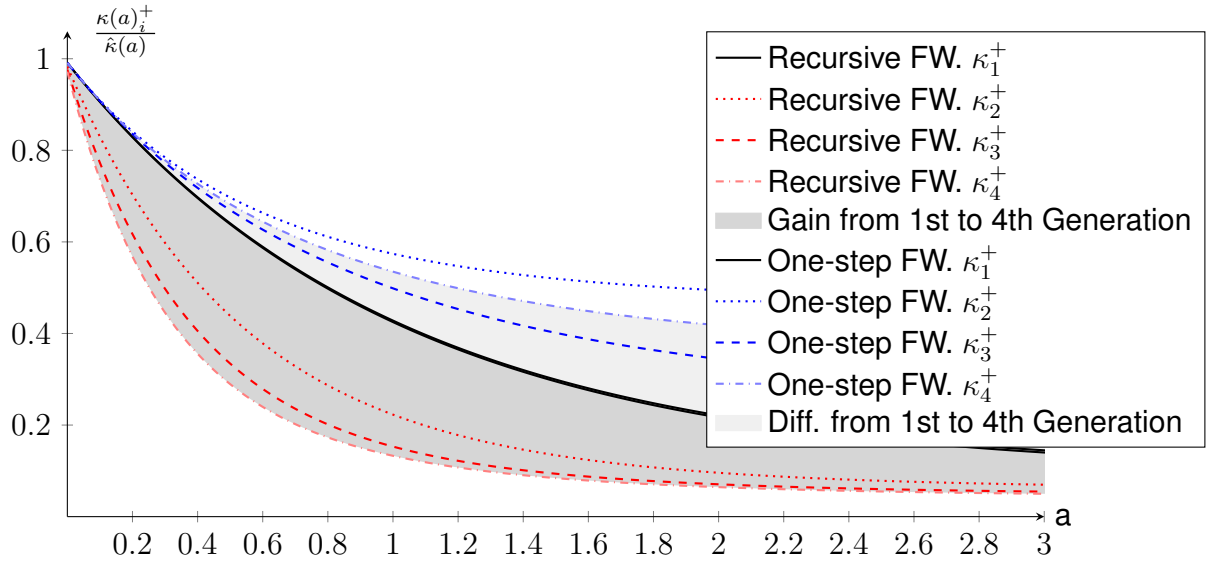
Figure 1.24: The figure shows the first four generations of recursive and one-step forward tracing. It visualizes the gain of recursive forward tracing and compares it to the reduced effect of one-step forward tracing over the first four generations. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

observed recoveries σ, which start tracing events, at the zeroth generation of $\kappa_0(a) = \hat{\kappa}(a)$ results in a lot of detected cases in the first generation of $\kappa_1^+(a)$. A lot of detected traces in $\kappa_1^+(a)$ result in fewer observed recoveries σ, which results in few detections within the second generation of $\kappa_2^+(a)$, resulting in more detections in the third generation of $\kappa_3^+(a)$ before detecting fewer infected individuals in the fourth generation of $\kappa_4^+(a)$.

Simulation The simulations of our individual-based model support the plots of fig. 1.24, which compare one-step and recursive forward tracing. Figure 1.25 shows the number of infected individuals within two populations with recursive forward tracing up to the depth of four and one-step forward tracing. In the case of recursive forward tracing, we had to introduce more infected individuals because the epidemic was not sustainable due to the increased effect of recursive tracing.

Figure 1.26 shows that the recursive forward tracing $\kappa^+(a)$ sampled from the population between time $t = 2.5$ till $t = 5.5$ corresponds to the second generation $\kappa_2^+(a)$ of the recursive forward tracing.

In the case of one-step tracing, we observe a slightly better effect than the one of the one-step $\kappa_2^+(a)$. This might be the case due to a combination of the small sample size and some individuals being sampled as a member of the first and third generation, which experience a larger one-step tracing effect than the second generation.
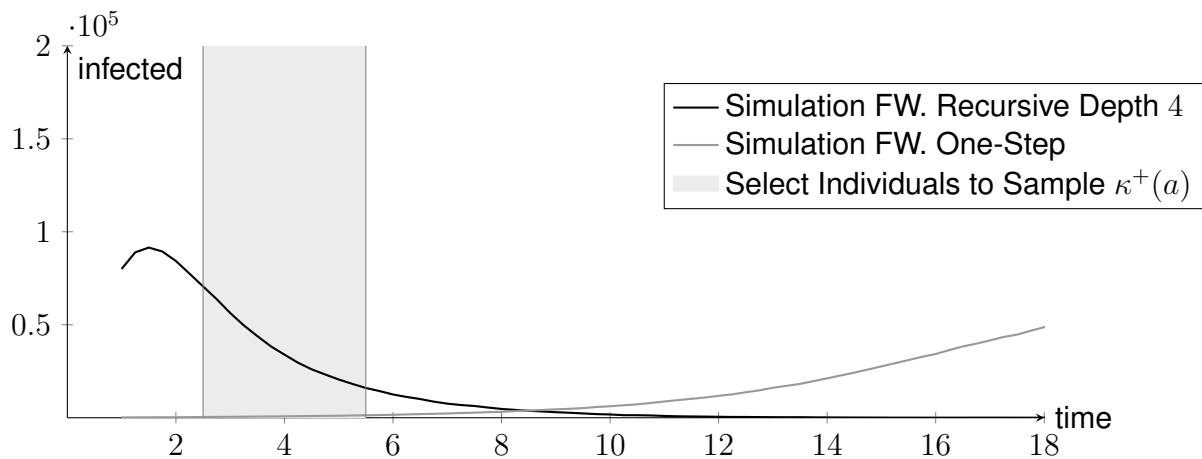
Figure 1.25: The number of infected individuals of two populations of $4 \cdot 10^5$ individuals. We used more initially infected individuals in the case of recursive tracing up to the depth of four, because the spread of the epidemic was not sustainable and the number of infected individuals decreased as soon as the recursive tracing took effect. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.
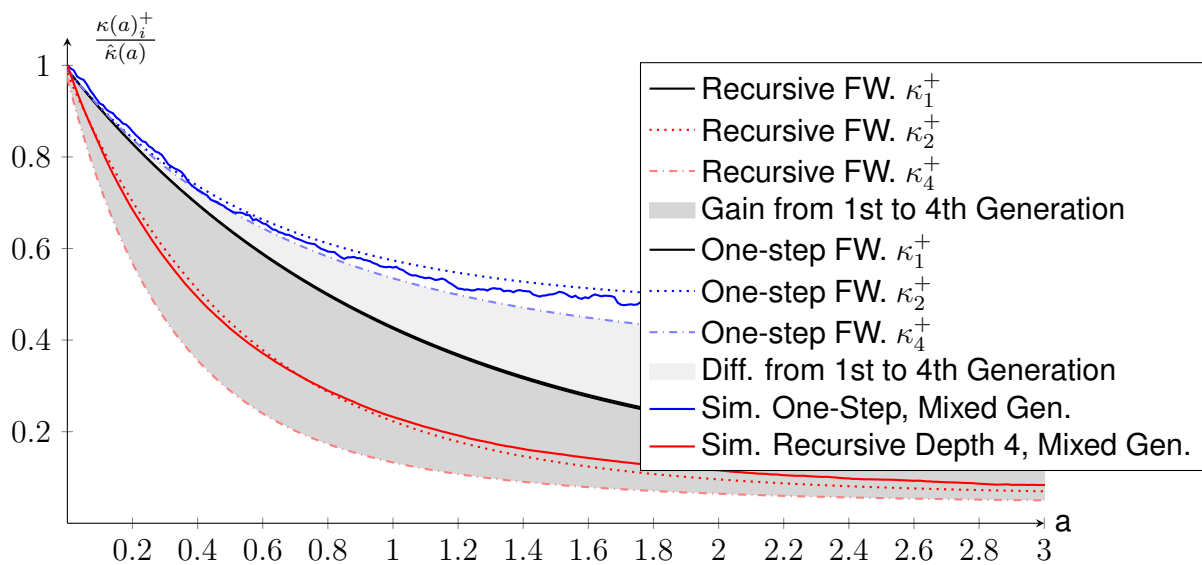

Figure 1.26: The probabilities of being infected sampled at time $t = 2.5$ to $t = 5.5$ corresponding to the populations plotted in fig. 1.25. The individuals used to sample the κ^+ corresponding to the simulation became infected between the time $t = 2.5$ and $t = 5.5$. Thus, they correspond to various generations of infected individuals. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

1.6.1 First-Order Approximation

We approximate the recursion by employing a first-order approximation. For technical reasons, we do not allow ϕ to have a point mass at $t = 0$. Later we will use the closed-form to express dependent expressions, like the reproduction number.

Proposition 24 *We assume the distribution of the tracing delay does not include masses at zero, i.e., $\lim_{\epsilon \to 0} \int_0^\epsilon \phi(a')da' \to 0$. Then the first-order approximation of $\kappa^+(a)$ for $i > 0$ reads:*

$$\kappa_i^+(a) = \hat{\kappa}(a) - pp_{obs}\hat{\kappa}(a)((1 - \hat{\kappa}) * \phi)(a)) + \mathcal{O}(p^2).$$

Proof: First, we recall the recursive equation from proposition 22 describing the forward tracing.

$$\kappa_i^+(a) = \hat{\kappa}(a) \left\{ 1 - p \frac{\int_0^\infty \int_0^a (-\kappa_{i-1}^{+}{}'(b+c) - \alpha\kappa_{i-1}^+(b+c)) \int_c^a \phi(a' - c)da'dc\, db}{\int_0^\infty \kappa_{i-1}^+(b)db} \right\}.$$

We aim at a first-order approximation of κ_i^+. Therefore, it is sufficient to replace $\kappa_{i-1}^+(a)$ on the right-hand side by the zero-order approximation that is $\hat{\kappa}(a) = e^{-(\alpha+\sigma)a}$. Thereby, the denominator simply reads $\int_0^\infty \hat{\kappa}(b)db = \frac{1}{(\sigma+\alpha)}$. The numerator requires more attention:

$$\int_0^\infty \int_0^a \left(-\kappa_{i-1}^{+}{}'(b+c) - \alpha\kappa_{i-1}^+(b+c) \right) \int_c^a \phi(a' - c)da'dc\, db$$

$$\approx \int_0^\infty \int_0^a (\alpha + \sigma)e^{-(\alpha+\sigma)(b+c)} - \alpha e^{-(\alpha+\sigma)(b+c)} \int_c^a \phi(a' - c)da'dc\, db$$

$$= \int_0^\infty \int_0^a \sigma e^{-(\alpha+\sigma)(b+c)\cdot} \int_c^a \phi(a' - c)da'dc\, db$$

Separate independent factors:

$$= \sigma \int_0^\infty e^{-(\alpha+\sigma)b}db \int_0^a e^{-(\alpha+\sigma)c} \int_c^a \phi(a' - c)da'\, dc.$$

Employ the antiderivatives when possible:

$$= \frac{\sigma}{-(\alpha + \sigma)} \left(0 - e^0 \right) \int_0^a \frac{1}{-(\alpha + \sigma)} \frac{d}{dc} e^{-(\alpha+\sigma)c} \int_c^a \phi(a' - c)da'\, dc.$$

Simplify and shift the innermost integral:

$$= \frac{-\sigma}{(\sigma + \alpha)^2} \int_0^a \frac{d}{dc} e^{-(\alpha+\sigma)c} \int_0^{a-c} \phi(a'')da''dc$$

Employ integration by parts:

$$= \frac{-\sigma}{(\sigma+\alpha)^2} \left\{ e^{-(\alpha+\sigma)c} \int_0^{a-c} \phi(a')da' \Big|_{c=0}^a - \int_0^a e^{-(\alpha+\sigma)c} \frac{d}{dc} \int_0^{a-c} \phi(a')da'dc \right\}.$$

We now use $\int_0^{a-a} \phi(a')da' = 0$ from the preconditions.

$$= \frac{-\sigma}{(\sigma+\alpha)^2} \left\{ -\int_0^{a-0} 1 \cdot \phi(a')da' + \int_0^a e^{-(\alpha+\sigma)c}\phi(a-c)dc \right\}$$

$$= \frac{\sigma}{(\sigma+\alpha)^2} \left\{ \int_0^a 1 \cdot \phi(a-c)dc - \int_0^a e^{-(\alpha+\sigma)c}\phi(a-c)dc \right\}$$

$$= \frac{\sigma}{(\sigma+\alpha)^2} ((1-\hat{\kappa}) * \phi)(a).$$

With $p_{\text{obs}} = \frac{\sigma}{\sigma+\alpha}$, we obtain the result:

$$\kappa_i^+(a) = \hat{\kappa}(a) \left\{ 1 - p\frac{\int_0^\infty \int_0^a (-\kappa_{i-1}^{+\,'}(b+c) - \alpha\kappa_{i-1}^+(b+c)) \int_c^a \phi(a'-c)da'dc\,db}{\int_0^\infty \kappa_{i-1}^+(b)db} \right\}$$

$$= \hat{\kappa}(a) \left\{ 1 - p\frac{\frac{\sigma}{(\sigma+\alpha)^2}((1-\hat{\kappa}) * \phi)(a)}{\frac{1}{\sigma+\alpha}} \right\} + \mathcal{O}(p^2).$$
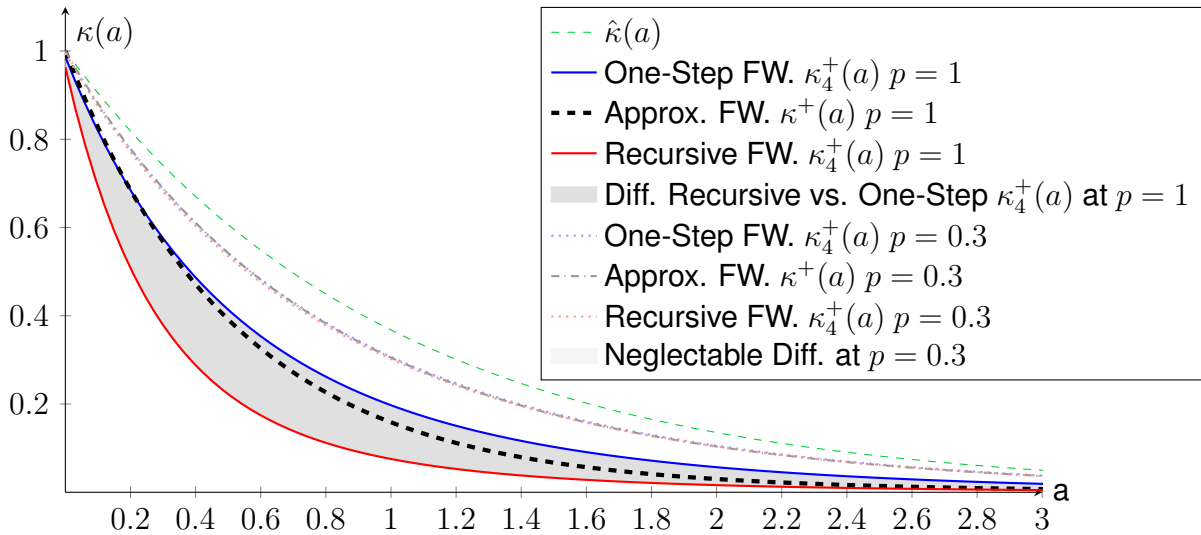
$\square$

Figure 1.27: The probabilities of being infective of the fourth generation $\kappa_4^+(a)$ of recursive and one-step tracing in the case of $p = 1$ and $p = 0.3$ compared to the approximation κ^+. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$.

Figure 1.27 visualizes the forward tracing approximation compared to the one-step and

recursive tracing of the fourth generation using the probability of recalling contacts $p = 1$ and $p = 0.3$.

As observed in the relative plots of the recursive forward tracing case, the difference between one-step and recursive tracing in the case of $p = 0.3$ is considerably smaller as the probability of $p = 0.3$ affects each generation and breaks the recursions. In that case, we observe a good fit for the forward tracing approximation.

In the proof for proposition 24, we constrained the delay to not including point masses at zero. This constraint simplifies the integration by parts. However, this assumption can be removed by an approximation argument. Particularly, also the case of no delay (T=0) can be handled by proposition 25. Thus, we recover the result of the model without delay [MH07, Proposition 3.2].

Proposition 25 *Removing the delay distribution $\phi(a)$ by replacing it with a fixed delay of length zero, i.e., $\phi(a) = \delta_0(a)$, allows for evaluation of the forward tracing approximation for $i > 0$ to*

$$\kappa_i^+(a) = \hat{\kappa}(a) - p\, p_{obs}\, \hat{\kappa}(a)(1 - \hat{\kappa}(a)) + \mathcal{O}(p^2) \text{ for } i > 0.$$

Proof: To prove the claim, we start right before using integration by parts. Furthermore, we start using $\delta_\epsilon(a)$ to avoid $\delta_0(a)$ at the integral border. Then, we show that in case of no delay the reason to perform integration by parts instead of a direct integration vanishes.

$$\sigma \int_0^\infty e^{-(\alpha+\sigma)b} db \int_0^a e^{-(\alpha+\sigma)c} \int_c^a \delta_\epsilon(a'-c) da'\, dc$$
$$= \frac{-\sigma}{-(\alpha+\sigma)} \int_0^a e^{-(\alpha+\sigma)c} \int_c^a \delta_\epsilon(a'-c) da'\, dc$$

Shifting the inner integral and writing $\int_0^{a-c} \delta_0(a') da'$ underlines that it always equals 1 within the range of the outer integral $c \in [0; a]$ for $\epsilon < a$. Now, we introduce an approximation argument and take the limes from above $\delta_0 = \lim_{\epsilon \to 0} \delta_\epsilon$. Using this argument, we write:

$$= \frac{-\sigma}{-(\alpha+\sigma)} \int_0^a e^{-(\alpha+\sigma)c} \cdot 1\, dc$$
$$= \frac{-\sigma}{-(\alpha+\sigma)} \frac{1}{-(\sigma+\alpha)} \left(e^{-(\alpha+\sigma)a} - 1 \right)$$
$$= \frac{\sigma}{(\sigma+\alpha)^2} (1 - \hat{\kappa}(a)).$$

This calculation proves the result of proposition 24 is valid in case of a fixed delay of length zero. $\qquad\square$
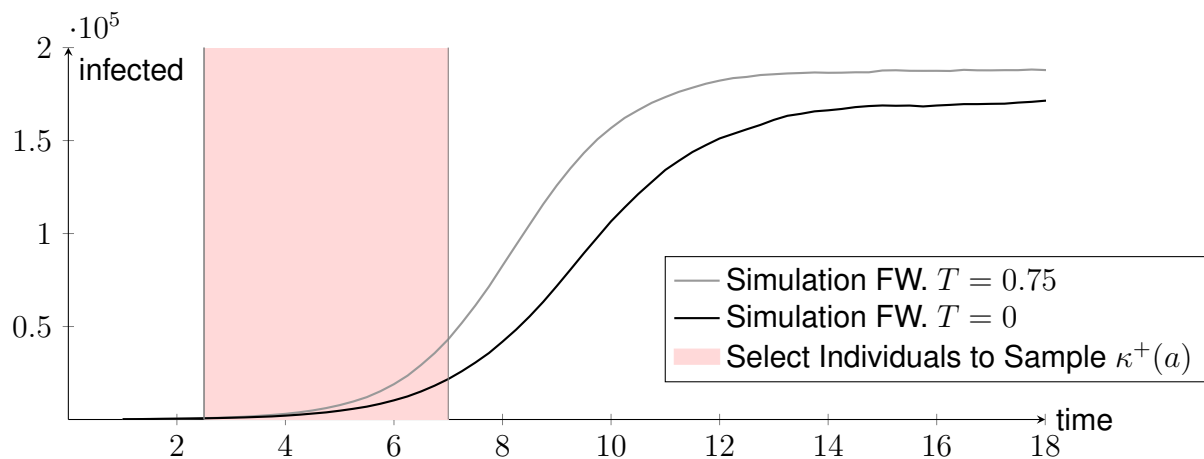
Figure 1.28: The count of infected individuals among a population of $4 \cdot 10^5$ with one-step forward tracing with a recall probability $p = 0.3$ and a contact rate $\beta = 2$, spontaneous recovery $\alpha = 0.1$, and supervised recovery $\sigma = 0.9$.

Simulation   We support the applicability of the approximation by IBM simulations. This time, we simulate a population of $4 \cdot 10^5$ individuals with one-step forward tracing at the fixed delay of $T = 0$ and $T = 0.75$. Figure 1.28 visualizes the epoch we used to sample the infected individuals we use to calculate $\kappa^+(a)$.

Figure 1.29 shows the matching shape of the approximation κ^+, $\kappa_4^+(a)$ calculated by the recursive model and the corresponding $\kappa^+(a)$ sampled from the data of the simulation.

Because we use a realistic probability of recalling a contact $p = 0.3$, we observe a good fit of the approximation, the model, and the simulation.
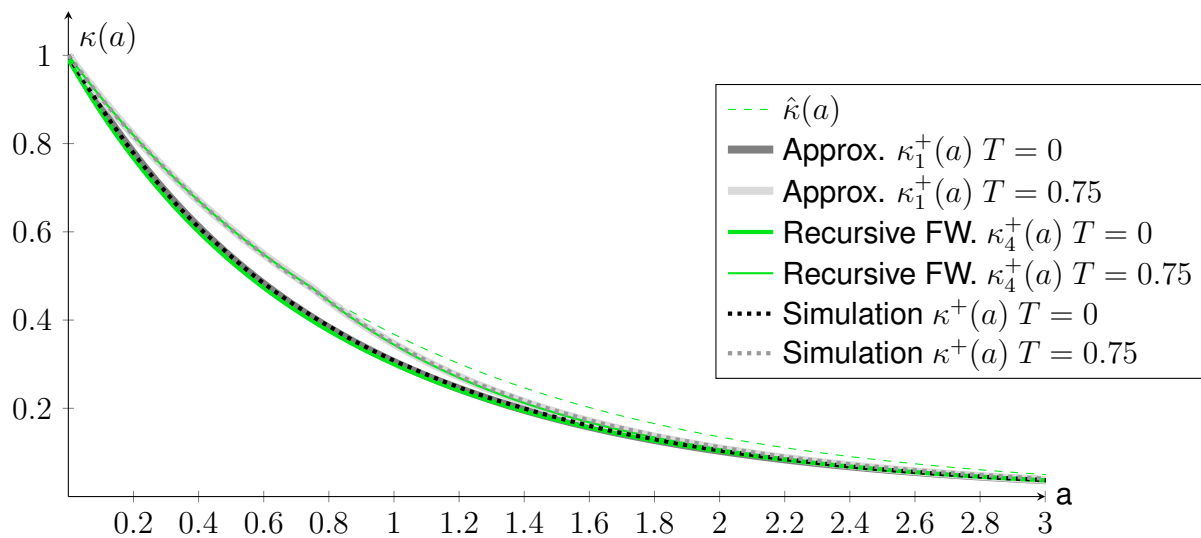
Figure 1.29: The sampled probability of being infective $\kappa^+(a)$ of the individuals of various generations from the epoch between $t = 2.5$ and $t = 7$ corresponding to fig. 1.28 at $T = 0$ and $T = 0.75$ compared to the approximation and the fourth generation of the recursive model. The green plots from the fourth order of the forward tracing recursion show the values shown in fig. 1.24. We show the forward tracing recursion solution to visualize the effect of different tracing delays without noise. The dashed black and gray lines $T = 0$ and $T = 0.75$ visualize the quality of the fit of the simulation.

The onset graphs, including tracing, branch away from the graph $\hat{\kappa}(a)$ without tracing. The time of this branching depends on the tracing delay T.

1.6.2 Specific Delays

Similar to backward tracing, we exemplify the equation by replacing the general distribution $\phi(a)$ with specific distributions, which allow further computation.

Fixed Delay

Proposition 26 *Replacing the delay distribution $\phi(a)$ by a fixed delay $\phi(a) = \delta_T(a)$ exemplifies the equation:*

$$\kappa_i^+(a) = \begin{cases} \hat{\kappa}(a) - p\,p_{obs}\hat{\kappa}(a)(1 - \hat{\kappa}(a - T)) + \mathcal{O}(p^2) & a \geq T \text{ and } i > 0, \\ \hat{\kappa}(a) + \mathcal{O}(p^2) & a < T \text{ or } i = 0. \end{cases}$$

Proof: This proposition is an immediate consequence of proposition 24,

$$\kappa_i(a) = \hat{\kappa}(a) - p\,p_{\text{obs}}\hat{\kappa}(a)((1 - \hat{\kappa}) * \phi)(a)) + \mathcal{O}(p^2),$$

and the evaluation of the delay $\phi * (1 - \hat{\kappa})(a) = (1 - \hat{\kappa}(a - t))$ while respecting that $\kappa(a) = 0$ for $a < 0$. $\square$

Exponential Delay

Proposition 27 *Using an exponential delay distribution $\phi(a) = \frac{1}{T}e^{\frac{-a}{T}}$, the first-order approximation of $\kappa^+(a)$ for $i > 0$ reads*

$$\kappa_i^+(a) = \hat{\kappa}(a) - p\,p_{obs}\hat{\kappa}(a)\left(1 - e^{-\frac{a}{T}} - \frac{\hat{\kappa}(a) - e^{-\frac{a}{T}}}{1 - T(\alpha + \sigma)}\right) + \mathcal{O}(p^2).$$

Proof: Again, we take the first-order approximation from proposition 24. We insert the distribution and compute the part dealing with delay:

$$((1 - \hat{\kappa}) * \phi)(a))$$
$$= \int_0^a \left(1 - e^{-(\alpha+\sigma)(a-\tau)}\right) e^{\frac{-\tau}{T}} \frac{1}{T} d\tau$$
$$= \frac{1}{T} \int_0^a e^{\frac{-\tau}{T}} - e^{-(\alpha+\sigma)a + \left(-(\alpha+\sigma) - \frac{1}{T}\right)\tau} d\tau$$
$$= \frac{1}{T} \left((-T)[e^{\frac{-\tau}{T}}]_0^a - \frac{\kappa(a)}{(\alpha+\sigma) - \frac{1}{T}} \left[e^{((\alpha+\sigma) - \frac{1}{T})\tau}\right]_0^a \right)$$
$$= 1 - e^{\frac{-a}{T}} - \frac{1}{T} \frac{\hat{\kappa}(a)}{(\alpha+\sigma) - \frac{1}{T}} \left(\hat{\kappa}(-a)e^{\frac{-a}{T}} - 1\right)$$

$$=1 - e^{\frac{-a}{T}} - \frac{e^{\frac{-a}{T}} - \hat{\kappa}(a)}{T(\alpha + \sigma) - 1}.$$

Inserting into the first-order approximation now proves the assumption.   □
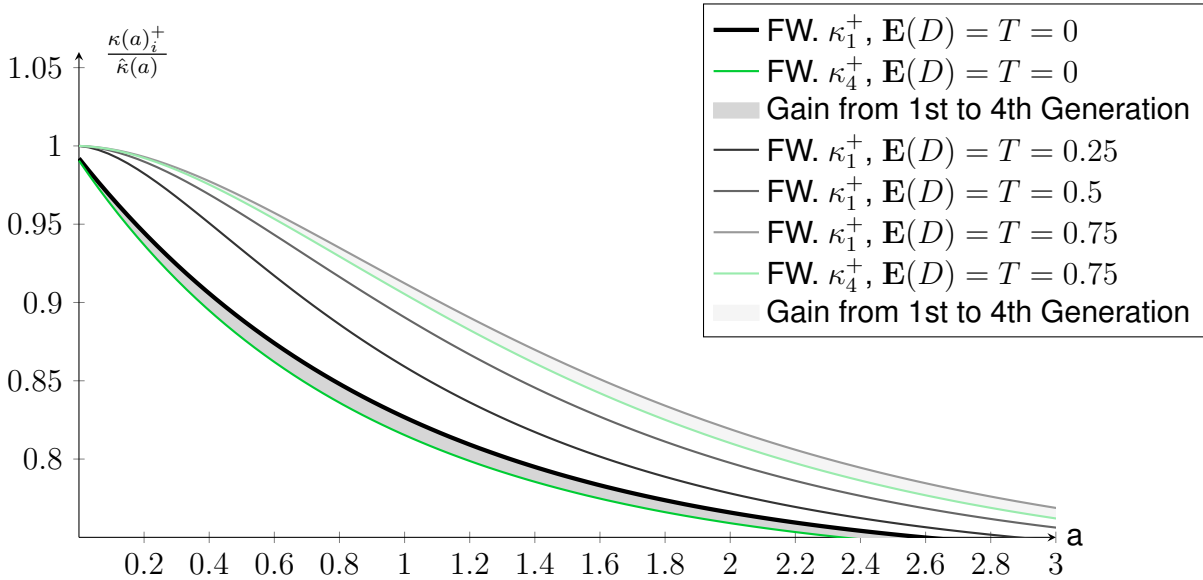

Figure 1.30: Recursive forward tracing with exponential delay from zero to $\mathbf{E} = T = 0.75$. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 0.3$.

Figure 1.30 visualizes recursive forward tracing with exponential tracing delays from no delay to $\mathbf{E} = 0.75$. In the case of an exponential delay, we do not see the effect of the tracing branching away from the no-tracing case in a clearly visible junction. Instead of junctions, we observe a more gradual start of the tracing. Figure 1.31 compares recursive forward tracing with fixed delay $T = 0.75$ to recursive forward tracing with exponential delay $\mathbf{E} = 0.75$. We observe that the gradual start of the exponential delay allows for a lot of tracing until the fixed delay tracing starts.
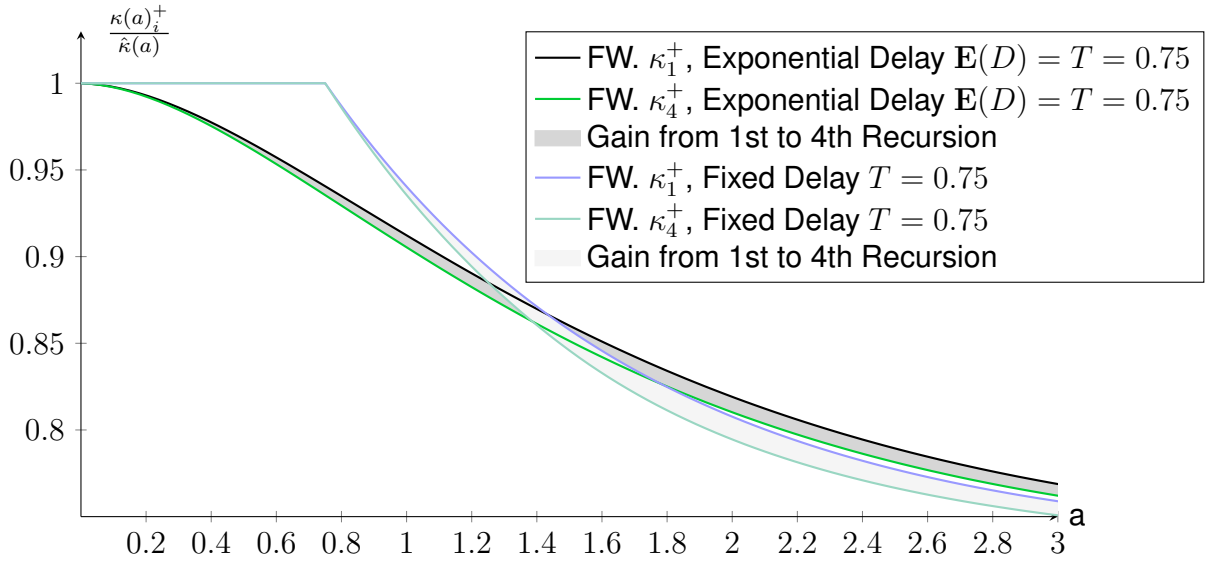
Figure 1.31: We compare the first and fourth generation of recursive forward tracing with exponential delay with $\mathbf{E}(D) = T = 0.75$ to the first and fourth generation of recursive forward tracing with fixed delay $T = 0.75$. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, and $p = 0.3$.
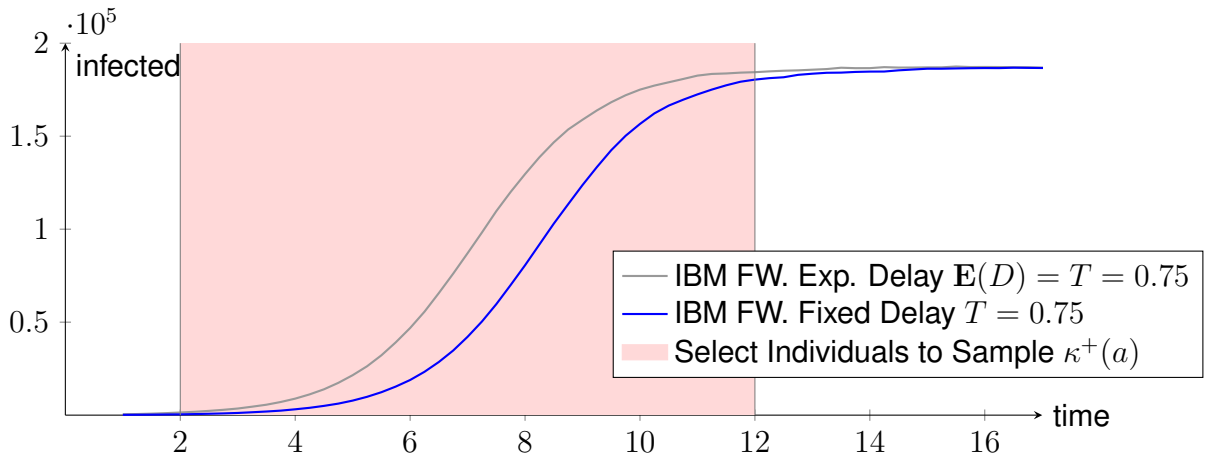

Figure 1.32: The infected individuals of recursive forward tracing until the fourth generation with exponential delay compared to recursive fixed delay forward tracing at $p = 0.3$, the fixed delay $T = 0.75$, and the exponential delay $\mathbf{E}(D) = T = 0.75$. The other parameters used are $\beta = 2$, $\sigma = 0.9$, and $\alpha = 0.1$.

Simulation   The IBM simulations support the shape of $\kappa_4^+(a)$ shown in fig. 1.31. We simulated two populations using recursive tracing up to the depth of four with fixed and exponential delay. To counter the variation, we sampled from $t = 2$ till $t = 12$. The sampled $\kappa^+(a)$ for the exponential and the fixed delay forward tracing show the effect of exponential delay before the fixed delay progressed and the slightly stronger effect of fixed delay tracing
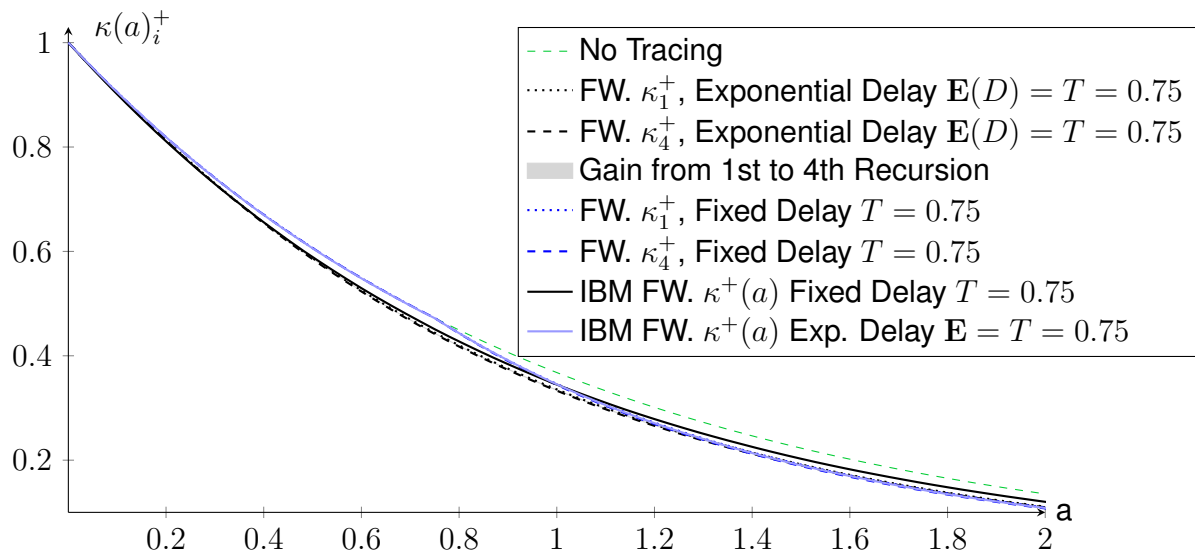
at later times of infection a.

Figure 1.33: The sampled probability of being infective $\kappa^+(a)$ of recursive forward tracing until the fourth generation with exponential delay compared to recursive fixed delay forward tracing at $p = 0.3$ and $T = 0.75$ fig. 1.32. For reference, we plotted the probabilities of being infective $\kappa_1^+(a)$ and $\kappa_4^+(a)$ of recursive forward tracing with exponential and fixed delay. The other parameters used are $\beta = 2$, $\sigma = 0.9$, and $\alpha = 0.1$.

1.6.3 Reproduction Number in Case of Forward Tracing

We recall corollary 7 introducing the effective reproduction number, which states the number of infectees per infector in the onset case respecting contact tracing. In the case of forward tracing, we conclude the following corollary.

Corollary 28 *The effective reproduction number in the case of plain forward tracing reads*

$$R_i^+ = \int_0^\infty \beta \kappa_i^+(a) da.$$

Using the first-order forward tracing approximation, we can expand $\kappa_i^+(a)$ for $i > 0$:

$$R_i^+ \approx \int_0^\infty \beta \left(\hat{\kappa}(a) - p\, p_{obs} \hat{\kappa}(a)((1 - \hat{\kappa}) * \phi)(a) \right) da + \mathcal{O}(p^2).$$

For R_0^+ the forward tracing resembles the no tracing case:

$$R_0^+ = R_0.$$

Reproduction Number in the Case of Fixed Delay Forward Tracing

Instead of using the general forward tracing approximation to calculate the reproduction number, we can solve the integral using the exemplified approximations, and we start using the approximation for a fixed delay $\phi(a) = \delta_T(a)$.

Proposition 29 *The effective reproduction number in the case of forward tracing with fixed delay reads for $i > 0$:*

$$R_i^{+,f} = \frac{\beta}{\alpha + \sigma} - p\, p_{obs} \beta \frac{\hat{\kappa}(T)}{2(\alpha + \sigma)} + \mathcal{O}(p^2).$$

Proof: First, we recall the single-step approximation of the reproduction number using a fixed delay and introduce it into the effective reproduction number from corollary 7.

$$
\begin{aligned}
R_i^{+,f} &= \int_0^\infty \beta \hat{\kappa}(a) - p\, p_{\text{obs}} \beta \hat{\kappa}(a)(1 - \hat{\kappa}(a - T)) da + \mathcal{O}(p^2) \\
&= \int_0^\infty \beta \hat{\kappa}(a) da - p\, p_{\text{obs}} \beta \int_0^\infty \hat{\kappa}(a)(1 - \hat{\kappa}(a - T)) da + \mathcal{O}(p^2).
\end{aligned}
$$

Then we concentrate on the integral, which describes the forward tracing update:

$$\int_T^\infty \hat{\kappa}(a) \cdot (1 - \hat{\kappa}(a - T))da = \int_T^\infty \hat{\kappa}(a) - \int_T^\infty \hat{\kappa}(a)\hat{\kappa}(a - T)da$$

$$= \left[\frac{-\hat{\kappa}(a)}{\alpha + \sigma}\right]_T^\infty - \left[\frac{-\hat{\kappa}(2a)}{2(\alpha + \sigma)\hat{\kappa}(T)}\right]_T^\infty = \frac{\kappa(T)}{2(\alpha + \sigma)}.$$

Using this result, we evaluate the integral:

$$R_i^{+,f} = \int_0^\infty \beta\hat{\kappa}(a)da - p\,p_{\mathsf{obs}}\beta \int_0^\infty \hat{\kappa}(a)(1 - \hat{\kappa}(a - T))da + \mathcal{O}(p^2)$$

$$= \frac{\beta}{\alpha + \sigma} - p\,p_{\mathsf{obs}}\beta\frac{\hat{\kappa}(T)}{2(\alpha + \sigma)} + \mathcal{O}(p^2).$$

$\square$

Corollary 30 *We can express the reproduction number in the case of fixed delay forward tracing for $i > 0$ using the basic reproduction number:*

$$R_i^{+,f} = R_0 - p\,p_{obs}R_0\frac{\hat{\kappa}(T)}{2} + \mathcal{O}(p^2).$$

Reproduction Number in the Case of Exponential Delay Forward Tracing

Proposition 31 *The effective reproduction number in the case of forward tracing with exponential delay reads for $i > 0$:*

$$R_i^{+,e} = \frac{\beta}{\alpha + \sigma} - p\,p_{obs}\frac{\beta}{\alpha + \sigma}\frac{1}{2(1 + T(\alpha + \sigma))} + \mathcal{O}(p^2).$$

Proof: First, we recall the single-step approximation of the reproduction number using an exponential delay and introduce it into the effective reproduction number from corollary 7.

$$R_i^{+,e} = \int_0^\infty \beta\hat{\kappa}(a) - p\,p_{\mathsf{obs}}\beta\hat{\kappa}(a)\left(1 - e^{-\frac{a}{T}} - \frac{\hat{\kappa}(a) - e^{-\frac{a}{T}}}{1 - T(\alpha + \sigma)}\right)da + \mathcal{O}(p^2).$$

Then, we expand the term and solve the resulting smaller integrals:

$$R_i^{+,e} = \int_0^\infty \beta\hat{\kappa}(a)da - p\,p_{\mathsf{obs}}\beta\left\{\int_0^\infty \hat{\kappa}(a)da - \int_0^\infty \hat{\kappa}(a)e^{-\frac{a}{T}}da\right.$$

$$\left. - \int_0^\infty \frac{\hat{\kappa}(a)\hat{\kappa}(a)}{1 - T(\alpha + \sigma)}da + \int_0^\infty \frac{\hat{\kappa}(a)e^{-\frac{a}{T}}}{1 - T(\alpha + \sigma)}da\right\}da + \mathcal{O}(p^2).$$

Now, we solve the individual integrals:

$$\int_0^\infty \beta\hat{\kappa}(a)da = \int_0^\infty \beta\hat{\kappa}(a)da = \frac{\beta}{\alpha + \sigma} = R_0,$$

$$\int_0^\infty \hat{\kappa}(a)da = \int_0^\infty \hat{\kappa}(a)da = \frac{1}{\alpha + \sigma},$$

$$\int_0^\infty \hat{\kappa}(a)e^{-\frac{a}{T}}da = \frac{T}{1 + (a + \sigma)T},$$

$$\int_0^\infty \frac{\hat{\kappa}(a)\hat{\kappa}(a)}{1 - T(\alpha + \sigma)}da = \frac{-1}{2(\alpha + \sigma)(T(\alpha + \sigma) - 1)},$$

$$\int_0^\infty \frac{\hat{\kappa}(a)e^{-\frac{a}{T}}}{1 - T(\alpha + \sigma)}da = \frac{-T}{(1 + T(\alpha + \sigma))(T(\alpha + \sigma) - 1)}.$$

Now, we insert the solutions into the integral:

$$R^{+,e} = \frac{\beta}{\alpha + \sigma} - p\,p_{\text{obs}}\beta\left\{\frac{1}{\alpha + \sigma} - \frac{T}{1 + (a + \sigma)T}\right.$$

$$\left. - \frac{-1}{2(\alpha + \sigma)(T(\alpha + \sigma) - 1)} + \frac{-T}{(1 + T(\alpha + \sigma))(T(\alpha + \sigma) - 1)}\right\} + \mathcal{O}(p^2)$$

$$= \frac{\beta}{\alpha + \sigma} - p\,p_{\text{obs}}\beta\left\{\frac{2(1 + T(\alpha + \sigma))(T(\alpha + \sigma) - 1) - T \cdot 2(\alpha + \sigma)(T(\alpha + \sigma) - 1)}{2(\alpha + \sigma)(1 + T(\alpha + \sigma))(T(\alpha + \sigma) - 1)}\right.$$

$$\left. + \frac{1 + T(\alpha + \sigma) - T\,2(\alpha + \sigma)}{2(\alpha + \sigma)(1 + T(\alpha + \sigma))(T(\alpha + \sigma) - 1)}\right\} + \mathcal{O}(p^2)$$

$$= \frac{\beta}{\alpha + \sigma} - p\,p_{\text{obs}}\beta\left\{\frac{(\alpha + \sigma)T - 1}{2(\alpha + \sigma)(1 + T(\alpha + \sigma))(T(\alpha + \sigma) - 1)}\right\} + \mathcal{O}(p^2)$$

$$= \frac{\beta}{\alpha + \sigma} - p\,p_{\text{obs}}\frac{\beta}{\alpha + \sigma}\frac{1}{2(1 + T(\alpha + \sigma))} + \mathcal{O}(p^2).$$

$$\square$$

Corollary 32 *We can express the reproduction number in the case of exponential delay forward tracing for $i > 0$ using the basic reproduction number:*

$$R_i^{+,e} = R_0 - p\,p_{obs}R_0\frac{1}{2(1 + T(\alpha + \sigma))} + \mathcal{O}(p^2).$$

Comparison of Fixed and Exponential Delay

This comparison resembles the backward tracing case from section 1.5.3 'Comparison of Fixed and Exponential Delay'. Therefore, we recall the effective reproduction numbers in

the case backward tracing with fixed and exponential delay:

$$R_{-,\mathsf{f}} = R_0 - p\, p_{\mathsf{obs}} R_0^2 \frac{\hat{\kappa}(T)}{2} + \mathcal{O}(p^2),$$

$$R_{-,\mathsf{e}} = R_0 - p\, p_{\mathsf{obs}} R_0^2 \frac{1}{2(1 + T(\alpha + \sigma))} + \mathcal{O}(p^2).$$

We compare the backward tracing to the forward tracing case for $i > 0$:

$$R_i^{+,f} = R_0 - p\, p_{\mathsf{obs}} R_0 \frac{\hat{\kappa}(T)}{2} + \mathcal{O}(p^2),$$

$$R_i^{+,e} = R_0 - p\, p_{\mathsf{obs}} R_0 \frac{1}{2(1 + T(\alpha + \sigma))} + \mathcal{O}(p^2).$$

By comparing the additive terms of the approximation of the reproduction number, we observe a similarity apart from the additional factor of R_0 in the backward tracing case. Hence, we conclude the same advantage of the exponential delay over a fixed delay in the forward tracing case.

1.7 Full Tracing

Now we combine forward tracing κ^+ and backward tracing κ_- to full tracing κ. We construct a recursive formula consisting of the backward tracing from proposition 10 and the forward tracing from proposition 22. Note: In the case of $\kappa_0(a)$, i.e., the first generation, no forward tracing occurs. In that case, only the backward tracing is active due to the lack of infectors from a previous generation.

Proposition 33 (Recursive Full Tracing) *κ_i for $i > 0$ follows from the following recursive formula, in which the zeroth generation $\kappa_0(a)$ only experiences recursive backward tracing*

$$\kappa_0'(a) = -\kappa_0(a) \left\{ \alpha + \sigma + p\beta \left[(\phi * (1 - \kappa_0))(a) - \alpha(\phi * \kappa_0^{\#})(a) \right] \right\},$$
$$\kappa_0(0) = 1.$$

In the case of $i > 1$, the recursion formula for $\kappa_i(a)$ includes recursive forward tracing

$$\kappa_{i-1}(b)\kappa_i(a|b) = \kappa_0(a) \left\{ \kappa_{i-1}(b) - p \int_0^a \left(-\kappa_{i-1}'(b+c) - \alpha\kappa_{i-1}(b+c) \right) \int_c^a \phi(a' - c)da'dc \right\},$$
$$\kappa_i(a) = \frac{\int_0^\infty \kappa_i^+(a|b)\kappa_{i-1}(b)db}{\int_0^\infty \kappa_{i-1}(\tau)d\tau}.$$

Proof: Forward tracing is triggered by the infector, backward tracing is triggered by the infectee. For the focussed individual, these two groups are independent, and we combine their influence.

In the case of recursive forward tracing from proposition 22, we took the probability of being infective without tracing $\hat{\kappa}(a)$ to calculate $\kappa_{i-1}^+(b)\kappa_i^+(a|b)$ by multiplying the probability to not be detected by forward tracing. Because of the independence of the backward tracing from the forward tracing, this influence of backward tracing acts on each generation independent of the forward tracing. Thus, we use $\kappa_0(a) = \kappa_-(a)$ instead of the probability to be infective without tracing $\hat{\kappa}(a)$ to calculate full tracing. $\quad\square$

We plot the probability of being infective corresponding to recursive full tracing of the first and fourth generation $\frac{\kappa_i}{\kappa(a)}$ to compare it to the first and fourth generations of forward tracing $\frac{\kappa_i^+}{\kappa(a)}$ and one-step and recursive backward tracing $\frac{\kappa_-(a)}{\hat{\kappa}(a)}$ in fig. 1.34. Dividing the probability of being infective with tracing by the probability of being infective without tracing $\hat{\kappa}(a)$ highlights the tracing effect.

We observe an increased effect from recursive tracing from the first to the fourth generation, and we note that these generations do not exist in backward tracing. Nevertheless,
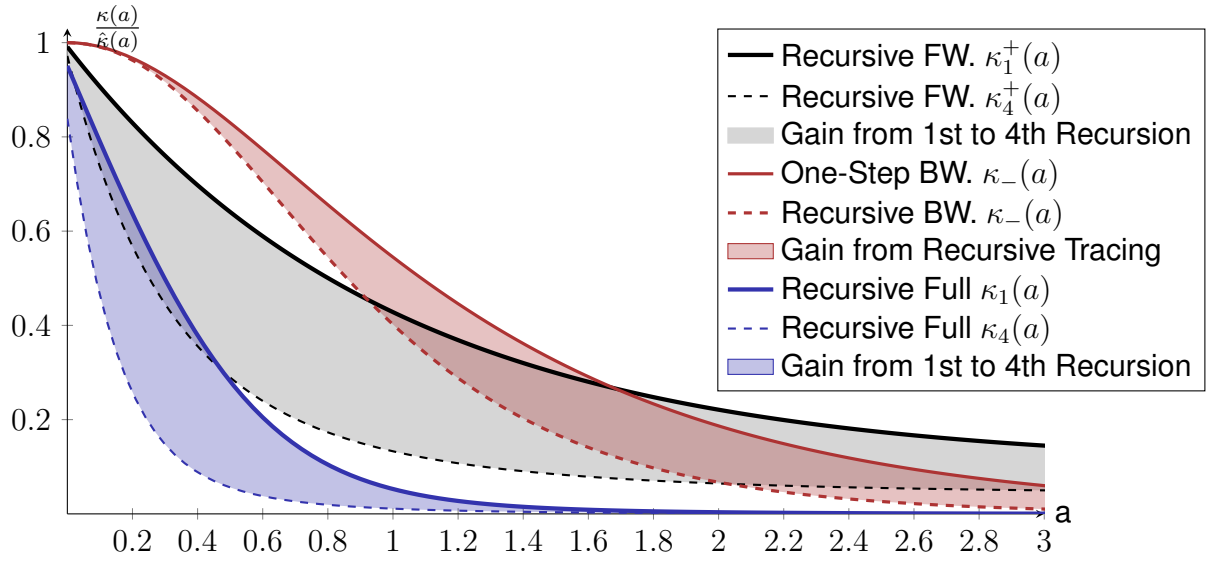
Figure 1.34: Comparison of the probability to be infective κ of recursive forward tracing, recursive full tracing and one-step and recursive backward tracing divided by the probability $\hat{\kappa}(a)$ of being infective without tracing to highlight the tracing effect. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$, and $T = 0$.

we observe the increased effect of the full tracing when the recursive backward tracing is combined with recursive forward tracing.

One-Step Tracing

We can apply the arguments of the recursive model proposition 33 on the one-step backward tracing from proposition 11 and forward tracing from proposition 23.

Proposition 34 (One-Step Full Tracing) *We characterize κ_i for $i > 0$ by the following recursive formula, in which the zeroth generation $\kappa_0(a)$ only experiences one-step backward tracing*

$$\kappa_0'(a) = -\kappa_0(a) \left\{ \alpha + \sigma + p\beta\sigma(\phi * \kappa_-^{\#})(a) \right\},$$

$$\kappa_0(0) = 1.$$

In the case of $i > 1$, the recursion formula for $\kappa_i(a)$ includes one-step forward tracing

$$\kappa_{i-1}(b)\kappa_i(a|b) = \kappa_0(a) \left\{ \kappa_{i-1}(b) - p\sigma \int_0^a \kappa_{i-1}(b+c) \int_c^a \phi(a'-c)da'dc \right\},$$

$$\kappa_i(a) = \frac{\int_0^\infty \kappa_i^+(a|b)\kappa_{i-1}(b)db}{\int_0^\infty \kappa_{i-1}(\tau)d\tau}.$$
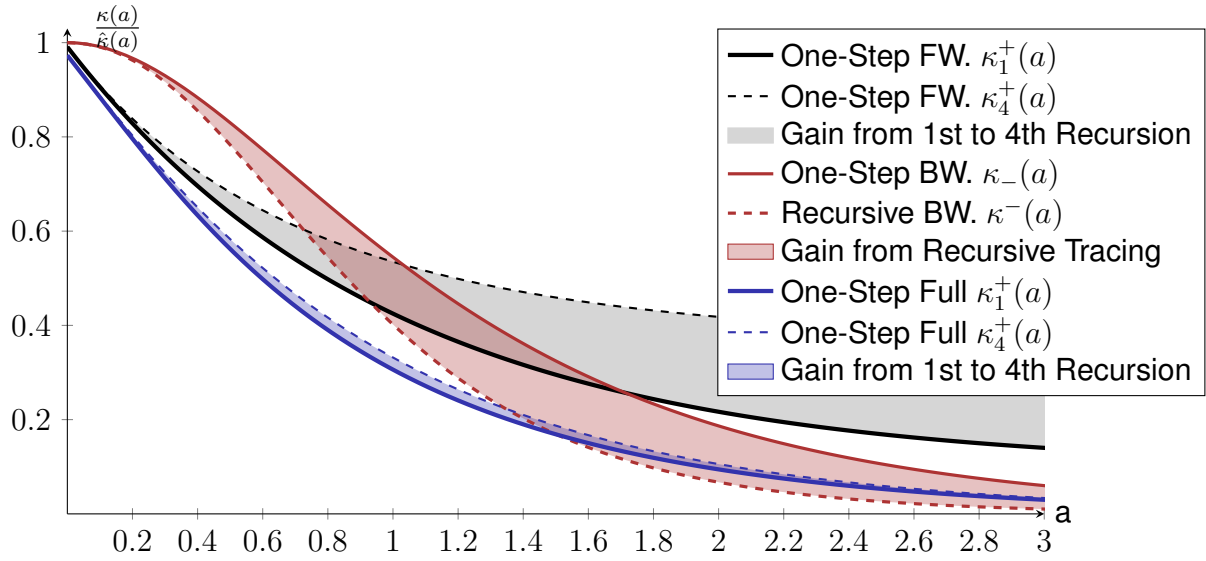
Figure 1.35: Comparison of one-step forward tracing, one-step backward and recursive backward tracing and full tracing. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, $p = 1$, and $T = 0$.

We complete the comparison of recursive forward, backward and full tracing from fig. 1.34 by comparing the corresponding one-step forward, backward and full-tracing probabilities of being infective divided by $\hat{\kappa}(a)$ in fig. 1.35. We observe the increased effect of the one-step full tracing compared to one-step forward and one-step backward tracing. We note that there are no generations in backward tracing as they are present in forward tracing. Therefore, the fourth generation of full and forward tracing can not be compared to recursive backward tracing.

Simulation

The simulations of our individual-based model support the recursive and one-step full tracing. Figure 1.36 shows the infected individuals within a population of $4 \cdot 10^5$ individuals with one-step full tracing. The highlighted area shows the times t used to sample the probability of being infective $\kappa(a)$.

Figure 1.37 shows the sampled probability of being infective. We observe the $\kappa(a)$ from the simulation matching the shape of the $\kappa_1(a)$ from the model.
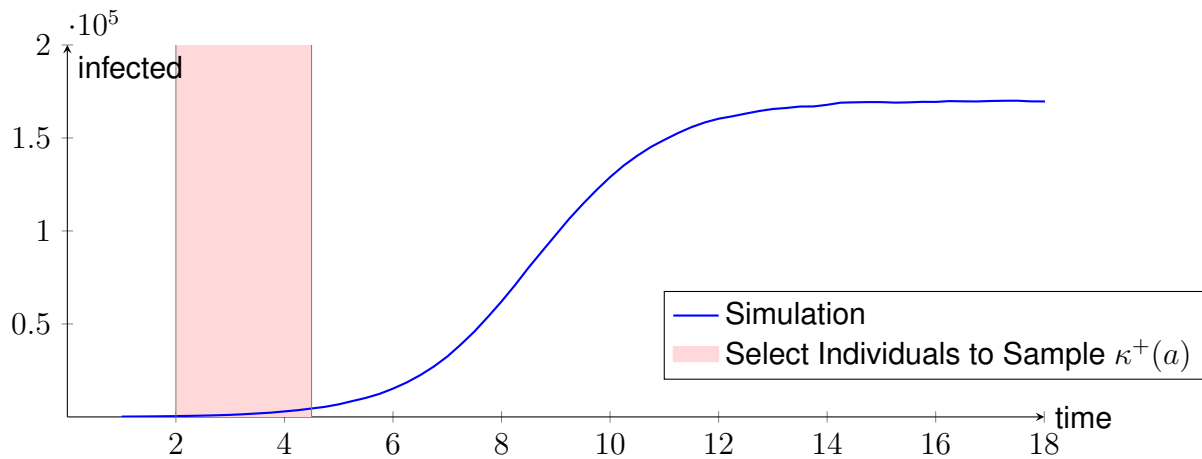
Figure 1.36: The number of infected individuals in a population of $4 \cdot 10^5$ individuals and one-step full tracing with fixed delay of $T = 0.5$. The other parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, and $p = 0.3$.
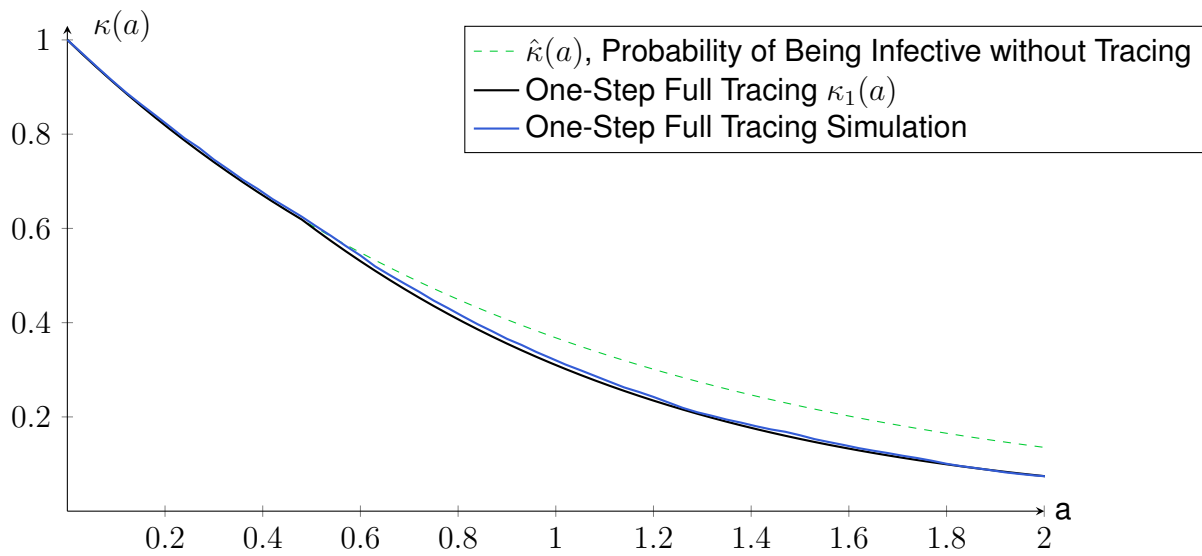

Figure 1.37: The sampled probability of being infective with one-step full tracing from an IBM simulation of a population of $4 \cdot 10^5$ individuals with fixed delay of $T = 0.5$. The parameters used are $\beta = 2$, $\sigma = 0.9$, $\alpha = 0.1$, and $p = 0.3$.

1.7.1 First-Order Approximation

The recursion involving an integro-differential equation of the recursive full tracing from proposition 33 can only be solved numerically. In proposition 33, we already noted the independence of the forward and the backward tracing.

When calculating the approximation for backward tracing in proposition 12 and the approximation for forward tracing in proposition 24, the first-order approximation turned out to be an additive correction. We will use this knowledge in proposition 35.

Proposition 35 *The first-order approximation of the zeroth generation reads*

$$\kappa_0(a) = \hat{\kappa}(a)\left\{1 - p\, p_{obs}\beta(1 * \phi * (1 - \hat{\kappa}))(a)\right\} + \mathcal{O}(p^2).$$

If the tracing delay does not have a positive mass at zero, i.e., $\int_0^\epsilon \phi(a')da' \to 0, \quad$ for $\epsilon \to 0$, we can use the approximation from proposition 24 for $i > 0$, where we used that fact to simplify the integration by parts.

$$\kappa_i(a) = \hat{\kappa}(a)\left\{1 - p\, p_{obs}\beta(1 * \phi * (1 - \hat{\kappa}))(a) - p\, p_{obs}(\phi * (1 - \hat{\kappa})(a)\right\} + \mathcal{O}(p^2).$$

Proof: The first-order approximation for $\kappa_0(a)$ consists of $\kappa_-(a)$ as no forward tracing occurs in the zeroth generation. In that case, only the backward tracing takes place. For all following generations, we add the forward tracing approximation from proposition 24

$$\kappa^+(a) = \hat{\kappa}(a) - p\, p_{\text{obs}}\hat{\kappa}(a)((1 - \hat{\kappa}) * \phi)(a)) + \mathcal{O}(p^2).$$

The approximation shows a clear distinction between forward and backward tracing. Both subtract from the total probability of being infectious. $\hfill\square$

The constraint of no mass of the distribution at zero can be removed using the argument of proposition 25 to allow for no delay approximations.

1.7.2 Specific Delays

Like in the case of backward tracing and the case of forward tracing, we can exemplify the computation for κ_i. We already observed the independent additive effect of the forward and backward parts of the tracing in the full tracing approximation in proposition 35.

Fixed Delay

The fixed delay approximation now consists of three cases, because it inherits the case distinction from backward tracing for the zeroth generation.

Proposition 36 (First Order Approximation for Fixed Delay) *Replacing the delay distribution $\phi(a)$ with a fixed delay $\phi(a) = \delta_T(a)$ exemplifies the equation to:*

$$\kappa_i(a) = \hat{\kappa}(a) \times$$

$$
\begin{cases}
\left(1 - p\, p_{obs} \left\{ \beta \left((a - T) - \frac{1 - \hat{\kappa}(a-T)}{\alpha + \sigma} \right) + (1 - \hat{\kappa}(a - T)) \right\} \right) & \text{for } a \geq T, i \geq 1 \\
\left(1 - p\, p_{obs}\, \beta \left\{ (a - T) - \frac{1 - \hat{\kappa}(a-T)}{\alpha + \sigma} \right\} \right) & \text{for } a \geq T, i = 0 \\
1 & \text{for } a < T
\end{cases}
$$

$$+ \, \mathcal{O}(p^2)$$

Proof: In the case of $a < T$, no tracing takes place, and hence $\kappa(a)$ and $\hat{\kappa}(a)$ coincide. In the case of the first generation $i = 0$, no forward tracing occurs, and we follow the backward tracing approximation for fixed delay from proposition 14 for $a \geq T$.

In the case of full tracing for $i > 0$, we add the fixed delay forward tracing approximation from proposition 26 for $a \geq T$. $\qquad\qquad\square$

Exponential Delay

We recall the exponential delay backward tracing from proposition 15 and the forward tracing from proposition 27. Then we use the exponential delay backward and forward tracing to replace the backward and forward tracing in the full tracing approximation from proposition 35.

Proposition 37 (First Order Approximation for Exponential Delay) *The full tracing first approximation with exponential delay for $i > 0$ is stated by:*

$$\kappa_-(a) = \hat{\kappa}(a) - p\, p_{obs}\beta\hat{\kappa}(a) \left\{ a + T\frac{T(\alpha+\sigma)}{1-T(\sigma+\alpha)} \left((1 - e^{-\frac{a}{T}}) - \frac{1-\hat{\kappa}(a)}{(T(\sigma+\alpha))^2} \right) \right\}$$
$$- p\, p_{obs}\hat{\kappa}(a) \left(1 - e^{-\frac{a}{T}} - \frac{\hat{\kappa}(a) - e^{-\frac{a}{T}}}{1 - T(\alpha+\sigma)} \right) + \mathcal{O}(p^2).$$

1.7.3 Reproduction Number

We recall the effective reproduction number from corollary 7 and its approximations in the form of the effective reproduction number approximation in the case of backward tracing corollary 16 and the effective reproduction number approximation in the case of forward tracing corollary 28.

We observed the additive effect of the first-order approximation of the backward and the forward tracing. This property remains after integrating the effect to calculate the reproduction number.

Inserting the first-order approximation of full tracing into the formula to calculate the reproduction number results in the following statement.

Corollary 38 *The effective reproduction number approximation in case of full tracing reads*

$$R_i^{\pm} = \int_0^{\infty} \beta \kappa_i(a) da$$

$$= R_0 - p\, p_{obs} \begin{cases} \{ & \int_0^{\infty} \beta^2 \hat{\kappa}(a)(1 * \phi * (1 - \hat{\kappa}))(a) da \\ + & \int_0^{\infty} \beta \hat{\kappa}(a)((1 - \hat{\kappa}) * \phi)(a) da \} + \mathcal{O}(p^2) \qquad \textit{for } i > 0 \\ \\ \int_0^{\infty} \beta^2 \hat{\kappa}(a)(1 * \phi * (1 - \hat{\kappa}))(a) da + \mathcal{O}(p^2) \quad \textit{for } i = 0 \end{cases} .$$

Because there is no forward tracing for the zeroth generation, the reproduction number approximation for the zeroth generation resembles the reproduction number approximation of backward tracing from corollary 16.

Reproduction Number with Fixed Delay

As before in the backward tracing and forward tracing case, we exemplify the effective reproduction number of the full tracing case employing a fixed delay.

Proposition 39 *The effective reproduction number approximation in case of fixed delay* $\phi(a) = \delta_T(a)$ *full tracing for* $i > 0$ *is given by:*

$$R_i^{\pm,f} = \frac{\beta}{(\alpha + \sigma)} - \frac{1}{2} p p_{obs} \hat{\kappa}(T) \frac{\beta}{(\alpha + \sigma)} \left(\frac{\beta}{(\alpha + \sigma)} + 1 \right) + \mathcal{O}(p^2).$$

Proof: The result follows along the lines of proposition 29 and proposition 17.   □

Corollary 40 *When expressing the fractions $\frac{\beta}{(\alpha+\sigma)}$ by R_0, the effective reproduction number in the case of full tracing for $i > 0$ and fixed delay reads:*

$$R_i^{\pm,f} = R_0 - \frac{1}{2} p\, p_{obs} \hat{\kappa}(T) R_0 (R_0 + 1) + \mathcal{O}(p^2).$$

Reproduction Number with Exponential Delay

Proposition 41 *The effective reproduction number approximation in case of exponential delay $\phi(a) = e^{-\frac{a}{T}}$ full tracing for $i > 0$ is given by:*

$$R_i^{\pm,e} = R_0 - p\, p_{obs} \frac{\beta^2}{(\alpha+\sigma)^2} \frac{1}{2(1+T(a+\sigma))} - p\, p_{obs} \frac{\beta}{\alpha+\sigma} \frac{1}{2(1+T(a+\sigma))} + \mathcal{O}(p^2).$$

Proof: The result follows along the lines of proposition 31 and proposition 19. $\qquad\square$

Corollary 42 *When expressing the fractions $\frac{\beta}{(\alpha+\sigma)}$ by R_0, the effective reproduction number in the case of full tracing for $i > 0$ and exponential delay reads:*

$$R_i^{\pm,e} = R_0 - p\, p_{obs} R_0 (R_0 + 1) \frac{1}{2(1+T(a+\sigma))} + \mathcal{O}(p^2).$$

When comparing the approximations of the effective reproduction number with exponential delay and with fixed delay in the case of backward tracing, we observed an increased effect of the exponential delay. We transferred the conclusion to the forward tracing case. Because the full tracing case adds the effect of forward tracing and backward tracing, we again conclude an increased effect of the exponential delay.

1.8 Summary, Discussion, and Outlook

We introduced backward and forward tracing before combining both tracing directions to achieve full tracing. We modeled one-step and recursive versions and compared them to the first-order approximation, which corresponded to the one-step tracing.

We took the approximation and exemplified it using fixed and exponential delays. This allowed for the comparison of these delays and the analysis of their influence on the effective reproduction number. In the first-order approximation, the effect of backward- and forward tracing behave additively. We used this additive behavior to combine both effects into the full tracing approximations and the reproduction number. That raises the question, which of these additive parts has a greater effect.

There is a long-lasting discussion, if the forward tracing or backward tracing direction is more effective [Koj+21]. If we inspect the two terms, we find that the reproduction number enters the backward term quadratically, while it is affine-linear for the forward tracing case. That might indicate that backward tracing is more important if R_0 is large, while forward tracing is more effective in the case of a relatively small basic reproduction number R_0. However, these conclusions are only valid if p is small. In the case of large p, it is more likely to follow longer infection chains, and in that, backward and forward tracing will be combined. Most likely, the conclusions stated here need to be adapted in that more general case.

In the next chapter, we will extend the onset model to the endemic case by extending the backward tracing to an age-structured model and using the forward tracing approximation to extend the model to cover full tracing.

2 Contact Tracing – The Endemic Case

The onset model for backward tracing in the previous sections concentrates on the onset of an epidemic when we can assume every contactee to be susceptible. The onset model assumes a constant probability of infecting a contactee. When leaving the onset case, not every contact transmits the infection. An infection event can not occur in a contact between two infected individuals, which becomes more likely, the higher the prevalence of a disease. We extend the onset model to an age-structured model to cover the endemic case. In the case of forward tracing, the paper [Sca+21] uses an age-structured model for that purpose. We start using the onset model for one-step backward tracing. The age-structure recovery rate is computed on the base of the probability to be infectious at age of infection as determined above. To cover full tracing, we add the one-step forward tracing approximation as a correction term. Furthermore, we also take the probability and the effect of a contact between two already infected, non-susceptible individuals into account. We call these contacts of two non-susceptible individuals cluster contacts. Cluster contacts increase the tracing probability without infecting new individuals.

In the course of the following chapter, we will describe the endemic model without delay and add the delay in a second step. Then we will compare the results of the endemic model to the IBM simulation and the onset model. In the case of the IBM simulation and the endemic model, this comparison includes the number of susceptible individuals and the probability of being infective $\kappa(a)$ of the endemic case.

Part of the contact are cluster contacts. As both contacts in a cluster contact are already infected, the infection is not transmitted, that is, no new infectee is produced. An infected individual has fewer infectees, and hence the intensity of backward tracing is also reduced. Section 2.3 will investigate this effect in further detail. On the other hand, cluster contact tracing can be added, which replaces the loss of backward tracing, which we investigate in section 2.3. I.e., the loss of backward tracing when entering the endemic case results in a higher probability to be infective, but the added cluster contact tracing compensates this loss resulting in a $\kappa(a)$ of similar shape compared to the $\kappa(a)$ of the onset case. Nevertheless, section 2.4 shows that the age structure of infected individuals differs in the onset and the endemic case. The age structure of the onset case is determined by the exponential

growth. In the endemic case, the age structure is determined by the influence of the tracing because the number of infected individuals reached an equilibrium.

2.1 Model

In the previous chapter, we achieved full tracing in the onset case by adding the backward tracing integro-differential equation result into the recursive process of forward tracing. The backward tracing was a time-independent, but age-dependent result, and the recursive forward tracing resulted in generations.

By adding a time dependency and respecting the current number of infected individuals, we turn the onset model of one-step backward tracing into an age-structured population model suited for the endemic case. The time-dependent nature of the endemic backward tracing model makes it difficult to integrate it into the recursive forward tracing model. On the other hand, we can reduce the forward tracing recursion to one-step tracing. In that case, we can solve the endemic backward tracing model numerically and then apply the correction of the one-step forward tracing. That procedure integrates an approximation of the forward tracing recursion result into the differential equation of the endemic one-step backward tracing model, instead of using the backward tracing model within the forward tracing recursion like in the previous chapter.

We simplify the model by reducing the complexity to one-step tracing. The analysis of the infection tree is localized around the infectee and its infector. The infector always exists, and if the infectee becomes an infector, it has its own infectees. Instead of building the whole tree, we only consider this local structure of the predecessor and successor of the current individual.

The cluster contacts, which emerge in the endemic case, turn the global infection tree, which we considered in the onset case, into a mesh as already infected individuals contact each other. Nevertheless, the local structure of one-step forward tracing or one-step backward tracing only considering the predecessor, the infector, and the possible successor, the infectee, does not observe the mesh. The theory of the previous chapter still works when considering one-step tracing, because the local structure, when considering only direct neighbors of the current individual always resembles a tree and that still holds true, when considering cluster contacts.

In the previous chapter where we investigate the onset case, we observed that especially in the case of low tracing probability, the recursive tracing over multiple levels of the tree becomes unlikely. Considering this, concentrating on one-step tracing and approximating the mesh locally and treat it as a local tree seems to be a feasible approach.

A Note about the Contact Rate β

In the onset model (chapter 1), β incorporated not only the contact rate, but it was a combined contact and infection rate. As we have varying infection rates in the endemic model due to a varying number of susceptible individuals, but a constant contact rate, we have to add a factor respecting the probability of a contacted individual to be susceptible.

2.1.1 Introduction of the Model without Delay

We first introduce an endemic model of contact tracing without delay to explain the transition from the onset model of contact tracing with delay to an endemic model. After this intermediate step, we will add a delay.

Visualizing the Probability of Being Infective

In previous sections, we calculated the likelihood of being infective $\kappa(a)$, which we represent by fig. 2.1. This $\kappa(a)$ was independent of the time t as long as we stayed within the onset of an infection. As we now extend this to the endemic case, we take $\kappa(a)$ and add time dependency to achieve $\kappa(a, t)$.
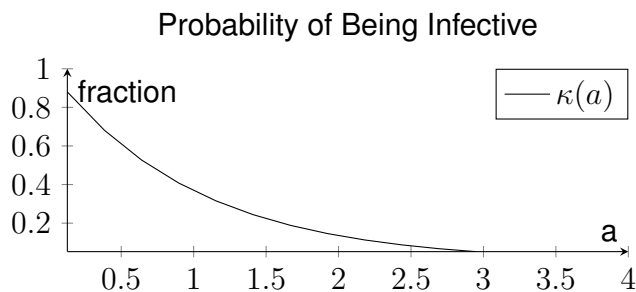
Figure 2.1: An example of the probability of being infective $\kappa(a)$ without tracing.

In the case of the endemic model, the probability of being infective does not only depend on the age since infection, but also on time. The time-dependency is introduced by a time-dependent likelihood of being traced modeled by the tracing probability. This tracing probability is time-dependent because of the time-dependent fraction of susceptible and infective individuals of the total population. The fraction of susceptible and infective individuals influences the spread and the tracing because only contacts between an infective and a susceptible individual spread the infection. However, contacts between infected individuals can result in tracing events, which we call cluster contact tracing.

In fig. 2.2, we introduce cohorts [IM17, p. 9, p. 79]. The cohorts consist of a group of individuals, which became infected at the same time or time period. When the time t progresses, the age a of the cohort progresses at the same time. Figure 2.2 is a scalar field. The probability of being infective corresponding to a cohort can be read at the diagonal starting at the time the cohort was infected[1]. We omitted a three-dimensional plot of the scalar field and annotate the black cohort with a graph showing its corresponding likelihood of being infective.

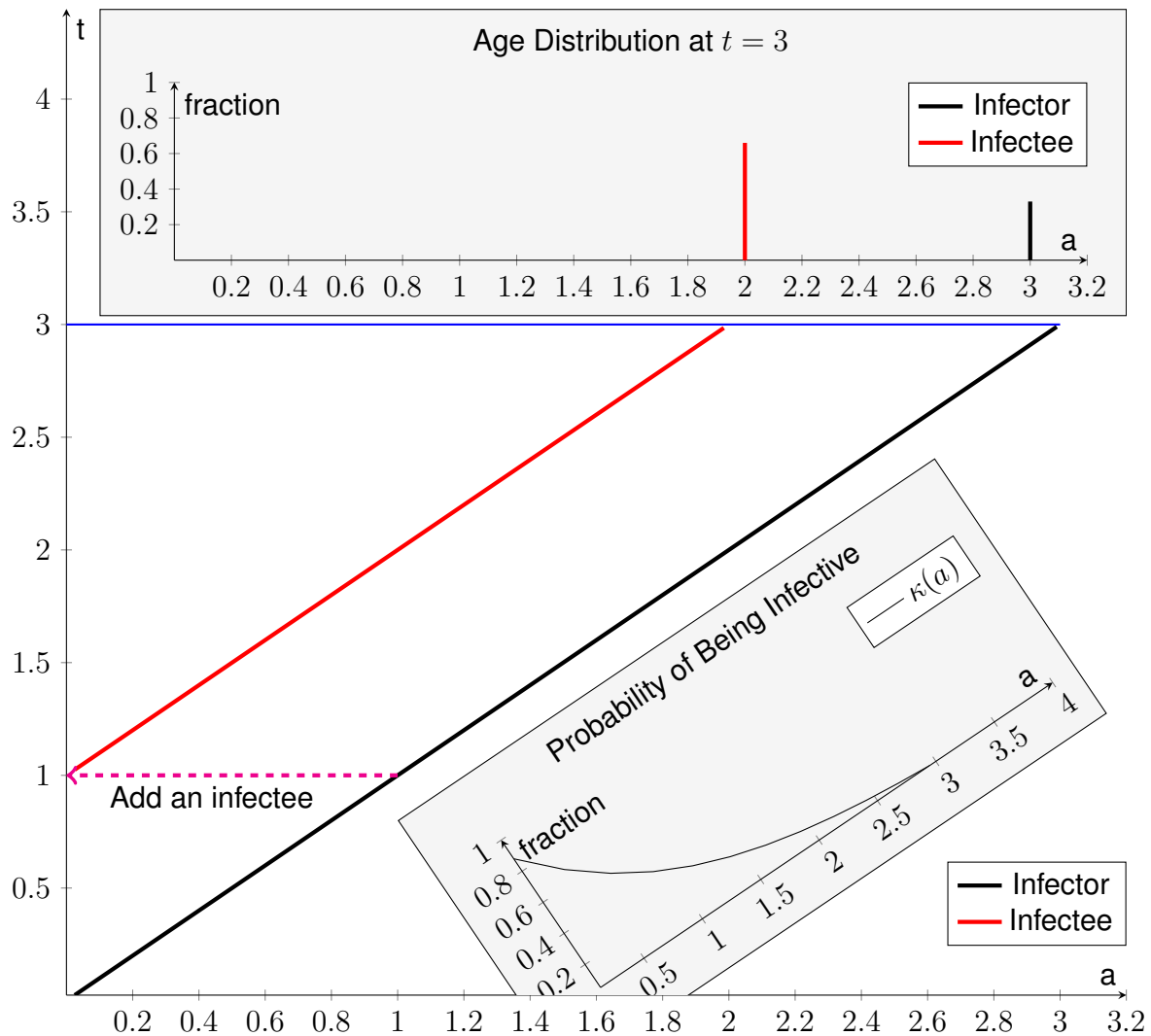[1]This is the characteristic of the corresponding transport equation.

Figure 2.2: The probability of being infective $\kappa(a,t)$ showing the cohorts of the index cases and of an infection at $t = 1$. It shows the values of the likelihood of being infective $\kappa(a,a)$ for one cohort and an normalized evaluation of $\kappa(a,t)$ at time $t = 3$ multiplied by the number of infected individuals per cohort, which resembles an age distribution.

Furthermore, we annotated the age distribution at time t. We can evaluate the number of infected individuals at time t of age a by evaluating the probability of being infective $\kappa(t,a)$ and multiplying it by the number of infected individuals of the cohort at time $t - a$. Plotting the number of infected individuals at time t and normalizing the graph results in the age distribution.

Visualizing an Infection

Beside introducing the cohorts and age distribution, fig. 2.2 shows an example of an infection to relate the cohorts and age distribution within the scalar field.

We visualize the probability of being infective $\kappa(a,t)$ for two selected age groups. The example only consists of supervised σ and unsupervised α recovery. We do not include tracing, time dependency, or conditional probability. At first, we neglect all effects to show the onset case visualized in fig. 2.1 in an endemic setting fig. 2.2 and introduce the notation of visualizing the probability of being infective within the endemic model.

For each index case, there is the probability of being infective. We visualize this probability at time $t = 0$. The characteristic of the infector is shown in black. As time t progresses, the age a increases. Therefore, the individual progresses at a diagonal. The probability of being infective of this index case is $\kappa(a,a) = \kappa(a)$, and it decays exponentially. The identity $t = a$ holds true, because the infection started at time $t = 0$. The likelihood of being infective on this diagonal is plotted next to it as an annotation and resembles the probability of fig. 2.1. This cohort contacts other individuals at a rate β resulting in a continuous increase of new infectees. At time 1 we visualize one of these infections resulting in a new cohort. This visualized infection is an example to represent the continuous process. At the time of infection of an infectee, we know that the infector is infective. This fact affects the infector's probability to be infectious. This does not influence the $\kappa(a)$ of a cohort, but it will be respected later, when implementing the contact tracing. Now, we just add an infectee.

The characteristic of the cohort resembling the infectee is shown in red, and we assume the same exponential decay of the likelihood of being infective as shown in the subgraph $\kappa(a)$.

At time 3 we evaluate $\kappa(a,t)$. We only show the cohorts spawned at 0 and 1. Based on the probability of being still infective we can calculate the fraction of infected individuals per cohort. This is achieved by multiplying the number of initially infected individuals by the corresponding probability of being infective at time $\kappa(a,3)$.

Example: When evaluating the number of infected individuals $i(2,3)$, we have to check the number of newly infected individuals at time $t - a$. These infections at time $t - a$, $i(0, t - a) = 1$, correspond to the evaluated age group a at time t. Assigned to that cohort is a probability of surviving/being infective $\kappa(a,t)$. Multiplying the initial number of infected individuals and their probability of being infective results in the number of surviving/infective individuals.

2.1.2 Model without Delay

This model is an intermediate step before introducing the delay. In opposite to the onset model, the endemic model respects the finite size of the population, because contacts between infected individuals are possible during the progress of the endemic. The probability of contacting a susceptible individual is dependent on the size of the population N, the number of infected individuals $I(t)$ at time t and the resulting number of susceptible individuals $S(t)$. We model the interactions in a SIS model.

In the case of contacts between infected and susceptible individuals, which we treated in the onset model, we adapt for the changing number of infected individuals by multiplying the contact rate β by $\frac{S(t)}{N}$, the fraction of the susceptible individuals at time t within the population N. Furthermore, we introduce contacts between infected individuals. In the case of these cluster contacts, we respect the age distribution of the contacted individuals at the time of contact. We respect the fact that in the case of an II contact, a cluster contact, both individuals are infected. We include the observed and unobserved recovery, the backward tracing and the cluster contact tracing in one differential equation to calculate κ_-, all respected tracing effects beside forward tracing. To approximate forward tracing, we take the first-order approximation of forward tracing without delay from proposition 25 to calculate the probability of being infectious with forward tracing κ_f. We need a gradient on the characteristic of a cohort to progress the number of infected individuals by a forward Euler step. The forward tracing approximation does not provide this gradient, therefore we introduce the derivative along the direction of the characteristic $D_{\mathbb{1}}$. All infected individuals at time t, $I(t)$, then infect new individuals at an infection rate stated by the contact rate, the probability to contact a susceptible individual $\frac{S(t)}{N}$ and the probability of actually infecting the contact π_{IS}.

We first state our model, and provide a derivation afterwards.

Model 43 *Full tracing not only including the onset of an epidemic by considering the probability of contacting already infected individuals can be modeled by the following equations:*

$$
(\partial_t + \partial_a)\kappa_-(a,t) = -\kappa_-(a,t) \Bigg(\quad \alpha + \sigma
$$
$$
+ p \int_0^a \beta \cdot \frac{S(t-c)}{N} \cdot \kappa_-(c,t) \cdot \sigma dc
$$
$$
+ p \int_0^a \beta \int_0^\infty \frac{i(\tilde{a}, t-c)}{N} \cdot \frac{\kappa_-(c+\tilde{a}, t)}{\kappa_-(\tilde{a}, t-c)} \sigma d\tilde{a}dc \quad \Bigg)
$$
$$
\kappa_-(0,t) = 1
$$
$$
\kappa_f(a,t) = \kappa_-(a,t)\left\{1 - p\,p_{obs}(1 - \hat{\kappa}(a))\right\}
$$

$$(\partial_t + \partial_a)i(a,t) = -i(a,t) \cdot \left(\frac{-D_{\mathbb{1}}\kappa_f(a,t)}{\kappa_f(a,t)} \right)$$

$$I(t) = \int_0^\infty i(t,a)da$$

$$i(t,0) = \beta\pi_{IS}\frac{S(t)}{N}I(t)$$

$$S(t) = N - I(t).$$

Derivation: We start assembling the equation for $\kappa(a,t)$ by introducing the observed and unobserved recovery σ and α. The boundary value describing the probability of being infective at the start of the infection is 1:

$$(\partial_t + \partial_a)\kappa_-(a,t) = -\kappa_-(a,t)(\alpha + \sigma),$$

$$\kappa_-(0,t) = 1.$$

This equation will result in an exponential decay of the likelihood of being infective. In the next step, we include direct backward tracing. Direct backward tracing operates on a cohort and incorporates the influence of its infectees. These infectees are created at the contact rate β. While being infective, they are recovered at the observed recovery rate σ recalling the infector at a probability of p. In the onset case, this resulted in the rate $p \cdot \beta \int_0^a \kappa(c)\sigma dc$, because there are infectees at the age 0 to a, which can recover.

Now, introducing the time, we have $\kappa(c,t)$. Because we not only concentrate on the onset of the outbreak, there is a probability of contacting already infected individuals, which lowers the probability of producing infectees. Figure 2.3 exemplifies the contact, considering a fraction of the population being non-susceptible. The cohort $k(a,t)$ was produced at time $t - a$. We will incorporate this effect by respecting the fraction of susceptible individuals among the population at that time, i.e., $\frac{S(t-c)}{N}$.

$$p \int_0^a \beta \cdot \frac{S(t-c)}{N} \cdot \kappa_-(c,t) \cdot \sigma dc.$$

The neglected contact between already infected individuals has to be respected now. We will call this contact between infected individuals cluster contact. Cluster contacts only influence the tracing rate. Although no new individuals were infected, this contact can be recalled during an independent tracing event leading to accidental detections.

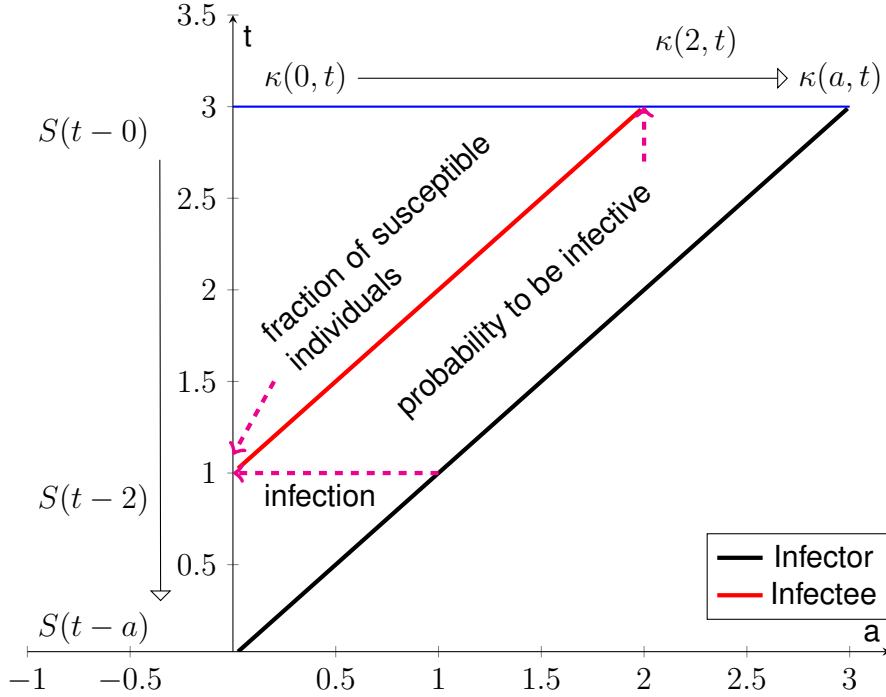These cluster contacts result in less backward tracing, because there are fewer new infec-

Figure 2.3: The cohorts of an infective contact, which results in tracing events without delay: $p \int_0^a \beta \cdot \frac{S(t-c)}{N} \kappa_-(c,t) \cdot \sigma dc$

tions, because there are fewer susceptible individuals. This decrease in the rate of backward tracing will be compensated by tracing events due to cluster contacts, i.e., tracing by other already infected contacts.

Contrary to the classic backward tracing, cluster contacts do not start an infection and the individuals have age since infection $a > 0$ at the time of the contact. Furthermore, there is no limit on the age of an infected individual. This undefined age means we do not only have to respect the fraction of infected individuals of the population at the time of the contact taking place $\frac{I(t-c)}{N}$ to distinguish cluster contacts from infective contacts.

Additionally, we have to take into account the age distribution $i_P(\tilde{a}, t)$ of the contactee $i_P(\tilde{a}, t - c) = \frac{i(\tilde{a}, t-c)}{\int_0^\infty i(a', t-c)da'} = \frac{i(\tilde{a}, t-c)}{I(t-c)}$ at the time of the contact. Figure 2.4 visualizes an example of such an age distribution of the contactee within a cluster contact taking place at time $t = 1$. At the time of a possible tracing event, when evaluating the corresponding infectivity likelihood $\kappa(c + \tilde{a}, t)$, which is distributed by the age distribution, we have to normalize the likelihood to be one at the time of the contact by dividing by $\kappa_-(\tilde{a}, t - c)$:

$$p \int_0^a \beta \frac{I(t-c)}{N} \int_0^\infty \frac{i(\tilde{a}, t-c)}{\int_0^\infty i(a', t-c)da'} \cdot \frac{\kappa_-(c+\tilde{a}, t)}{\kappa_-(\tilde{a}, t-c)} \sigma d\tilde{a}dc$$

$$= p \int_0^a \beta \frac{1}{N} \int_0^\infty i(\tilde{a}, t-c) \cdot \frac{\kappa_-(c+\tilde{a}, t)}{\kappa_-(\tilde{a}, t-c)} \sigma d\tilde{a}dc.$$
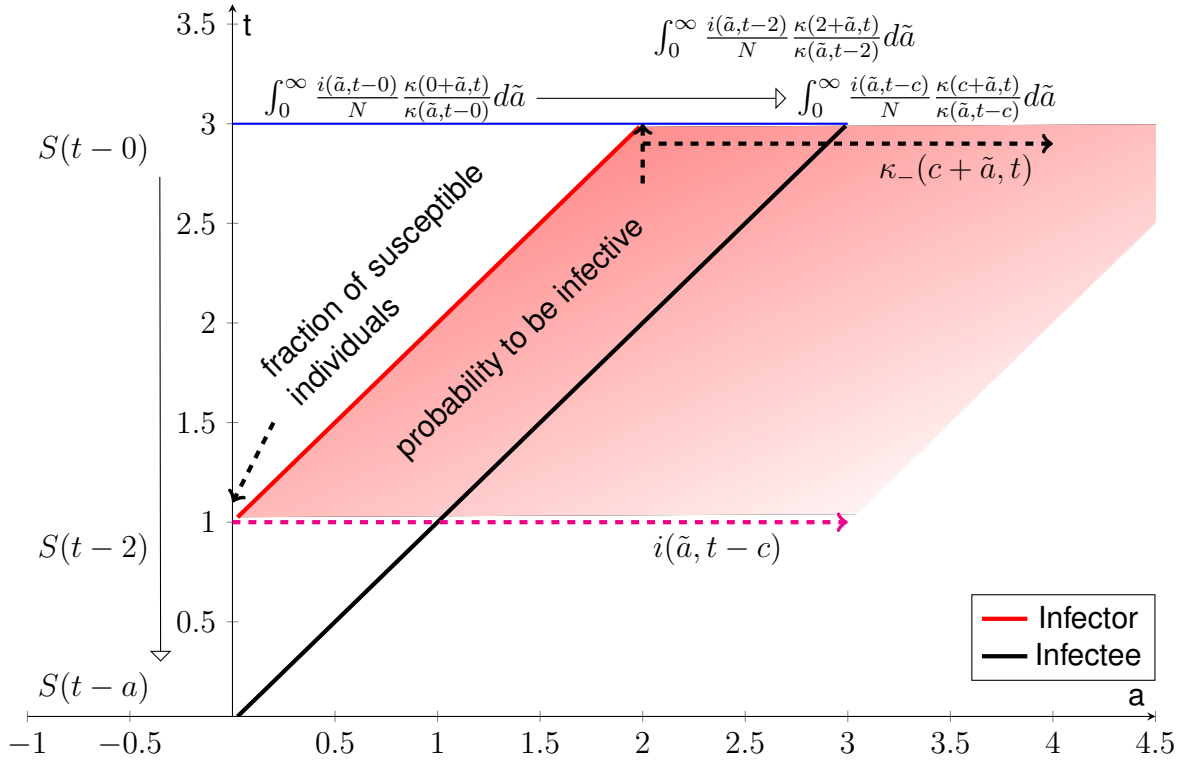
Figure 2.4: A cluster contact, which results in tracing events. In opposite to an infective contact, we don't have a single probability $\kappa(a,t)$ of an infectee, which was infected at the time of contact. Now, we have an integral over all $\kappa(c+\tilde{a},t)$ of possible cohorts $i(\tilde{a}, t-c)$ of the contactee at the time of contact $t-c$. Therefore, we have to normalize the probability at the time of contact to 1 by dividing each probability by $\kappa(\tilde{a}, t-c)$.

Furthermore, $i(\tilde{a}, t-c)$ can then be written as a product of the initial size of the cohort $i(0, t-c-\tilde{a})$ multiplied by their survival probability $\kappa(\tilde{a}, t-c)$.

$$= p \int_0^a \beta \frac{1}{N} \int_0^\infty i(0, t-c-\tilde{a}) \cdot \kappa_-(c+\tilde{a}, t)\sigma d\tilde{a}dc.$$

Now, the forward tracing step is applied. Therefore, we remember the one-step approximation of the recursive full tracing from proposition 35:

$$\kappa_i(a) = \hat{\kappa}(a) \left\{ 1 - p\, p_{\text{obs}}\beta(1 * \phi * (1-\hat{\kappa}))(a) - \underbrace{p\, p_{\text{obs}}(\phi * (1-\hat{\kappa}))(a)}_{\text{forward tracing}} \right\}.$$

In the onset case, the forward tracing part of this approximation can be applied as a linear correction term to one-step backward tracing. This method is now used to add the one-step forward tracing correction into the endemic equation, to get an approximation for the full tracing $\kappa_f(a, t)$.

The approximation of the correction consists of $\kappa_-(a, t)$, the probability of the infectee to be still infective. This probability is multiplied by the approximated probability of the infector being recovered $(1 - \hat{\kappa})(a)$, which is multiplied by the probability $p_{\text{obs}} = \frac{\sigma}{\sigma + \alpha}$ of being an observed recovery and the probability p, which is the probability of recalling the contact during a tracing event.

$$
\begin{aligned}
\kappa_f(a, t) &= \kappa_-(a, t) \cdot P(\text{not traced via infector}) \\
&= \kappa_-(a, t) \cdot (1 - P(\text{traced via infector})) \\
&= \kappa_-(a, t) \cdot (1 - P(\text{infector not infectious}) \cdot p_{\text{obs}} \cdot p) \\
&\approx \kappa_-(a, t) \cdot (1 - (1 - \hat{\kappa}(a)) \cdot p_{\text{obs}} \cdot p) \\
&= \kappa_-(a, t) - \kappa_-(a, t)(1 - \hat{\kappa}(a)) \cdot p_{\text{obs}} \cdot p.
\end{aligned}
$$

Using this new probability of being infective in case of full tracing $\kappa_f(a, t)$, we calculate our hazard rate. The hazard rate is the normalized change of infected individuals in the direction of the characteristic $(1, 1)$ of a cohort $\frac{-D_{\mathbb{1}} \kappa_f(a,t)}{\kappa_f(a,t)}$, where $D_{\mathbb{1}} \kappa_f(a, t) = \left. \frac{\partial}{\partial \tau} \kappa_f(a + \tau, t + \tau) \right|_{\tau = 0}$. Using this information, the PDE for the distribution of the infected individuals $i(a, t)$ can be constructed:

$$
(\partial_t + \partial_a) i(a, t) = -i(a, t) \cdot \left(\frac{-D_{\mathbb{1}} \kappa_f(a, t)}{\kappa_f(a, t)} \right).
$$

Obviously, the integral over the infected individuals at time t results in the total number of infected individuals at that time:

$$
I(t) = \int_0^\infty i(t, a) da.
$$

The boundary value for the distribution of the infected individuals is the number of newly infected individuals. The number of newly infected individuals is calculated by the contact rate β, the probability of infecting a contact π_{IS}, which will be set to 1 within the thesis, and the probability of the other contact being susceptible $\frac{S(t)}{N}$.

$$
i(t, 0) = \beta \pi_{IS} \frac{S(t)}{N} I(t).
$$

Finally, calculating the number of susceptible individuals knowing the number of infected individuals is trivial:

$$S(t) = N - I(t).$$

$\square$

2.1.3 Introduction of the Model with Delay

The previous section was introducing the model without delay to provide an overview before starting more complicated operations. In this section, we will introduce the tracing delay. In fig. 2.5 we visualize an example after an individual recalls an infective contact or a cluster contact, there is a tracing delay before the identified individual is treated.
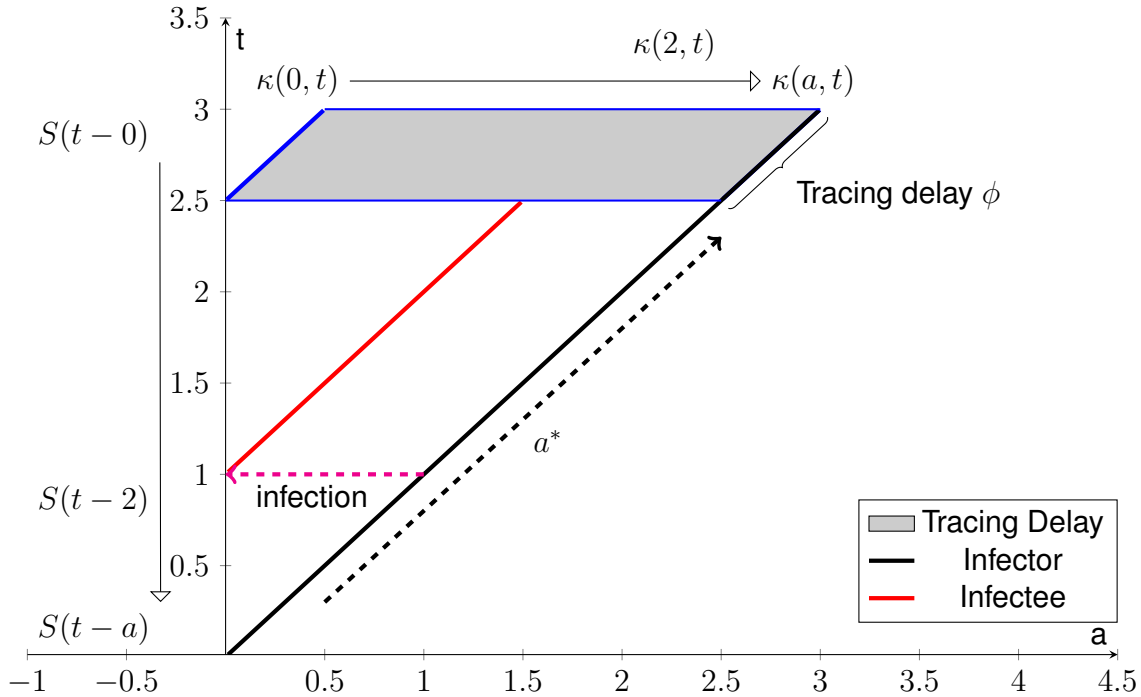
Figure 2.5: An infective contact, which results in tracing events. After the red infectee is detected it triggers a tracing event. The gray area visualizes the tracing delay. The age of the black cohort of the infector is denoted by a^*. A diagonal line visualizes the cohort because the time progresses while the age of individuals within a^* grows.

Figure 2.5 shows the delay in the direction of the cohorts, a diagonal direction. It does not delay in a single direction like age or time. Our coordinates live in the direction a and t. Within the cohort, we introduce a^*. When aging $a^* + 1$, it means $a + 1$ and $t + 1$.

Definition 44 Convolution in the direction of the characteristic of a cohort.

$$(\phi * f)(a, t) := (\phi(\cdot) * f(a + \cdot, t + \cdot))(a, t)$$
$$= \int_0^a \phi(\tau) \cdot f(a - \tau, t - \tau)d\tau$$

2.1.4 Model with Delay

Using the previous definition and the tracing defined in the previous section, we derive the following model.

Model 45 (Tracing with Delay) *This model includes full tracing with tracing delay. It not only includes the onset of an epidemic by considering the probability of contacting already infected individuals. The tracing delay influences the backward tracing T_b and the cluster contacts T_c. Furthermore, it influences the approximation of the forward tracing correction T_{fc}.*

$$T_b(a,t) = p \int_0^a \beta \cdot \frac{S(t-c)}{N} \cdot \kappa_-(c,t) \cdot \sigma dc$$

$$T_c(a,t) = p \int_0^a \beta \int_0^\infty \frac{i(\tilde{a}, t-c)}{N} \cdot \frac{\kappa_-(c+\tilde{a}, t)}{\kappa_-(\tilde{a}, t-c)} \sigma d\tilde{a} dc$$

$$T_{fc}(a,t) = p \cdot p_{obs} \cdot ((1-\hat{\kappa}))(a), \qquad a \geq 0$$

$$T_{fc}(a,t) := 0, \qquad a < 0$$

$$(\partial_t + \partial_a)\kappa_-(a,t) = -\kappa_-(a,t)\left(\alpha + \sigma + (\phi * T_b)(a,t) + (\phi * T_c)(a,t)\right)$$

$$\kappa_-(0,t) = 1$$

$$\kappa_f(a,t) = \kappa_-(a,t)\left\{1 - (\phi * T_{fc})(a,t)\right\}$$

$$(\partial_t + \partial_a)i(a,t) = -i(a,t) \cdot \left(\frac{-D_{\mathbb{1}}\kappa_f(a,t)}{\kappa_f(a,t)}\right)$$

$$I(t) = \int_0^\infty i(t,a)da$$

$$i(t,0) = \beta \pi_{IS} \frac{S(t)}{N} I(t)$$

$$S(t) = N - I(t).$$

In the following chapter, we will deal with constant delays T. Therefore, we introduce the model with constant delays and explain how to introduce the delay and evaluate the constant delay.

Proposition 46 (Tracing with Constant Delay) *Full tracing with constant tracing delay T, not only including the onset of an epidemic by considering the probability of contacting already infected individuals, can be modeled by the following equations:*

$$(\partial_t + \partial_a)\kappa_-(a,t) = -\kappa_-(a,t)\left(\alpha + \sigma\right.$$

$$+ p\int_0^{a-T}\beta \cdot \frac{S(t-T-c)}{N}\cdot\kappa_-(c,t-T)\cdot\sigma dc$$

$$\left.+ p\int_0^{a-T}\beta\int_0^\infty \frac{i(\tilde{a},t-T-c)}{N}\cdot\frac{\kappa_-(c+\tilde{a},t-T)}{\kappa_-(\tilde{a},t-c-T)}\sigma d\tilde{a}dc\right)$$

$$\kappa_-(0,t) = 1$$

$$\kappa_f(a,t) = \kappa_-(a,t)\left\{1 - p\,p_{obs}((1-\hat{\kappa}))(a-T)\right\}$$

$$(\partial_t + \partial_a)i(a,t) = -i(a,t)\cdot\left(\frac{-D_1\kappa_f(a,t)}{\kappa_f(a,t)}\right)$$

$$I(t) = \int_0^\infty i(t,a)da$$

$$i(t,0) = \beta\pi_{IS}\frac{S(t)}{N}I(t)$$

$$S(t) = N - I(t).$$

Proof: As before, we start assembling the equation for $\kappa(a,t)$ by introducing the observed and unobserved recovery σ and α. The boundary value describing the probability of being infective at the start of the infection is 1.

$$(\partial_t + \partial_a)\kappa_-(a,t) = -\kappa_-(a,t)\left(\alpha + \sigma\right)$$

$$\kappa_-(0,t) = 1.$$

In the next step, we include direct backward tracing respecting the likelihood of contacting a susceptible individual as we did in the previous model 43 and call it T_b before adding the delay. T_b consists of the probability p of recalling the infector, the contact rate β, the rate of supervised recovery σ and an integral integrating the time dependent probability of being infectious $\kappa_-(a,t)$ and the probability of contacting a susceptible individual $S(t)$ since the start of the infection till the current age a.

$$T_b(a,t) = p\int_0^a \beta \cdot \frac{S(t-c)}{N}\cdot\kappa_-(c,t)\cdot\sigma dc.$$

Now use the definition 44 of the convolution along the cohort to calculate the tracing delay ϕ_T.

$$\phi_t * T_b(a,t) = T_b(a-T, t-T)$$

$$= p \int_0^{a-T} \beta \cdot \frac{S(t-c-T)}{N} \cdot \kappa_-(c, t-T) \cdot \sigma dc.$$

When modeling the cluster contacts $T_c(a,t)$, we have to respect the probability of meeting an already infected individual $\frac{I(t-c)}{N}$. Additionally, we have to respect the age distribution $i_P(\tilde{a}, t)$ at the time $t-c$ described by $i_P(\tilde{a}, t-c) = \frac{i(\tilde{a}, t-c)}{\int_0^\infty i(a', t-c)da'} = \frac{i(\tilde{a}, t-c)}{I(t-c)}$ and the corresponding infectivity likelihood $\kappa(c+\tilde{a}, t)$ of the contacted individual, which is distributed by the age distribution. We visualize an example cluster contact in fig. 2.6. As we define the cluster contacts as contacts between infected individuals, we normalize this infectivity likelihood of a cohort of age $\tilde{a}$ at the time of the contact to 1 at the time of the contact $t-c$ by dividing the infectivity likelihood by $\kappa(\tilde{a}, t-c)$.
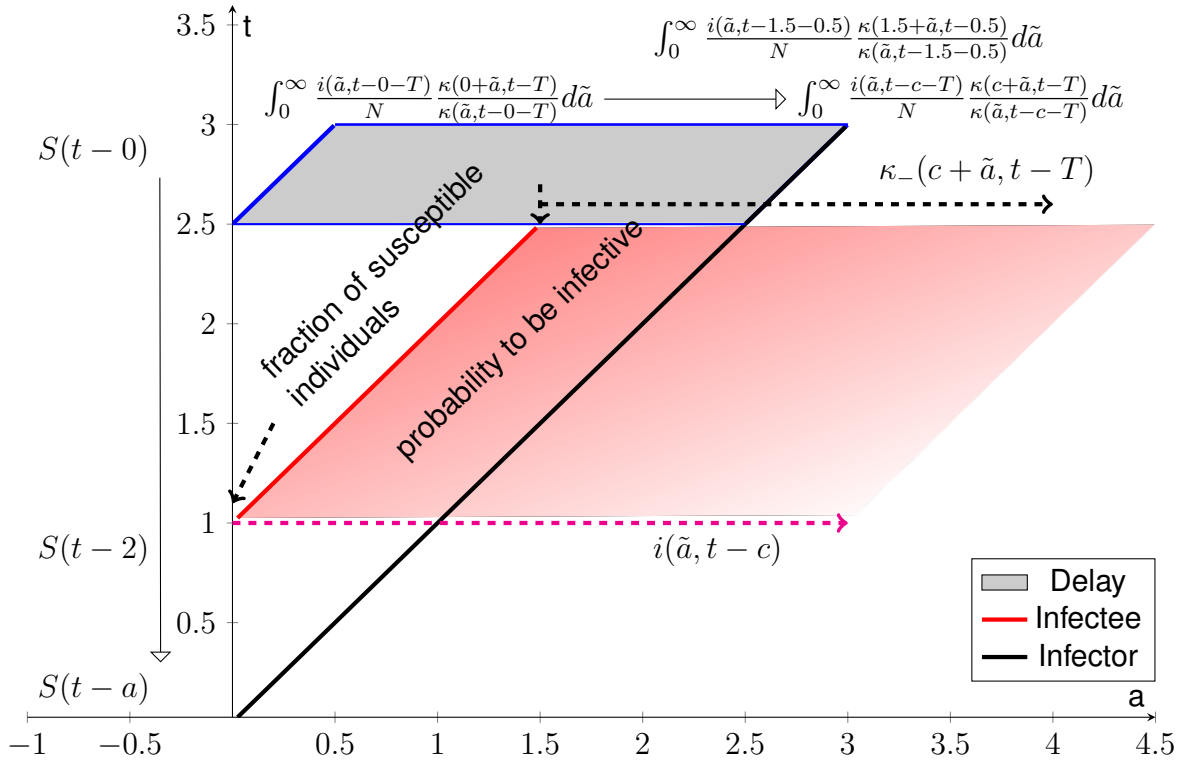
Figure 2.6: A cluster contact, which results in a tracing event similar to fig. 2.4. This time, there is a constant tracing delay, which introduces a delay between the detection and the influence of the tracing event. In opposite to an infective contact, we do not have $\kappa(a,t)$ of an infectee, which was infected at the time of contact. Now, we have an integral over all $\kappa(c+\tilde{a}, t-T)$ of possible cohorts $i(\tilde{a}, t-c-T)$ of the contactee at the time of contact $t-c-T$. Dividing by $\kappa_-(\tilde{a}, t-c-T)$ respects the conditional probability, because the contactee of the cluster contact is infective at that time, which is independent of its age.

$$T_c(a,t) = p \int_0^a \beta \frac{I(t-c)}{N} \int_0^\infty \frac{i(\tilde{a}, t-c)}{\int_0^\infty i(a', t-c)da'} \cdot \frac{\kappa_-(c+\tilde{a}, t)}{\kappa(\tilde{a}, t-c)} \sigma d\tilde{a} dc$$

$$= p \int_0^a \beta \frac{1}{N} \int_0^\infty i(\tilde{a}, t-c) \cdot \frac{\kappa_-(c+\tilde{a}, t)}{\kappa(\tilde{a}, t-c)} \sigma d\tilde{a} dc.$$

Again use the convolution from definition 44 to calculate the tracing delay ϕ_T.

$$\phi_t * T_c(a,t) = T_c(a-T, t-T)$$

$$= p \int_0^{a-T} \beta \frac{1}{N} \int_0^\infty i(\tilde{a}, t-c-T) \cdot \frac{\kappa_-(c+\tilde{a}, t-T)}{\kappa(\tilde{a}, t-c-T)} \sigma d\tilde{a} dc.$$

Now, the forward tracing step from proposition 35 is applied as in model 43, the model without delay. The correction consists of the probability of the infectee still being alive $\kappa_-(a,t)$, and the infector having an observed recovery, i.e., is not infective anymore $(1 - \hat{\kappa}(a))$ multiplied by the probability of an observed recovery $\frac{\sigma}{\sigma+\alpha}$ and the probability of recalling the contact p. Taking the tracing delay into account, which was started by the observed recovery of the infector, we get $p \cdot p_{\text{obs}}(1 - \hat{\kappa}(a-T))$.

$$\kappa_f(a,t) = \kappa_-(a,t) - \kappa_-(a,t) \cdot p \cdot p_{\text{obs}}(1 - \hat{\kappa}(a-T)).$$

Using this new probability of being infective in case of full tracing $\kappa_f(a,t)$, we calculate our hazard rate. The hazard rate is the normalized change of infected individuals in the direction of the characteristic $(1,1)$ of a cohort $\frac{-D_{\mathbb{1}}\kappa_f(a,t)}{\kappa_f(a,t)}$. Using this information, the PDE for the distribution of the infected individuals $i(a,t)$ can be constructed:

$$(\partial_t + \partial_a)i(a,t) = -i(a,t) \cdot \left(\frac{-D_{\mathbb{1}}\kappa_f(a,t)}{\kappa_f(a,t)} \right).$$

The rest of the model is identical to the model without delay. □

2.2 Compare Onset, Endemic, and IBM Model

This section will compare the IBM simulation[1], the onset model, and the endemic model output to show the fit between the onset model, the endemic model and the IBM simulation in the onset and in the endemic case.

We will start by taking a look at full tracing with cluster contacts and delay. To understand the accuracy of the endemic model in case of backward tracing and the approximation in case of forward tracing, we also compare the backward and forward tracing case.
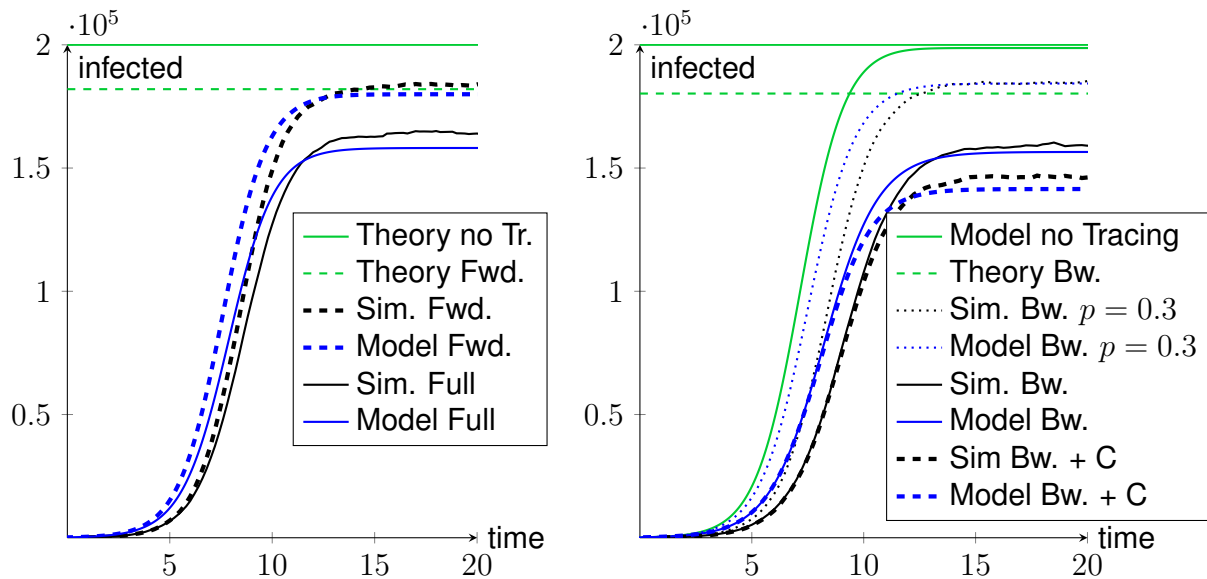
Figure 2.7: Left: Full tracing with cluster contacts (solid line) and forward tracing (dashed line) with a recall probability $p = 0.3$, a delay $T = 0.5$, a contact rate $\beta = 2$, supervised recovery rate $\sigma = 0.9$, and unsupervised recovery rate $\alpha = 0.1$. The simulation is shown in black, the numerical result of the endemic model in blue and for comparison the theoretical limits without tracing and respecting the forward tracing approximation in green.
Right: The solid black line shows the backward tracing simulation without cluster contacts and a recall probability $p = 1$ to enlarge the effect of tracing. The solid blue line shows the numeric result, the dashed blue line adds cluster contacts, while the dotted blue line shows backward tracing without cluster contacts but with $p = 0.3$ for better comparability.

Figure 2.7 shows the graphs of infected individuals. These graphs permit the classification of time points into the onset of the infection, the endemic case, and the transition between these states. Exponential growth characterizes the onset. When collecting the information to calculate the probability of being infective, the individuals' whole lifespan must be respected. The equilibrium indicates the endemic case. The time, which can not be assigned

---

[1]A description of the simulation based on an individual-based model, IBM, can be found in Appendix 2 – Programs section 6.1.1.

to one of these cases will be the transition.

The left graph shows the forward tracing and the full tracing case. Both cases use a recall probability of $p = 0.3$. This value is a compromise. Low values of p within the approximation result in a small impact of tracing but a higher accuracy. In the case of lower values of p, the higher accuracy results from a higher accuracy of the approximations used, and the higher accuracy results from less influence of the different generations of the tracing. The simulation does not separate generations. Thus, traced individuals are prohibited from triggering a tracing event themselves in the one-step tracing case. But, the higher the recall probability, the higher the impact, which is easier to visualize.

On the other hand, the analytic approximation enables us to calculate analytic limits in the equilibrium case, which become inaccurate for bigger values of p. Using the formula to calculate the reproduction number $R = \int_0^\infty \beta \hat{\kappa}(a) da$ for the onset case, we can derive an equation from the reproduction number to fit the endemic case by adding the probability of contacting a susceptible individual $\frac{N-I(t)}{N}$. Furthermore, we know that the equilibrium case is characterized by a reproduction number of exactly one. On average, every infected individual infects another individual. This process results in a constant number of infected individuals. The resulting equation can be solved to achieve $I(t)$, for times t representing the equilibrium case[1].

Remark 47 (Infected Individuals Calculated from Reproduction Number) *We know the formula for the reproduction number. The reproduction number equals 1 in the endemic case. Therefore, we can solve the formula $1 = R = \int_0^\infty \beta \frac{N-I(t)}{N} \hat{\kappa}(a) da$ for $I(t)$ in the endemic case to calculate the number of infected individuals.*

2.2.1 Overall Match in Case of Full Tracing with Cluster Contacts

The black and blue solid lines in fig. 2.7 (left panel) show the simulation and numeric solution of the endemic model, illustrating the number of infected individuals during an outbreak and the steady state of an epidemic. The solid lines cover full tracing, i.e., forward and backward tracing, and also cluster contacts are traced.

Figure 2.8 shows the corresponding probability of being infective $\kappa(a)$ at the age of infection a.

In the onset case, we note the very good match of the endemic model and the onset model. The logarithmic scale shows some divergence of the simulation when $\kappa(a)$ approaches zero.

---

[1]In the chapter 5 'Appendix' section 5.1.1 describes how to calculate the equilibrium case without tracing, with forward or with backward tracing.
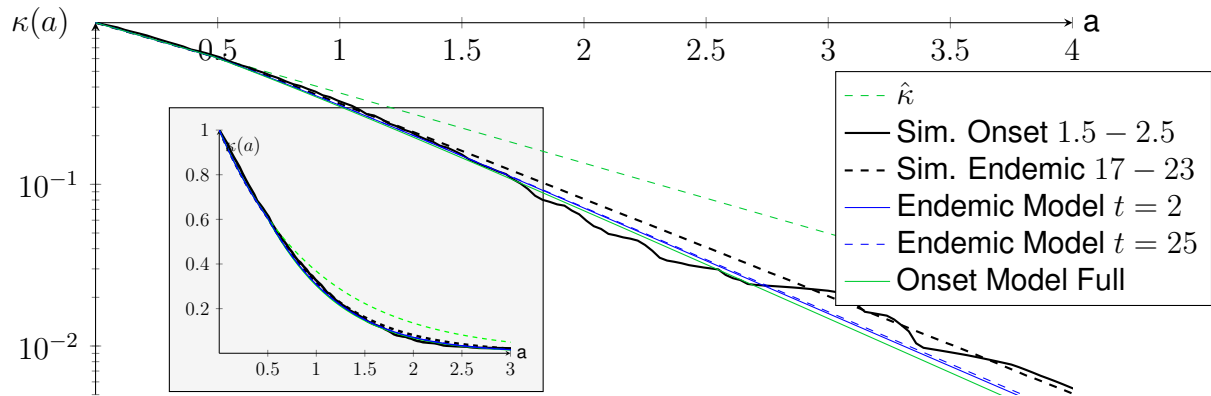
Figure 2.8: $\kappa(a)$ of full tracing with cluster contacts and delay of $T = 0.5$. Furthermore, unsupervised recovery rate $\alpha = 0.1$, supervised recovery rate $\sigma = 0.9$, contact rate $\beta = 2$. The endemic model onset cohort starts at time 2; the endemic model endemic cohort starts at age 25. The simulation samples are drawn between time 1.5 and 2.5, and between 17 and 23, when the spread of the infection reaches a steady state.

Discussion

Overall, the match of the infection probabilities is satisfying. When concentrating on the simulation of the endemic case, there are effects not covered in the mathematical model or the simulation. Let us review some of them.

We do not cover reinfections. In a population having such a high number of infected people, which recover, some traced people will have been recovered and infected again until they are traced. We programmed the IBM simulation to ignore these contacts during tracing to disable this effect if the infector of the infectee changed since the initial contact. In the sense of tracing, this are lost contacts resulting in less effective tracing. The same is true for one-step tracing. Infectees or infectors removed by one-step tracing can not trigger tracing events themselves. We do not respect this effect for backward tracing and in the case of forward tracing, the IBM simulation always considers a mix of generations.

A certain part of the small divergence in the graph of infected individuals of full and forward tracing can be traced to the forward tracing being an approximation, which slightly overestimates the effect.

To better understand the effects, the next sections 2.2.2 and 2.2.3 will show the fit of the endemic model concentrating on forward and backward tracing only. Afterward, we will focus on adapting the onset case to the endemic case and showing the influence of cluster contacts during this transition.
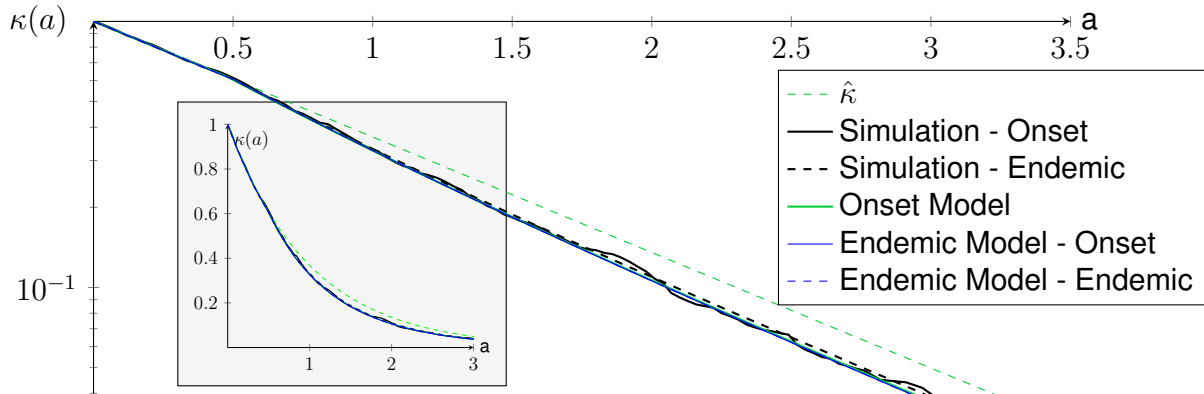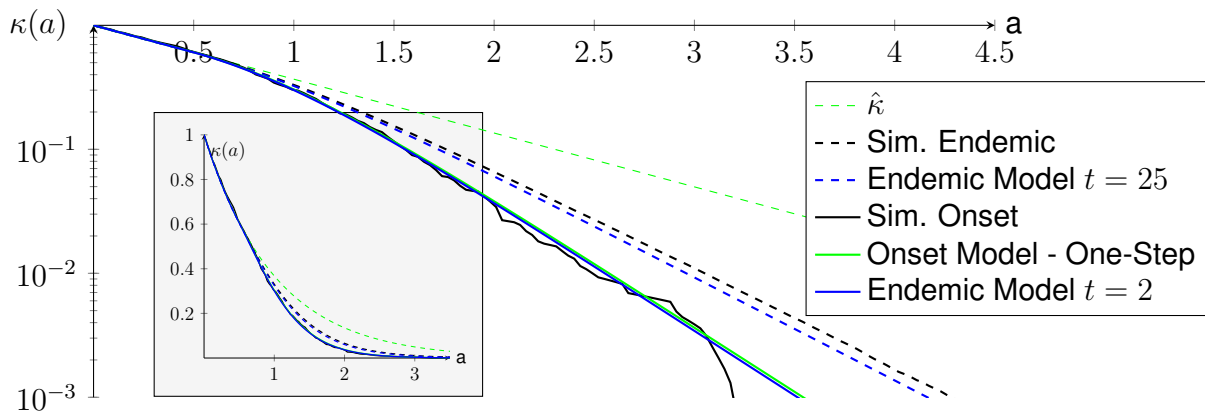
2.2.2 Forward Tracing

Figure 2.9: $\kappa^+(a)$ of forward tracing and delay of $T = 0.5$. Furthermore, unsupervised recovery rate $\alpha = 0.1$, supervised recovery rate $\sigma = 0.9$, contact rate $\beta = 2$. The endemic model onset cohort starts at time 2; the endemic model endemic cohort starts at age 25. In the case of the simulation, the samples are drawn between time 1.5 and 2.5 and between 17 and 23, when the spread reached a steady state.

The number of infected individuals of the simulation and the endemic model displayed in fig. 2.7 shows a small divergence around the theoretical limit. We calculate this theoretical limit in the appendix lemma 83.

When examining the probabilities of being infective after the age of infection a, the function $\kappa(a)$ from fig. 2.9 shows a very good resemblance of the onset model, endemic model, and simulation in the onset and the endemic case. Furthermore, the onset and endemic case are very similar.

Discussion

The number of infected individuals not being exact can be explained by the approximation used. The forward tracing correction within the endemic model from model 45 is done by an analytical term approximating the forward tracing derived in proposition 24. Because this is an approximation, it slightly overestimates the tracing and the approximation does not respect the mix of generations, which present during every epoch of the IBM simulation. Though, the low value of $p = 0.3$ successfully reduces these effects. In fig. 2.9, the divergence is neglectable and becomes visible in fig. 2.7, because the number of infected individuals depends on the integral of the probability to be infective and thus integrates the error.

Nevertheless, the forward tracing approximation is a very good approximation for $\kappa(a)$ for the onset and the endemic case.

The similarity between the onset and the endemic case of the model is evident due to the same approximation being used, which is relatively independent of the number of infected individuals. I.e., the impact of the forward tracing within the simulation does not change from the onset to the endemic case. The same is true for the IBM simulation. In the case of forward tracing, this can be explained. Because in the case of forward tracing, $\kappa(a)$ is largely independent of the contact rate and the probability of susceptible contacts. Every infectee has exactly one infector, which can trace his infectee. This process does not change when the epidemic changes from the onset to the endemic case.

2.2.3 Backward Tracing – Separation of the Different States

Figure 2.10: $\kappa(a)_-$ of backward tracing and delay of $T = 0.5$. Furthermore, unsupervised recovery rate $\alpha = 0.1$, supervised recovery rate $\sigma = 0.9$, contact rate $\beta = 2$, and $p = 1$. The endemic model onset cohort starts at time 2; the endemic model endemic cohort starts at age 25. In the case of the simulation, the samples are drawn between time 1.5 and 2.5 and between 17 and 23, when the spread reached a steady state.

Again, we see a good resemblance between the model and the simulation. In opposite to the forward tracing shown in section 2.2.2, the backward tracing in fig. 2.10 shows a clear separation between the onset and the endemic case. In section 2.3, we will explain this separation, and we will show in the paragraph 'Cluster Contact Tracing Substitutes Lost IS Tracing' that cluster contacts remove this clear separation.

Discussion

The separation of the onset and endemic cases in the case of backward tracing can be explained by the rate of contacting susceptible individuals. During the onset case, the

number of infected individuals is small. Every contact can be expected to be an IS contact, which results in a new infection[1].

In the case of backward tracing, more infective contacts result in more possibilities to be traced. When the reproduction number in the endemic case drops to 1, every infected individual can on average only be traced by a single infectee. This low average number of backward tracing cases is lower than the contact rate $\beta = 2$ during the onset case.

The deviation between the endemic cases of the simulation and the endemic model shows a slightly higher probability of still being infective in the simulation case. This deviation is only visible on the logarithmic scale and does not obfuscate the separation of the onset case and the endemic case. A big part of it can be traced back to the one-step tracing. Traced individuals can not trigger a tracing event themselves. That explains that the difference gradually starts at around $2 \cdot T$. There is the tracing delay to detect the infectee by backward tracing. If this infectee wasn't removed by tracing, it could trigger a tracing event influencing the current individual after another delay of T.

2.2.4 Overall Discussion

The full tracing shows good accuracy while there is a visible deviation of $\kappa(a)$ and a small deviation in the number of infected individuals. When taking a closer look at the $\kappa(a)$ of backward and forward tracing, backward tracing shows only a slight deviation in the endemic case. We note, the backward tracing is separated in the onset and the endemic case. On the other hand, in the case of the number of infected individuals of the full tracing, the forward tracing shows the limitations of the approximation, which is responsible for a big part of the small deviation visible in the number of infected individuals.

The next subsections will cover some of the observed effects. Section 2.3 will show how the backward tracing of the endemic model changes when leaving the onset case, which is equivalent to the onset model, and entering the endemic case. We will then proceed to the absence of this separation in the case of additional cluster contact tracing.

These cluster contacts added to combined forward and backward tracing result in full tracing with cluster contacts. We observe cluster contacts to increase another slight inaccuracy of $\kappa(a)$, because the model diverges slightly from the simulation. If a high percentage of the population is infected, there is a possibility that a cluster contact was recovered and infected again before a tracing event started. We call this effect reinfection and do not model this. The IBM simulation does not trace contacts, which have been infected by a third individual

---

[1]When ignoring the tracing, respecting an infection probability is equivalent to adapting different values for β. Contacts not resulting in an infection will not be traced. They are equivalent to contacts not taking place, and the model can therefore neglect these contacts to model the spread.

since the contact considered in the tracing event, to match the endemic model as close as possible.

Although not influencing the tracing directly, it does have an impact, because it does reduce the effective number of contacts. On the other hand, we conclude that it is only a minor effect taking place at unrealistic high infection rates. In the case of such a high percentage of the population being infected, other methods than contact tracing might be more successful and should be considered. One possible alternative to contact tracing might be identifying and sampling core groups, so-called clustering. At high percentages of infection, the probability of drawing an infected individual might be comparable to the probability of an infected individual to recall his contacts.

Finally, section 2.4 will show that similar $\kappa(a)$ do not mean similar age structures in the onset and the endemic case.

2.3 The Adaption to the Endemic Case

The endemic model (model 45) adapts the onset model (proposition 33 'Full Tracing') to the endemic case. We describe the characteristics of this adaption. Because forward tracing is independent of the onset or endemic case, there is no forward tracing adaption to analyze. Furthermore, the endemic model adds the same forward approximation, which is used in the onset model introduced in section 2.2.2 (Forward Tracing).
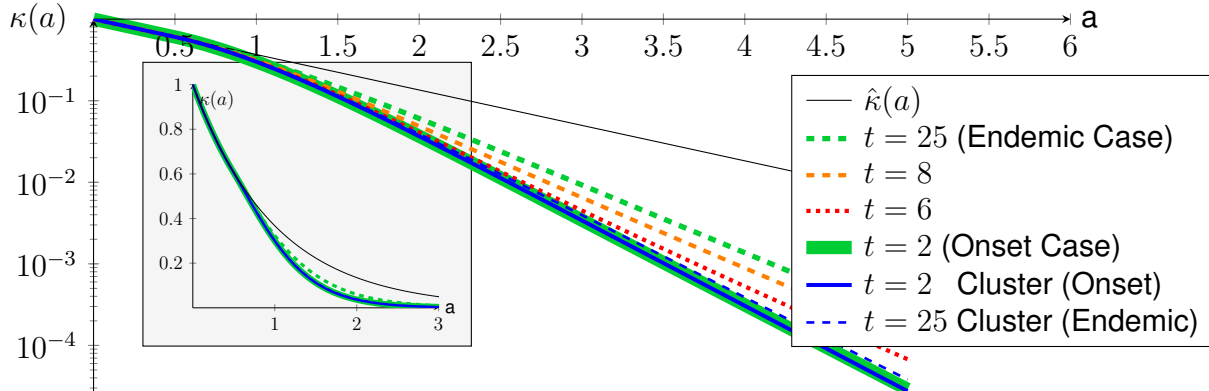
Figure 2.11: $\kappa(a)_-$ of backward tracing between the onset case at $t = 2$ and the endemic case at $t = 25$ of the endemic model with fixed delay $T = 0.5$. Furthermore, unsupervised recovery rate $\alpha = 0.1$, supervised recovery rate $\sigma = 0.9$, contact rate $\beta = 2$, and $p = 1$. Green depicts the onset of the endemic model as a solid line and the endemic case as a dashed line. Red and orange depict the transition between the onset and the endemic case. Blue depicts the endemic model with backward tracing and cluster contact tracing. The endemic case with cluster contact tracing is very similar to the onset case in contrast to the separation of endemic and onset without cluster contact tracing. Thus, the following three plots almost overlay: The endemic model with backward tracing in the onset case and the onset and the endemic case of the endemic model with backward and cluster contact tracing.

We analyze the transition of the endemic model from the onset to the endemic case by concentrating on the probability $\kappa(a)$ first. Because $\kappa(a)$ describes the probability of being infected after time a, we determine $\kappa(a)$ for individuals born at time t and select values of t from the onset ($t = 2$), the transition ($t = 6$ and $t = 8$), and the endemic case ($t = 25$).
Figure 2.11 shows how $\kappa(a)$ changes during the transition from the onset case to the endemic case, and it shows how the $\kappa(a)$ of the endemic case resembles the onset in the case of additional cluster contact tracing. I.e., there is an apparent difference between the onset and the endemic case in the case of backward tracing only. In the case of additional cluster contacts, this difference is gone.

The Endemic Model Resembles the Onset Model in the Onset Case

The backward tracing part of the endemic model (model 45) adapts the onset model from proposition 11 (One-Step Backward Tracing) to the endemic case. The endemic case is modeled by including the possibility of infected individuals contacting each other, which is excluded in the onset case. The addition of a time axis and introducing an age-structure allows for calculating a graph of infected individuals at the time of the epidemic t. Despite these additions, the endemic model has to fit to the results of the onset model in the onset case.

Proposition 48 (The Endemic Model Resembles Onset Model in the Onset Case)
The backward tracing with or without the cluster contact tracing subset of the endemic model from model 45 equals the onset model from proposition 11 (One-Step Backward Tracing) in the onset case when $t \to 0$.

Proof: The onset case is characterized by exponential growth. We observe $t \to 0 \Rightarrow \frac{S}{N} \to 1$. If every contact is an IS contact, there will not be contacts between two infected individuals (cluster contacts). Thus:

$$\kappa_-(0,t) = 1$$
$$T_b(a,t) = p \int_0^a \beta \cdot \frac{S(t-c)}{N} \cdot \kappa_-(c,t) \cdot \sigma dc$$
$$T_c(a,t) = p \int_0^a \beta \int_0^\infty \frac{i(\tilde{a},t-c)}{N} \cdot \frac{\kappa_-(c+\tilde{a},t)}{\kappa_-(\tilde{a},t-c)} \sigma d\tilde{a} dc$$
$$(\partial_t + \partial_a)\kappa_-(a,t) = -\kappa_-(a,t)\left(\alpha + \sigma + (\phi * T_b)(a,t) + (\phi * T_c)(a,t)\right)$$
$$\overset{\frac{S(t)}{N} \to 1}{\Rightarrow} \kappa'_-(a) = -\kappa_-(a)\left\{\alpha + \sigma + p\beta\sigma(\phi * \kappa_-^{\#})(a)\right\}, \quad \kappa_-(0) = 1.$$

$\square$

Figure 2.7 shows the number of infected individuals and helps to classify a time t into the onset or endemic case. If cluster contact tracing is not activated, the equilibrium shows a higher number of infected individuals. $\kappa(a)$ of the endemic case shows a higher probability to be infective (as shown in fig. 2.11). Without the tracing of cluster contacts, every cluster contact prohibits the possible tracing by an infective IS contact. In the limit of the onset case, no cluster contacts are prohibiting IS contacts and the resulting backward tracing and the endemic model resembles the onset model.

Cluster Contact Tracing Substitutes Lost IS Tracing

Figure 2.7 shows the good match of the onset and the endemic case in the case of active cluster contact tracing. Without cluster contact tracing, we observed less effective backward tracing because fewer infections per infector during the endemic case result in less tracing. On the other hand, the contact rate β stays constant. Instead of an IS contact, cluster contacts between infected individuals occur and substitute the common backward tracing.

Proposition 49 *Let $T_b(a,t)$ and $T_c(a,t)$ denote the backward tracing and cluster contact tracing rates used in the endemic model without delay (model 43):*

$$T_b(a,t) = p \int_0^a \beta \cdot \frac{S(t-c)}{N} \cdot \kappa_-(c,t) \cdot \sigma dc,$$
$$T_c(a,t) = p \int_0^a \beta \int_0^\infty \frac{i(\tilde{a},t-c)}{N} \cdot \frac{\kappa_-(c+\tilde{a},t)}{\kappa_-(\tilde{a},t-c)} \sigma d\tilde{a} dc.$$

Then cluster contact tracing T_c replaces the lost backward tracing T_b by IS contacts in the endemic case. The overall effectivity of the tracing both terms stays constant in a first-order approximation

$$T_c(a,t) + T_b(a,t) \approx p \int_0^a \beta \cdot \kappa_-(a) \cdot \sigma dc \quad \text{for all } a \geq 0.$$

Proof: In a first-order approximation, the sum of the integrals stays constant, because the integrands are similar apart from a factor resulting in a linear combination of backward tracing and cluster contact tracing.
First, we remember the one-step backward tracing from proposition 13.

$$\kappa_-(a) = \hat{\kappa}(a) - p\, p_{\text{obs}}\beta\hat{\kappa}(a)(1 * \phi * (1 - \hat{\kappa})(a)) + \mathcal{O}(p^2).$$

We observe the use of the zeroth-order approximation. The zeroth-order approximation does not include tracing and therefore is independent of the onset and endemic case.
Then we use the zeroth-order approximation of $\kappa_-(a)$ to approximate the backward tracing $T_b(a,t)$ within the integro differential equation, before tracing the backward tracing and the cluster-contact tracing back to the backward tracing of the onset case:

$$T_b(a) = p \int_0^a \beta \cdot \hat{\kappa}_-(c) \cdot \sigma dc.$$

After using the first-order approximation for T_b and observing T_c being a linear combination of shifted and scaled first-order approximations, we observe the sum $T_c(a,t) + T_b(a,t)$ of

first-order approximations being independent of the number of infected individuals.

$$T_b(a,t) = p \int_0^a \beta \cdot \underbrace{\frac{S(t-c)}{N} \cdot \kappa_-(c,t)}_{B(c,t)} \cdot \sigma dc,$$

$$T_c(a,t) = p \int_0^a \beta \underbrace{\int_0^\infty \frac{i(\tilde{a},t-c)}{N} \cdot \frac{\kappa_-(c+\tilde{a},t)}{\kappa_-(\tilde{a},t-c)}}_{C(c,t)} \sigma d\tilde{a} dc.$$

Replace $\kappa_-(c,t)$ by its zeroth order approximation $\hat{\kappa}(c)$:

$$B(c,t) = \frac{S(t-c)}{N} \cdot \kappa_-(c,t)$$

$$\Rightarrow B(c,t) \approx \frac{S(t-c)}{N} \cdot \hat{\kappa}_-(c).$$

Now, replace $\kappa_-(c,t)$ by its zeroth order approximation $\hat{\kappa}(c)$ like in the first order approximation of backward tracing:

$$C(c,t) = \int_0^\infty \frac{i(\tilde{a},t-c)}{N} \cdot \frac{\kappa_-(c+\tilde{a},t)}{\kappa_-(\tilde{a},t-c)} d\tilde{a}$$

$$\Rightarrow C(c,t) = \int_0^\infty \frac{i(\tilde{a},t-c)}{N} \cdot \frac{\hat{\kappa}(c+\tilde{a})}{\hat{\kappa}(\tilde{a})} d\tilde{a} + \mathcal{O}(p)$$

Apply $\frac{\hat{\kappa}_-(c+\tilde{a})}{\hat{\kappa}_-(\tilde{a})} = \frac{e^{-(\sigma+\alpha)(c+\tilde{a})}}{e^{-(\sigma+\alpha)(\tilde{a})}} = e^{-(\sigma+\alpha)(c)} = \hat{\kappa}_-(c).$

$$= \int_0^\infty \frac{i(\tilde{a},t-c)}{N} \cdot \hat{\kappa}(c) d\tilde{a} + \mathcal{O}(p)$$

Calculate the trivial linear combination.

$$= \frac{I(t-c)}{N} \cdot \hat{\kappa}_-(c) + \mathcal{O}(p)$$

$$T(c,t) = C(c,t) + B(c,t) \approx \left(\frac{S(t-c)}{N} + \frac{I(t-c)}{N} \right) \cdot \hat{\kappa}(c) = 1 \cdot \hat{\kappa}(c).$$

□

In the case of the onset, the endemic and the onset model are identical. When approaching the endemic case, fewer infections per infector are taking place. The number of infections per infector approaches 1, and the backward tracing becomes less effective. However, the lost backward tracing is replaced by cluster contact tracing.

The reduced number of infections while keeping a tracing effect similar to the onset results

in fewer infected individuals, as seen in fig. 2.7.

2.4 The Age Structure

In the last part, we investigated the probability of being infective in the case of backward tracing without and with cluster contact tracing. With cluster contact tracing, we obtained the probability $\kappa_-(a)$ independent of the onset and the endemic case. Now, we will investigate the age structures in the onset and the endemic case.

The age structures of the onset and the endemic case differ. We approximate the age structures in the no tracing case. Without tracing, the probability of being infective in the endemic case after time a is $\kappa(a)$, which is also the zeroth-order approximation of backward tracing. In the onset case, the structure is determined by the contact rate. Figure 2.12 shows this approximation compared to the numerical age structure from the endemic model[1].
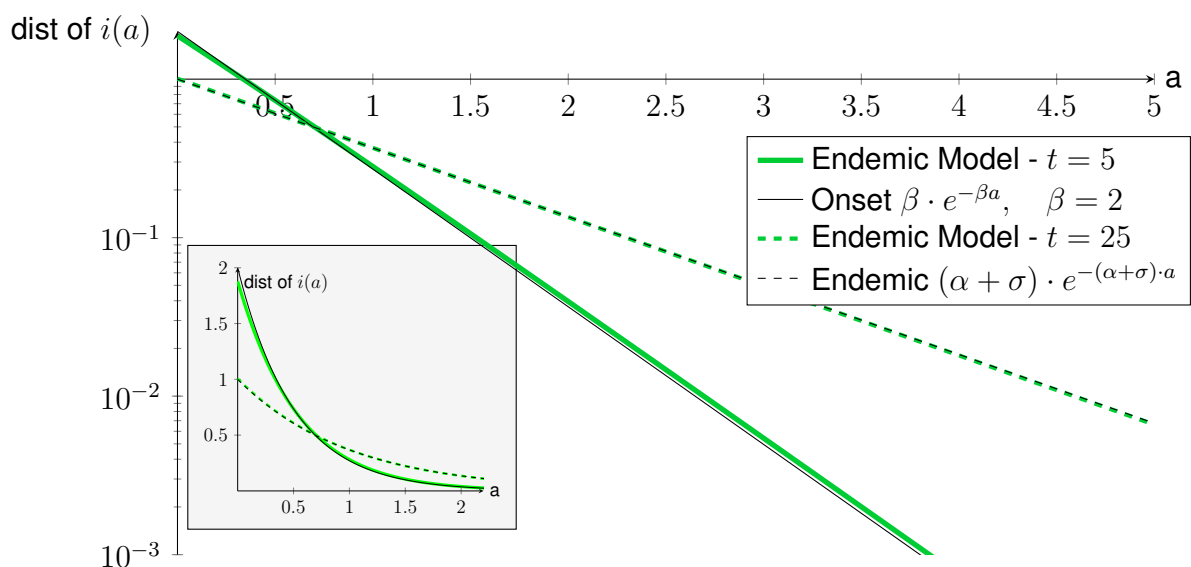
Figure 2.12: Distribution of $i(a)$ without tracing during the onset at time $t = 5$ and during the endemic phase at time $t = 25$ using a contact rate $\beta = 2$, supervised recovery rate $\sigma = 0.9$, and unsupervised recovery rate $\alpha = 0.1$.

The Onset Age Structure without Tracing

We use a PDE to calculate the age structure of the onset of the spread like it is modeled by the onset model. [MH07, p.233, proof of proposition 3.6]

Proposition 50 (The Onset Age Structure without Tracing) *The onset age structure without tracing can be described by* $i_P(a) = \beta \cdot e^{-\beta a}$

---

[1] Figure 2.7 shows the number of infected individuals from the start of an endemic until the equilibrium in case of no tracing.

Proof:

We take a partial differential equation modeling the infected individuals z using the hazard rate derived from $\hat{\kappa}(a)$.

$$\partial_t z + \partial_a z = \left(\frac{\hat{\kappa}'(a)}{\hat{\kappa}(a)} \right) z.$$

The boundary condition is defined by the contact rate β and the size of the population $\int_0^\infty z(t,a) da$.

$$z(t,0) = \beta \int_0^\infty z(t,a) da.$$

Choose the ansatz:

$$z(t,a) = e^{\hat{\lambda} t} u(a).$$

Insert the ansatz into the partial differential equation:

$$\Rightarrow \qquad e^{\hat{\lambda} t} \hat{\lambda} u(a) + e^{\hat{\lambda} t} u'(a) = \left(\frac{\hat{\kappa}'(a)}{\hat{\kappa}(a)} \right) e^{\hat{\lambda} t} u$$

$$\Rightarrow \qquad \hat{\lambda} u(a) + u'(a) = \left(\frac{\hat{\kappa}'(a)}{\hat{\kappa}(a)} \right) u$$

$$\Rightarrow \qquad u' = \left(\left(\frac{\hat{\kappa}'(a)}{\hat{\kappa}(a)} \right) - \hat{\lambda} \right) u$$

$$\Rightarrow \qquad u(a) = C e^{\int_0^a \frac{\hat{\kappa}'(\tau)}{\hat{\kappa}(\tau)} - \hat{\lambda} d\tau}$$

$$= C e^{ln(\hat{\kappa}(a)) - ln(\hat{\kappa}(0)) - \hat{\lambda} a}$$

Simplify the term by using the boundary condition $\hat{\kappa}(0) = 1 \Rightarrow ln(\hat{\kappa}(0)) = 0$.

$$= C e^{ln(\hat{\kappa}(a)) - \hat{\lambda} a}$$

$$= C \hat{\kappa}(a) e^{-\hat{\lambda} a}.$$

This turns the ansatz into:

$$z(t,a) = e^{\hat{\lambda} t} C \hat{\kappa}(a) e^{-\hat{\lambda} a}$$

$$= C \hat{\kappa}(a) e^{\hat{\lambda}(t-a)}.$$

105

Now use the boundary condition of the PDE to remove C and evaluate λ.

$$z(t,0) = e^{\hat{\lambda}t}C\hat{\kappa}(0)e^{-\hat{\lambda}0} = \beta\int_0^\infty z(t,a)da$$

$$= \beta\int_0^\infty e^{\lambda t}C\hat{\kappa}(a)e^{-\hat{\lambda}a}da$$

$$\Rightarrow \qquad \kappa(0)e^{-\hat{\lambda}0} = 1 = \beta\int_0^\infty \hat{\kappa}(a)e^{-\hat{\lambda}a}da$$

$$\Rightarrow \qquad \frac{1}{\beta} = \int_0^\infty e^{-(\alpha+\sigma)a}e^{-\hat{\lambda}a}da$$

$$\Rightarrow \qquad \hat{\lambda} = -\alpha - \sigma + \beta.$$

To gain the age distribution we have to normalize $u(a)$:

$$i_P(a) = \frac{u(a)}{\int_0^\infty u(a)da} = \frac{e^{-\hat{\lambda}a}\hat{\kappa}(a)}{\int_0^\infty e^{-\hat{\lambda}a}\hat{\kappa}(a)da}$$

$$= \frac{e^{-\hat{\lambda}a}\hat{\kappa}(a)}{\frac{1}{\beta}}$$

$$= \beta e^{-\hat{\lambda}a}\hat{\kappa}(a)$$

$$= \beta e^{-\beta a}.$$

$$\square$$

The Endemic Age Structure

On the other hand, the endemic age structure equals the shape of the probability of being infective $\kappa(a)$.

The age structure $i_P(a)$ being equal to the shape of $\kappa(a)$ in the endemic state is expected. The endemic case is characterized by reaching the equilibrium state. Therefore, the time t can be neglected. Normalizing $\kappa(a)$ and multiplying this probability by the total number of infected individuals at that time $I(t)$ results in the number of infected individuals of age a at time t denoted by $i(a,t)$. In the endemic equilibrium, $i(a,t)$ will be independent of time. We obtain:

$$i(a) = I \cdot \frac{\kappa(a)}{\int_0^\infty \kappa(a)da}.$$

Now, dividing by the number of infected individuals results in the distribution of individuals still being infective at age a

$$i_P(a) = \frac{i(a)}{I} = \frac{\kappa(a)}{\int_0^\infty \kappa(a)da}.$$

This consideration leads to the following proposition, which provides an alternative proof for the endemic case without tracing.

Proposition 51 (Endemic Age Structure without Tracing) *In the endemic case, we approach an equilibrium. Thus, $i(a,t)$, $I(t)$ and $\kappa(a,t)$ do not depend on t. With $p = 0$, we can describe the age distribution $i_P(a)$ by $i_P(a) = \frac{i(a,t)}{I(t)} = \frac{\kappa(a,t+a)}{\int_0^\infty \kappa(a,t+a)da}$.*

Proof:

As in the onset case, we take the partial differential equation modeling the infected individuals z using the hazard rate derived from $\hat{\kappa}(a)$. This time, we expect a result independent of time t and choose the ansatz $z(t,a) = u(a)$.

$$\partial_t z(t,a) + \partial_a z(t,a) = \left(\frac{\hat{\kappa}'(a)}{\hat{\kappa}(a)} \right) z(t,a)$$

$$\Rightarrow \qquad u'(a) = \left(\frac{\hat{\kappa}'(a)}{\hat{\kappa}(a)} \right) u(a)$$

$$\Rightarrow \qquad u(a) = C e^{\int_0^a \frac{\hat{\kappa}'(a)}{\hat{\kappa}(a)} dt}$$

$$= C e^{ln(\hat{\kappa}(a)) - ln(\hat{\kappa}(0))} = C \hat{\kappa}(a).$$

Normalizing results in the distribution and removes the unknown variable C.   □

The difference between the onset age structure and the endemic age structure is visualized in fig. 2.12 and summarized in the following remark.

Remark 52 *Without tracing, $p = 0$, we can describe the age structure in the onset and endemic case. The onset distribution of infected individuals $i_P(a) = \beta e^{-\beta a}$ is characterized by the contact rate β. The endemic distribution $i_P(a) = \hat{\kappa}(a) = e^{-(\alpha+\sigma)a}$ is determined by the recovery rate $\sigma + \alpha$.*

In opposite to proposition 50 (The Onset Age Structure without Tracing) the proposition 51 (The Endemic Age Structure) did not need to employ the boundary condition to eliminate a free variable. If analyzing the boundary condition despite this fact, we observe it can be used to determine the number of infected individuals, because it resembles the calculation from remark 47 (Compare Onset, Endemic, and IBM Model). Compare the calculations to deduce the number of infected individuals from the reproduction number.
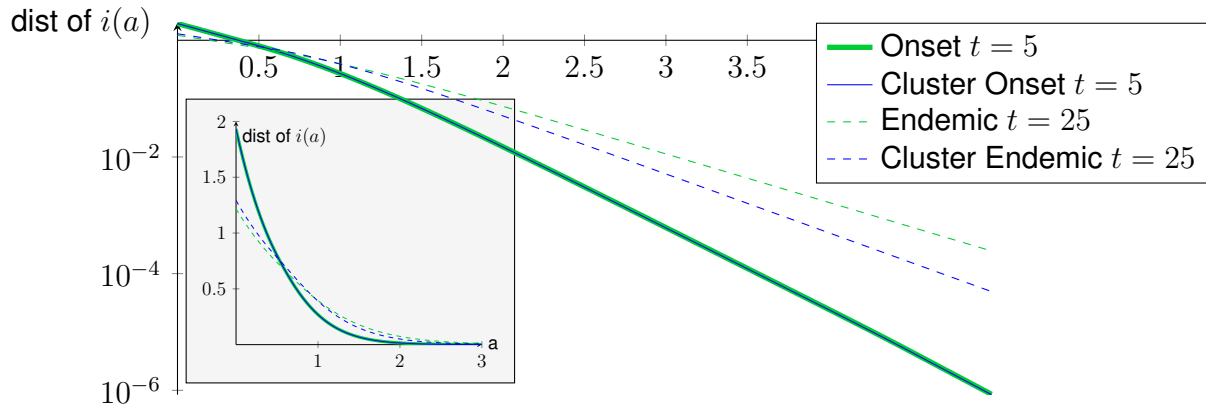
Figure 2.13: We use the endemic model to calculate the distribution of $i(a)$ with backward tracing and optional cluster contact tracing during the onset at time $t = 5$ and during the endemic phase at time $t = 25$ using a contact rate $\beta = 2$, supervised recovery rate $\sigma = 0.9$ and unsupervised recovery rate $\alpha = 0.1$.

The Age Structure with Tracing

In the endemic model (model 45), we obtain the onset age structure and the endemic age structure at an arbitrary time. However, evaluating arbitrary points in time does not always result in artifact free distributions. During the onset, very early times result in old individuals not existing within the age distribution yet[1]. Therefore, we start at time $t = 5$.

Figure 2.13 shows the age distribution with backward tracing in the onset and the endemic case compared to cluster contact tracing in the onset and the endemic case. The age distributions of the endemic and onset case are separated. Adding cluster contact tracing does not stop this separation; it does not result in similar age distributions of the onset and the endemic case. There is a visible effect of cluster contact tracing on the age distribution, but this time there are no 'lost' IS-contacts replaced by cluster contact tracing.

Previously, when concentrating on the probability of being infective $\kappa(a)$ shown in fig. 2.11, adding cluster contact tracing resulted in similar probabilities of being infective at age a of the infection, when comparing the onset and endemic case.

The analytic formulas (see fig. 2.12) show that the age structure of the onset is dominated by the initial spread, which is defined by the contact rate β. On the other hand, the endemic case is characterized by the strength of the recovery rate.

---

[1]A graph showing the age distribution at $t = 2$ looking similar to the age distribution at $t = 5$ can be found in the appendix (section 5.1.4). Because the maximum age at $t = 2$ is limited by that time, the distribution of that graph stops at age 2.

2.5 Summary and Outlook

In this chapter, we created a model to cover the endemic case, which lead to adapting the backward tracing of the onset model to the endemic case. We included the cluster contact tracing, which consists of the non-infective contacts between infected individuals. Because these contacts are non-infective, they do not result in backward tracing.

First, we observed a good fit between our model and our IBM simulation.

Then, we observed that in a first-order approximation, the tracing by the cluster contacts compensates the smaller effect of backward tracing in the endemic case. When including cluster contact tracing, the probability of staying infective per individual stays approximately constant, not being dependent on the onset or endemic case.

When analyzing the age structure, other effects are determining the shape. During the onset phase, the spread of the infection dominates the structure, while in the endemic phase, the recovery rate, which is influenced by cluster contact tracing, shapes the age structure.

Outlook

Currently our evaluations of the model use a likelihood to infect a contact of 1. During the onset phase, there is no difference between not infecting an individual and not contacting an individual at all. This is not necessary true in the endemic case. As an educated guess, a smaller likelihood would result in the cluster contact to more than compensate the loss in backward tracing. This smaller effect might persist even if we introduce a non-infective contact tracing class, because a big part of these non-infected contacts will stay susceptible for a longer time and will not be able to trigger any tracing cases.

Apart from non-infective-contacts, which might become infected until the contact is analyzed within a tracing event, there are reinfected contacts, contacts which recover and become infected again before being covered and analyzed in a tracing event. These higher order interactions should not be likely in realistic scenarios. To match the model at these unrealistic incidence rates, we added an option into the IBM simulation to only trace the contacts, which actually took part in an infection or in a cluster contact within the current infection of the individual. Analyzing these interactions could make the mathematical model slightly more precise.

It would be interesting to analyze the mix of generations we observe within the IBM simulation. The interaction of different generations plays a significant role, when analyzing one-step tracing. In a real-world scenario, we can not classify all individuals of an epoch as a member of a single generation. Furthermore, if the infectee or infector was traced

determines if the infectee or infector can trigger a tracing event, which is highly dependent on the generation.

3 Gradient Matching

In this chapter, we employ Gradient Matching to recover unobserved state variables of a dynamical system with known parameters. E.g., as a practical example in chapter 4 we observe an LC oscillator, we know the capacity, the inductor, and the resistance, but we only record the voltage. By employing Gradient Matching we recover the corresponding unobserved current.

Gradient Matching is usually employed to fit the parameters of a differential equation according to a given dataset. In contrast to classical methods, the differential equations are not solved during the process of fitting the parameters. Instead, the algorithm works solely on the equation itself.

As the data samples $x_i = x(t_i)$ at time point t_i, $i = 1, \ldots, n$, of an observed process are known, let us assume we can calculate a derivative $\dot{x}_i$ numerically. We have an ODE-model for the data $\dot{x} = f(x; \mu)$ with parameter vector μ. Finding the best parameters of the model $f(.)$ boils down to solving the system of equations $\dot{x}_i = f(x_i)$. The classic problem of identifying ODE-parameters requires to solve the ODE and compare the solution and the observed data to optimize the set of parameters. This Gradient Matching approach does not require to solve the ODE. Furthermore, we will restrict the model $f(.)$ to consist of polynomials. The parameters of the model control the monomials linearly, and we can employ regression. If we minimize the number of parameters not equal zero by penalizing them while minimizing the residual, the penalized regression allows for a different view on the solution. Now, the process of identifying the correct parameters resembles a symbolic view on constructing the right-hand side of the ODE.

Approaches using the data and the derivative of the data to estimate parameters of an ODE have been around for a relatively long time. Many approaches explicitly underline the two steps of recovering a derivative from noisy observations and using this derivative to calculate parameters in a second step (see Himmelblau in 1967 [HJB67]). We consider a paper of Kate BAK [Bak+63] as an early example of this approach in 1963. In addition to numerical smoothing and differentiation, the paper mentions practical approaches by drawing a smooth line through the plotted data samples and using a tangent meter [Fra48] to obtain

the derivative, which is then used to determine rate constants[1]. Some advantages of Gradient Matching from that time might be obsolete in many cases. Since we use computers to solve the equations, we do not see big rate constants as a problem when solving an ODE. On the other hand, there are drawbacks of Gradient Matching, which became much more manageable. Gaining smooth derivatives does not require the effort it required before high throughput data became available. Squire [Squ71] picks up the approach using very few data points and concludes the difficulty to gain usable derivatives from experimental data. Apart from difficulties due to noise, the example shows an advantage of Gradient Matching to directly compute parameters within a regression, which does not require repeated evaluation of the differential equation. The paper of Varah [Var82] applies this approach to the Lotka-Volterra predator-prey model in 1982 and furthermore benchmarks the advantages and disadvantages compared to a classic method evaluating the ODE. In 2002 this is picked up by Ellner[ESS02] and established the term Gradient Matching. Brunel[Bru08][BCd14] elaborates this approach by analyzing the nonparametric estimators to achieve a smooth representation. He analyzes the consistency of the estimators and even covers the estimation of parameters of partially observed data. Brunton[BPK16] recently combined the Gradient Matching approach with compressive sensing to match arbitrary dynamical systems. Although he traces the idea back to symbolic regression [SL09], which is better known in the form of genetic algorithms in computer science, his work continues the path of Gradient Matching. His addition to the evolution of Gradient Matching fits the pattern defined by the legacy, even more, when he later applies his algorithm to biological systems [Man+16]. The compressive sensing approach, which has some similarities to our approach we applied in chapter 4, not only estimates the values of the parameters, but it also identifies the correct parameters. Furthermore, the approach allows integrating expert knowledge to guide the identification of an unknown differential equation. Partial knowledge can be used to improve the identification and estimation of the parameters.

In this Gradient Matching approach, the sensitive part is the process of approximating a good derivative. The fitting is a regression. We chose polynomials as a right-hand side, and the regression determines the parameters corresponding to the monomials. We penalize the number of monomials by using the l_0 norm, which means, when solving the regression the number of monomials having a parameter $\neq 0$ is multiplied with a constant and added to the residual error. This penalization reduces the number of parameters to the more essential parameters. Furthermore, this penalization limits the negative effect of

---

[1]A tangent meter like the version described by Frampton [Fra48] is more advanced than the commonly used set square with protractor (German: Geodreieck). The main difference: The tangent meter employs a prism at the contact point of the tangent and the curve, which enables the user to place the angle of the meter with higher accuracy. Furthermore, there is a reference to support the user in determining the angle of the tangent.

collinearity, because the more collinear the terms corresponding to two parameters are, and the higher the penalty, the higher the likelihood of one of the parameters to equal zero. The robustness against collinearity allows for overdetermined systems. If polynomials are used, the parameters of the monomials often correspond to the parameters of the analyzed models in many cases. On the other hand, no effort of actually solving the differential equation has to be made, because fitting the parameters is done by the regression.

The great advantage of assembling the right-hand side of the differential equations using simple operations, which do not employ complicated solvers, is paid by the need to differentiate the data. Differentiating amplifies the noise in opposite to solving the differential equation, which can be seen as an integration canceling the noise.
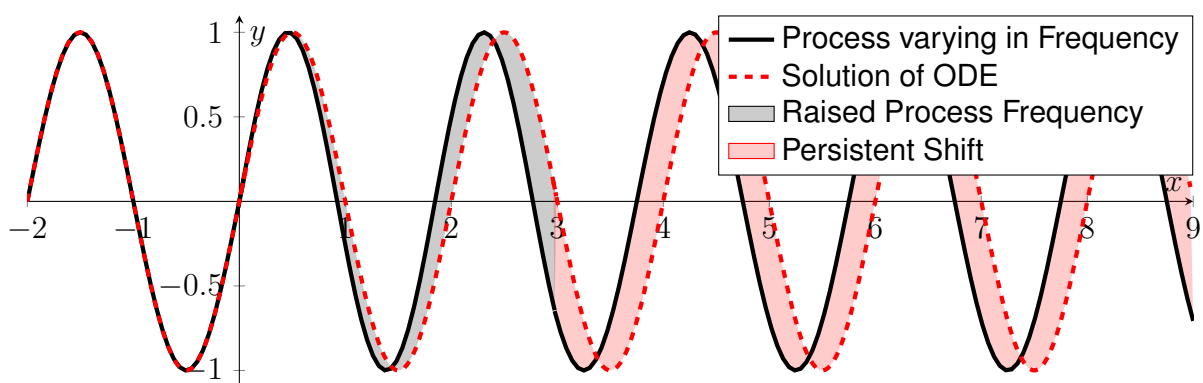
Figure 3.1: Gradient Matching vs classical method, exemplify the matching of an ODE of constant frequency to a process of slightly varying frequency. The frequency of the process is raised between $x = 0$ and $x = 3$, which results in a persisting shift for all following samples after $x = 3$.

In the case of classical methods, the parameters of an ODE are fitted by employing a maximum likelihood approach, which searches for the least difference between the solution of the ODE and the observed process, as visualized in fig. 3.1. Figure 3.1 visualizes a process of varying frequency fitted by an ODE solution only representing constant frequencies. The figure shows an epoch of the process of slightly higher frequency between $x = 0$ and $x = 3$ accumulating an error. Gray shading marks the accumulating error. It results in a phase shift, which remains constant during the following epoch of matching frequency. We mark this remaining phase shift by red shading. In the example of a fast oscillation of an unstable frequency, which is observed over a longer period, this accumulating error can result in a loss of synchronization between measurements and model. The global function averages noise, but has to cope with random changes in the process or rare effects not covered in the ODE. Conversely, Gradient Matching calculates the parameters by employing regression. This regression includes every data sample of the state variable and its corresponding

derivative independently of the rest of the data, the history and the future progress. I.e., in the theoretical limit case, every data sample could be a part of the solution on a different trajectory with different initial conditions.

If the exact derivative is given, the Gradient Matching algorithm is completely localized evaluating every sample on its own and then using these samples to estimate the parameters. If noise occurs as single altered data sample of the state variable (fig. 3.2), it will have a comparable local or limited effect on both approaches. In the classical approach, it will be a single error, as it differs from the solution of the solved ODE. In the case of Gradient Matching, it is wrong pairs of data and corresponding derivative, which will be overruled by correct pairs during the regression as it happens during integration in the classical method.
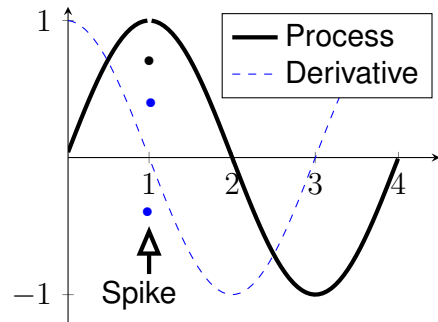
Figure 3.2: Local noise events of sample values result in a local noise spike in the derivative.
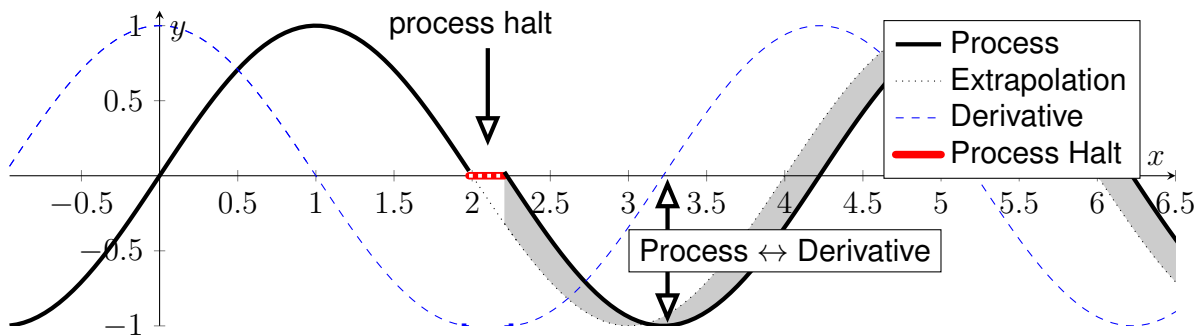
Figure 3.3: Gradient Matching vs classical method. The process halts from $x = 2.0$ or $x = 2.2$. Extrapolating the process would result in a persistent error, because it is shifted to the actual process. The correlation of the value to the derivative, which we concentrate on using Gradient Matching, persists and is not influenced by the halt of the process.

Figure 3.3 shows a process, which halted for a short period of time. The problem of constant additive errors in the time when integrating differential equations was already mentioned in [Bak+63]. After the halt, fig. 3.3 visualizes that the following derivatives still fit the corresponding data points after the additive error in time. Again, the Gradient Matching analyzes the data and the corresponding derivative. Events outside this local scope do not influence the analysis. It is translation invariant. The solution of the ODE, which extrapolates the previous data, will have an offset compared to all the following data.

Because we analyze the sample and its derivative in the case of Gradient Matching, even permuting each sample would not influence the outcome. Presumingly this behavior can be exploited to gain robustness against singular effects.

As the Gradient Matching solution is not dependent on initial values of the differential equation, the algorithm would not even be bothered by an effect resulting in the data jumping into a different trajectory having different initial conditions. In real-life experiments, the variation of initial values of the differential equations often results in very different outcomes. They can result in ill-posed problems. Figure 3.4 shows exponential growth. The initial values are close to zero. Therefore, small offsets due to noise result in very different trajectories and might
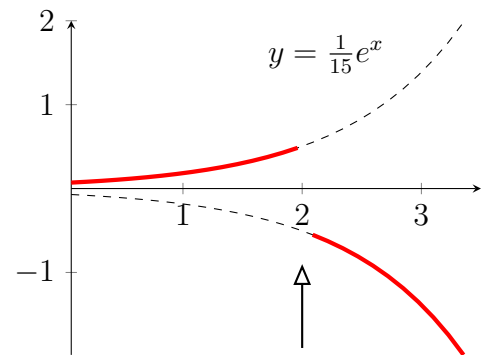
Figure 3.4: Singular event – jumping between trajectories.

even change the sign of the differential equation. When using a small measured value as initial value of the exponential growth differential equation, the noise will determine if the differential equation grows towards plus or minus infinity. The paper [Var82] points out the advantage of being independent of initial conditions. The local structure allows calculating each sample independent of each other. Because no ODE is solved, the first samples do not influence the rest of the trajectory. Noise at that initial stage does not prohibit the calculation of the major rest. In the end, variations in the initial phase will just add to the noise level of the regression. There are enhancements to the classic methods regarding this problem. The classic methods solve the differential equation numerically, starting at the initial conditions. They perform an optimization on the parameter of the differential equation by repeatedly solving the equation. The multiple shooting method [SB96][p. 516ff.] breaks the ill-posedness of the problem by starting at different times of the differential equation. A practical description and an updated comparison to other methods is provided by [PT07]. Instead of solving one differential equation over the whole space, the space is divided into smaller subspaces, which are solved independently having their own initial values.
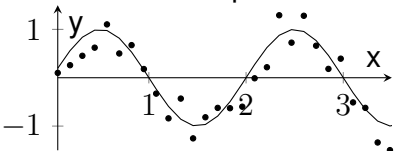
To further underline the feature of Gradient Matching to be independent of the initial conditions without having to rely on enhancements of the original idea, we exaggerate the problem and assume a system, which switches to a different trajectory. In the classical approach using the multiple shooting method, detection of the incidents, which affects the trajectory, is necessary. Otherwise, a larger chunk of the data would result in misleading parameters flawing the result of the calculation.

This change of the trajectory is illustrated in the exponential growth example from fig. 3.4. The Gradient Matching algorithm will not note that half of the samples are sampled of a positive and half of the samples of a negative trajectory of the exponential growth function. Varying the parameter of growth on the other hand, would result in the search of a new

parameter trying to approximate all variations of the data. The experiments performed on measured data in chapter 4 show a practical example and support this assumption. They deal with an oscillating circuit. That circuit is triggered by Dirac impulses repeatedly because the damping is high. If we had to integrate the differential equation, we would have to detect the Dirac impulses, which start the oscillation, i.e., a new trajectory. Because of the independence of the trajectories, Gradient Matching does not have to detect the impulses. It can just record the data and ignore the fact that it jumps to a new trajectory every time a new impulse is triggered.

This comparison between Gradient Matching and classical methods is summarized in table 3.1. Of course, a model describing the data, which relies on being reinitialized with different initial conditions from time to time, does not seem to be doing a very good job. On the other hand, when monitoring an oscillation, it seems obvious to not model the Dirac impulse triggering the oscillation in the differential equation. This impulse is only a side effect, and its existence within the model would distract from the actual dynamic happening. The same is true if the growth is monitored in biological experiments. Many times, an offset in time or a delay in bacteria growth is considered as noise. In many cases, not modeling the noise is not a flaw in the model, but a necessary step to focus on the essential dynamics.

In the following chapters, we will use Gradient Matching and restrict the right-hand side of the ODE to polynomials. We choose polynomials, because of the capability to describe the right-hand side of typical equations occurring in models like the Lotka-Volterra equations, spring/mass systems, LC oscillators, and chemical reactions. These models result in equations of the form $\dot{x} = a + bx + cx^2 + \dots$. If the pair $(x, \dot{x})$ is available by sufficiently smooth data points, linear regression can in general estimate the parameter.

Gradient Matching versus Classical Methods

Property	Classical Method	Gradient Matching
View	Works on solution of differential equations.	Works on the equations itself.
Visualization	Solution of ODE minimizes distance to data samples.	Works on the equation like symbolic regression and minimizes the error of the regression. $\dot{x} = a + b \cdot x + c \cdot x^2 + d \cdot x \cdot y \dots$ $\dot{y} = e + f \cdot y + \dots$

Tool	$\int$ Solving the differential equation is achieved by integrating the data.	$\frac{d}{dx}$ In case of Gradient Matching the data has to be derived.
Locality	Nonlocal	Local
Side effect	Sensitive to poisson noise and initial conditions.	Amplification of noise.
Computational Complexity	Multiple evaluations of ODE.	Linear + computations enhancing the derivative or identifying the existence of parameters.

Table 3.1: Gradient Matching versus Classical Methods.

3.1 Numerical Derivatives

When applying Gradient Matching in practice, there are various ways to achieve a numerical derivative of the data. Above, we already mentioned the disadvantage of Gradient Matching of amplifying noise when deriving the data. Furthermore, we should try to maintain one of the advantages of Gradient Matching, the local approach. On the other hand, there is no per-sample local numerical derivative. Often, respecting more samples results in a less local approach; this is especially true if smoothing or averaging is used. However, various algorithms are tradeoffs between being more localized or being more noise resistant.

A numerical derivative can be achieved by fitting a polynomial locally like Ellner [ESS02] or [SB75], applying splines [Var82], smoothing the function, and applying a relatively stable approach like the five-point stencil or using optimization and a total variation approach like Chartrand[Cha11].

Thereby, the total variation approach will keep discontinuities, i.e., jumps from one trajectory to another. The drawback is the involved l_1 norm, which is not as easy to handle within optimizers. On the other hand, a traditional way of smoothing by applying splines or polynomials will be a less localized approach and usually not model discontinuities.

3.2 Latent Function Recovery Using Gradient Matching

In this chapter, we concentrate on a dynamical system of two equations. While only observing one component $x(t)$ of the trajectory, we use Gradient Matching to construct a functional, which allows constructing a locally unique solution of the second component $y(t)$ of the trajectory.

This tool helps in cases of partially observed data. Sometimes it is harder to observe some components of the data compared to other components. This scenario results in missing components or components of low quality. To provide an example, we can imagine measuring the current and the voltage of an LC-oscillator. Measuring the voltage is a lot easier than measuring the current, if the current is measured by employing a shunt resistor (see the example provided in chapter 4, Gradient Matching Experiment).

After showing the existence and local uniqueness of the classical solution in the exact (noise-free) case, we add noise and show the applicability of the implicit function theorem at the location of the 'true' solution. The implicit function theorem hints the existence of an approximating solution near the unknown 'true' $y(t)$ trajectory for sufficiently small levels of noise. This proof is performed by showing that the optima underlie a continuous transition as the level of noise rises. Furthermore, we exemplify the theory by the experiments performed in chapter 4. On the other hand, we note that local optima might appear, which will prohibit reaching the correct solution in case of higher noise levels, especially in more complex dynamical systems.

3.3 Notation

Function spaces

C^0	Continuous functions.
$C_0(\Omega)$	Continuous functions with compact support within Ω. In literature, $C_c(\Omega)$ is a common alternative notation.
C^n	Continuous function having n continuous first derivative y.
$C^{n,a}$	In addition to C^n Hölder spaces [Eva10][pp.240-241] fulfil the Hölder condition $|f(x) - f(y)| \leq C\|x - y\|^\alpha$.
L^p	Function space of measurable functions using the norm $\|f\|_{L^p} := \left(\int_\Omega |f|^p d\mu\right)^{\frac{1}{p}} < \infty$.
$W^{n,p}$	The Sobolev space. A function within the space and its first n derivatives have a finite L^p norm.
$W_0^{1,2}$	Closure of a smooth function with zero boundary condition using the Sobolev norm $\equiv \overline{C_0^\infty}^{\|\cdot\|_{W^{1,2}}}$.

Operator Spaces

$\mathcal{L}$	Bounded, linear Operators

Variables

$\eta(t,\lambda)$	Poisson disturbance with parameter λ.
$n(t)$	The function $n(t) \in W_0^{1,2}$ describes the noise. The subscript $n_x(t)$ or $n_y(t)$ denotes the corresponding state variable $x(t)$ or $y(t)$.
$x(t)$	The observed function, a solution of the first equation of the dynamical system.
$\tilde{x}(t)$	$\tilde{x}(t) = x(t) + n(t, \mu = 0, \sigma)$ - The observed data $x(t)$ with Gaussian noise added.
$y(t)$	The unknown second component of the dynamical system.
$\tilde{y}(t)$	Approximation of $y(t)$.
$\hat{y}(t)$	True $y(t)$.
$u(t)$	Difference to $\hat{y}$.

Functions

$\mathbb{1}_\Omega$	Indicator function.
$f_1(x, y)$	Right-hand side function of the first equation.
$f_2(x, y)$	Right-hand side of the second equation of the dynamical system.
$f_{2,x}$	The derivative in direction of x of f_2.

Table 3.2: Notation – Gradient Matching.

3.4 Model

This chapter deals with a dynamical system in two components, $x(t)$ and $y(t)$. However, the 'true' function $\hat{y}(t) \in W^{1,2}$ of the original system is assumed to be unobserved. This 'true' function fulfills the second equation exactly.

This system raises the question, whether there is sufficient information to recover an approximation $\tilde{y}(t)$ on a given period of time $t \in (0, \tau)$ only by observing $x(t)$ and knowing the right-hand side of the differential equations. The right-hand side of the differential equations describing the dynamical system consists of the functions $f_1, f_2 \in C^{0,1}_{\text{loc}}(\mathbb{R}^2, \mathbb{R})$ corresponding to the first component $x(t)$ and the second unknown component $y(t)$. The time period is restricted by the finite time τ of observation of the trajectory $x(t)$. Furthermore, it raises the question, whether this approximation stays near the 'true' function $\hat{y}(t) \in W^{1,2}$ when noise is introduced. A pursued behavior would be a continuous shift of the approximated optimum when increasing the noise. Discontinuous behavior would start at higher noise levels, when $\tilde{y}$ leaves the shrinking neighborhood of $\hat{y}$ and enters a different local optimum. A not desirable behavior would be an ill-posed behavior being inherently discontinuous.

Problem without Noise

The trajectory consisting of the two funcitons, the observed $x(t) \in C^1$ and the unobserved, 'true' $\hat{y}(t) \in W^{1,2}$, solves the following dynamical system.

The function $x(t)$ has to be differentiable to allow being analyzed by Gradient Matching tools.

Problem 53 (Dynamical System) *The function $x(t) \in C^1$ and the 'true' function $\hat{y}(t) \in W^{1,2}$ solve the following dynamical system. For $x(t) \in C^1$, $f_1, f_2 \in C^{0,1}_{loc}(\mathbb{R}^2, \mathbb{R})$ and $\tau \in \mathbb{R}_+$ find $\tilde{y}(t) \in W^{1,2}$ such that:*

$$x'(t) = f_1(x(t), \tilde{y}(t)), \quad t \in [0, \tau], \tag{3.1}$$

$$\tilde{y}'(t) = f_2(x(t), \tilde{y}(t)), \quad t \in [0, \tau]. \tag{3.2}$$

We note, the recovered $\tilde{y}(t)$ does not have to match the 'true' function $\hat{y}(t)$. The dynamical system with a given observed function $x(t)$ might not have a unique solution $\tilde{y}(t)$ as we show in an example in section 3.7.4.

When searching for a function $\tilde{y}(t)$ fulfilling these equations, we intend to minimize the residuals $x' - f_1(x, y)$ and $y' - f_2(x, y)$. Therefore, we define a functional $J(y(t))$ in definition 54, adding the residuals. This functional has a global minimum at $J(\hat{y}(t)) = 0$ as the 'true' solution fulfills the equation without any residuum.

Definition 54 Define the functional $J : W^{1,2} \to \mathbb{R}$ describing the error from problem 53 by

$$J(y(t)) = \frac{1}{2}\|x'(t) - f_1(x(t), y(t))\|^2_{L^2(0,\tau)} \tag{3.3}$$

$$+ \frac{1}{2}\|y'(t) - f_2(x(t), y(t))\|^2_{L^2(0,\tau)}, \tag{3.4}$$

$$I = [0, \tau].$$

To meet the structure of our problem, we require from now on the following assumptions.

Hypothesis 55 (General Preconditions) *In the remaining part of the chapter, we assume there exists a $\hat{y}$ fulfilling the problem 53 resulting in $J(\hat{y}) = 0$, in the case of no noise. Furthermore, we assume that $y_n \in W^{1,2}$ is a sequence with:*

1. *$\|y_n\|_{L^\infty} \leq \zeta$,*

2. *$\lim_{n \to \infty} J(y_n) = 0$,*

3. *$\forall t : (f_{1,y}(x(t), y(t)))^2 \geq \epsilon > 0 \Rightarrow \|f_{1,y}\|^2 > 0.$*

3.4.1 Deriving the Objective Function at the True Optimum

Before we proceed in section 3.4.2 to prove the existence of a classical solution of the objective function, we will derive the objective function from definition 54. We show $\hat{y}$, which is defined as the original, 'true' function and fulfills $J(y) = 0$, is an isolated optimum. Therefore, we assume $y(t) \in W^{2,2}$ within this subsection and calculate the first and second derivative of the objective function at the location of the 'true' function $\hat{y}(t)$.

The second derivative will indicate that $\hat{y}(t)$ is an isolated optimum, and we will proceed in section 3.4.2 to prove the existence of a classical solution of the objective function. That proof for the existence of the classical solution does not rely on the derivatives of this subsection.

First Derivative of the Objective Function

We calculate the derivation in the direction of a test function using test functions with compact support $\phi \in C_0^\infty$ [RR04, p.124 def. 5.1]. The boundary values of our domain are not of further interest, and the support prohibits them from introducing more terms into the weak derivative.

Lemma 56 *First derivative of functional J stated in definition 54 using the test function* $\phi \in C_{0\,(0,\tau)}^\infty$ *and* $y \in W^{2,2}$.

$$\frac{\partial}{\partial \epsilon} J(y + \epsilon\phi)|_{\epsilon=0} = -\int_0^\tau \left((x' - f_1) \cdot f_{1,y} + (y'' - f_{2,x} - f_{2,y}) + (y' - f_2) \cdot f_{2,y} \right) \phi \, dt.$$

Proof:

$$
\begin{aligned}
\frac{\partial}{\partial \epsilon} J(y + \epsilon\phi)|_{\epsilon=0} &= \int_0^\tau \frac{\partial}{\partial \epsilon} \frac{1}{2} \left(x'(t) - f_1(x(t), y(t) + \epsilon\phi(t)) \right)^2 \\
&\quad + \frac{\partial}{\partial \epsilon} \frac{1}{2} \left(y'(t) + \epsilon\phi' - f_2(x(t), y(t) + \epsilon\phi(t)) \right)^2 dt \bigg|_{\epsilon=0} \\
&= \int_0^\tau (x' - f_1(x, y + \epsilon\phi)) \cdot (-1) f_{1,y}(x, y + \epsilon\phi)\phi \\
&\quad + (y' + \epsilon\phi' - f_2(x, y + \epsilon\phi)) \cdot (\phi' - f_{2,y}(x, y + \epsilon\phi)\phi) \, dt|_{\epsilon=0} \\
&= \int_0^\tau (x' - f_1(x, y)) \cdot (-1) f_{1,y}(x, y)\phi \\
&\quad + (y' - f_2(x, y)) \cdot (\phi' - f_{2,y}(x, y)\phi) \, dt
\end{aligned}
$$

$$= \int_0^\tau (x' - f_1(x, y)) \cdot (-1) f_{1,y}(x, y) \phi$$

$$- \frac{d}{dt} (y' - f_2(x, y)) \phi - (y' - f_2(x, y)) \cdot (f_{2,y}(x, y) \phi) \, dt.$$

□

The exact solution, $\hat{y}$, results in the derivative of the objective function to evaluate to zero, $\frac{\partial}{\partial \epsilon} J(\hat{y} + \epsilon \phi)|_{\epsilon=0} = 0$.

This can be verified when inspecting the last step of the calculation. The factors $(x' - f_1(x, y))$ and $(y' - f_2(x, y))$ evaluate to zero, and the derivative $\frac{d}{dt} (y' - f_2(x, y))$ evaluates to zero, because the inner term $y' - f_2(x, y)$ evaluates to zero for all values of x and y.

Second Derivative of the Objective Function

We simplify the equations by inserting the exact solution. I.e., in case of the second equation of the dynamical system: $\hat{y}' = f_2(x, \hat{y}) \Rightarrow 0 = \hat{y}' - f_2(x, \hat{y})$.

In the case of the first equation of the dynamical system: $x' = f_1(x, \hat{y}) \Rightarrow 0 = x' - f_1(x, \hat{y})$.

Lemma 57 (Second Derivative, if y Matches the True, Exact Function $\hat{y}$, $y = \hat{y}$)

$$\left. \frac{\partial^2}{\partial \epsilon^2} J(y + \epsilon \phi) \right|_{\epsilon=0} = \int_0^\tau (f_{1y}(y) \cdot \phi)^2 + (\phi' - f_{2y}(y) \cdot \phi)^2 \, dt.$$

Proof:

$$\left. \frac{\partial^2}{\partial \epsilon^2} J(y + \epsilon \phi) \right|_{\epsilon=0} = \frac{\partial^2}{\partial \epsilon^2} \frac{1}{2} \int_0^\tau (x' - f_1(y + \epsilon \phi))^2 dt \tag{3.5}$$

$$+ \left. \frac{\partial^2}{\partial \epsilon^2} \frac{1}{2} \int_0^\tau ((y + \epsilon \phi)' - f_2(y + \epsilon \phi))^2 dt \right|_{\epsilon=0} \tag{3.6}$$

To avoid bulky equations, we employ auxiliary calculations for term 3.5 and term 3.6.

Term 3.5:
$$\int_0^\tau \frac{d}{d\epsilon} (x' - f_1(y + \epsilon \phi))(-1) f_{1,y}(y + \epsilon \phi) \phi \, dt$$

$$= \int_0^\tau (f_1(y + \epsilon \phi) - x') \cdot f_{1yy}(y + \epsilon \phi) \phi^2 dt$$

$$+ \int_0^\tau f_{1y}(y + \epsilon \phi) \phi f_{1y}(y + \epsilon \phi) \phi \, dt$$

$$\Rightarrow \int_0^\tau f_{1y}(y + \epsilon \phi)^2 \phi^2 dt.$$

In the exact case the term $(f_1(x, \hat{y}) - x')$ equals zero and can be neglected, which is also true in case of $\hat{y}'$ and f_2.

Term 3.6:

$$\int_0^\tau \frac{\partial}{\partial \epsilon} (y' + \epsilon \phi' - f_2(y + \epsilon \phi)) \cdot ((-1) f_{2y}(y + \epsilon \phi) \phi + \phi') dt$$

$$= \int_0^\tau (\phi' - f_{2y}(y + \epsilon \phi)\phi)((-1) f_{2y}(y + \epsilon \phi)\phi + \phi')$$

$$+ (ny' - f_2(y + \epsilon \phi) + \epsilon \phi)((-1) f_{2yy}(y + \epsilon \phi)\phi) dt$$

$$\Rightarrow \int_0^\tau (\phi' - f_{2y}(y + \epsilon \phi)\phi)^2 dt.$$

Combining the two auxiliary calculations and using $\epsilon = 0$ results in the statement.   □

We can employ the second derivative to show $\hat{y}$ is a localized optimum.

Lemma 58 (Strict Local Optimum $\hat{y}$) *If y is a minimum with $J(y) = 0$ matching the 'true' solution $y = \hat{y}$, it is a strict local optimum.*

Proof: We take the derivative from lemma 57 and the prerequisites from hypothesis 55 to show the second derivative is positive:

$$\frac{\partial^2}{\partial \epsilon^2} J(y + \epsilon \phi)|_{\epsilon=0} = \int_0^\tau f_{1y}(x, y)^2 \phi^2 + (\phi' - f_{2y}(x, y)\phi)^2 dt$$

$$\geq \int_0^\tau (f_{1,y}(x, y))^2 \phi^2 dt$$

Now, according to hypothesis 55.

$$\geq \epsilon \int_0^\tau \phi^2 \, dt = \epsilon \|\phi\|_{L^2} > 0.$$

The positive second derivative indicates a strict local minimum.   □

3.4.2 Existence of the Classical Solution

Before adding noise, we deal with the exact model. In a real-world scenario, numerical solvers are appropriate, which converge iteratively. In the case of the exact model, we prove that $\lim_{n \to \infty} J(y_n) = 0$ implies the existence of a converging subsequence.

Theorem 59 *If a sequence $y_n \in W^{1,2}$ satisfies hypothesis 55, then there is a converging subsequence $y_{n_l} \to y_\infty$ in L^2 so that $y_\infty \in C^1$ and $(x(t), y_\infty(t))$ is a classic solution of problem 53.*

Before we prove this theorem, several lemmata are required. These lemmata bound y and its sequence elements y_n in $W^{1,2}$, then y_∞ in L^∞ and y_∞ in $W^{1,2}$. After those lemmata, we analyze the convergence of the elements.

In the following, ζ_i denote non-negative constants.

Lemma 60 (Priori Bound in $W^{1,2}$) *Let $y \in W^{1,2}(I)$ for $I = [0, \tau]$. Let furthermore the function y and therefore $J(y)$ be bounded ($\|y\|_{L^\infty} \leq \zeta_1, |J(y)| \leq \zeta_2$) as stated in problem 53, definition 54, and hypothesis 55.*
Then there is ζ_3 only depending on ζ_1, ζ_2 such that $\|y\|_{W^{1,2}} \leq \zeta_3(\zeta_1, \zeta_2)$.

Proof: The function $f_2 \in C^{0,1}_{\text{loc}}(\mathbb{R}^2, \mathbb{R})$. Furthermore, x is a fixed function in C^1, and y is bounded in L^∞. Hence, the function $f_2(x, y)$ is bounded in L^2 by a constant ζ_4.
Using this bound and employing the triangle inequality, the bounded functional $J(x, y)$ yields an estimate for $\|y'\|_{L^2}$.

$$\|x' - f_1(x,y)\|_{L^2}^2 + \|y' - f_2(x,y)\|_{L^2}^2 = J(x,y) \leq \zeta_2$$
$$\Rightarrow \qquad \|y' - f_2(x,y)\|_{L^2} \leq \sqrt{\zeta_2}$$
$$\Rightarrow \qquad \|y'\|_{L^2} = \|y' - f_2(x,y) + f_2(x,y)\|_{L^2} \leq \|y' - f_2(x,y)\|_{L^2} + \|f_2(x,y)\|_{L^2}$$
$$\leq \sqrt{\zeta_2} + \zeta_4 = \zeta_5.$$

Now use $\|y\|_{L^\infty} \leq \zeta_1$ to estimate $\|y\|_{L^2}$:

$$\zeta_6 = \left(\int_I \|y\|_{L^\infty}^2 \right)^{\frac{1}{2}} \leq \left(\int_I \zeta_1^2 \right)^{\frac{1}{2}} = |I| \cdot \zeta_1.$$

Combining both estimates ζ_5, ζ_6:

$$\|y\|_{W^{1,2}} = \|y\|_{L^2} + \|y'\|_{L^2} \leq \zeta_6 + \zeta_5 \leq \zeta_3.$$

ζ_3 depends on ζ_6 and ζ_5 which depend on ζ_1 and ζ_2 proving $\|y\|_{W^{1,2}} \leq \zeta_3(\zeta_1, \zeta_2)$. $\square$

Regarding y_n is bounded in L^∞, but the minimizing sequence only controls y_n in L^2, an L^∞ bound for y_∞ has to be assured.

Lemma 61 *Let $y_n \in L^2 \cap L^\infty$ be bounded in L^∞, $\|y_n\|_{L^\infty} \leq \zeta$, and assume there is $y_\infty \in L^2$ such that $y_n \xrightarrow{n\to\infty} y_\infty \in L^2$, then $\|y_\infty\|_{L^\infty}$ is bounded such that $\|y_\infty\|_{L^\infty} \leq \zeta$.*

Proof: Let μ be the Lebesgue measure on $\mathbb{R}_+$. If $\|y_\infty\|_{L^\infty} > \zeta$ then $\exists \epsilon > 0 : \mu(\{t | y_\infty(t) \geq \zeta_1 + \epsilon) > 0$.

Let $\epsilon > 0$, and define $\Omega_\epsilon = \{t | y_\infty(t) \geq \zeta + \epsilon\}$. Then, if $\mu(\Omega_\epsilon) > 0$, i.e. the set of values exceeding the bound is not a null set:

$$\begin{aligned}
\|y_n - y_\infty\|_{L^2} &= \int_I (y_n - y_\infty)^2 dx \\
&\geq \int_{\Omega_\epsilon} (y_n - y_\infty)^2 dx \\
&\geq \int_{\Omega_\epsilon} \epsilon^2 dx > 0.
\end{aligned}$$

This contradicts $\|y_n - y_\infty\|_L^2 \xrightarrow{n\to\infty} 0$ and therefore proves $\|y_\infty\|_{L^\infty} \leq \zeta$ $\square$

We remember hypothesis 55. As $J(y_n) \xrightarrow{n\to\infty} J(y_\infty) = 0$ shows, the function y_∞ is fulfilling the objective function. Therefore, we can concentrate on subterms of the objective function from definition 54, when proving the regularity of y_∞.

Problem 62 *We concentrate on line 3.4, the second part, of the functional J from definition 54, and turn it into a weak formulation:*

$$K : W^{1,2} \cap L^\infty \times W_0^{1,2} \to \mathbb{R}$$

$$\begin{aligned}
K(y, \phi) &= \int_0^T -y \cdot \phi'(t) - f_2(x, y) \cdot \phi(t) dt \\
&= (-1) \int_0^T y \cdot \phi'(x) + f_2(x, y) \cdot \phi(t) dt.
\end{aligned}$$

After bounding y_n in $W^{1,2}$ in lemma 60 and extending this bound to an L^∞ bound for y_∞ in lemma 61 we show $y_\infty \in W^{1,2}$. We will show $y_\infty \in W^{1,2}$ using problem 62 and the boundedness of $\|y_\infty\|_{L^\infty}$ in lemma 63.

Furthermore, lemma 63 shows that y_∞ solves $K(y, \phi)$ from problem 62 in distribution, which is the last intermediate step before proving that y_∞ is a classic solution (theorem 59) of the dynamical system problem 53.

Lemma 63 (Priori Bound in $W^{1,2}$) *Let $y_n \in W^{1,2} \cap L^\infty$, $y_\infty \in L^\infty$ with $J(y_n) \xrightarrow{n\to\infty} J(y_\infty) = 0$. Let furthermore the function y_n and the functional $J(y_n)$ be bounded: $||y_n||_{L^\infty} \le \zeta_1$, $||J(y_n)|| \le \zeta_2$.*

1. $y_\infty \in W^{1,2}$.

2. $K(y_n, \phi)$ converges: $K(y_n, \phi) \xrightarrow{n\to\infty} 0$, i.e., y_∞ is a solution of problem 62.

Proof: First use the convergence of $J(y_n)$ to simplify the equation used for both claims:

$$
\begin{array}{ccccccc}
\int & -y_n \cdot \phi' & -f_2(x, y_n) & \cdot\phi dt & = K(y_n, \phi) & \le & J(y_n) \cdot ||\phi||_{L^2} \\
& \downarrow & \downarrow & & & \downarrow & n \to \infty \\
\forall \phi \in W_0^{1,2} \int & -y_\infty \cdot \phi' & -f_2(x, y_\infty) & \cdot\phi dt & & \le & 0.
\end{array}
\tag{3.7}
$$

(1) We use lemma 61, which provides a bound for y_∞, and as x is bounded for $f_2 \in C^{0,1}$. We can find a constant ζ_3 bounding y_n in the Sobolev norm: $||y_n||_{W^{1,\infty}} \le \zeta_1 \Rightarrow ||y_n||_{W^{1,2}} \le \zeta_3$. Using this bound $y_\infty \in W^{1,2}$ can be shown.

$$
\Rightarrow \quad \int -y_\infty \cdot \phi' dt - \int f_2(x, y_\infty)\phi dt \le 0
$$

$$
\Rightarrow \quad \int -y_\infty \cdot \phi' dt \le \int f_2(x, y_\infty)\phi dt \le \underbrace{\left(\int f_2(x, y_\infty)^2 dt \right)^{\frac{1}{2}}}_{=\zeta_3} \underbrace{\left(\int \phi^2 dt \right)^{\frac{1}{2}}}_{||\phi||_{L^2}}
$$

$$
\Rightarrow \quad \int -y_\infty \cdot \phi' dt \le \zeta_3 ||\phi||_{L^2},
$$

$$
\int y_\infty \cdot \phi' dt \le \zeta_3 ||\phi||_{L^2}
$$

$$
\Rightarrow \quad \left| \int y_\infty \cdot \phi' dt \right| \le \zeta_3 ||\phi||_{L^2}.
$$

Now the equivalence of $y_\infty \in W^{1,2}(I)$ and $\left| \int_I y_\infty \phi' dt \right| \le \zeta ||\phi||_{L^2}$ [Bré11, p. 268 proposition 9.3] can be used.

(2) Recall eq. (3.7), which also shows $K(y_\infty, \phi) \le 0$ for all $\phi \in W_0^{1,2}$. We conclude:

$$
\begin{cases}
\text{be } \phi \in W_0^{1,2} \to K(y_\infty, \phi) \le 0 \\
-\phi \in W_0^{1,2} \Rightarrow K(y_\infty, \phi) = -K(y_\infty, \phi) \le 0
\end{cases}
$$

$$
\Rightarrow \quad K(y_\infty, \phi) = 0.
$$

So $K(y_\infty, \phi)$ converges to zero.   □

Lemma 63 now justifies the step $y'_\infty = g(t)$ in $\mathcal{D}'$ in the proof of theorem 59. The existence of an absolute continuous antiderivative is provided by lemma 64.

Lemma 64 (Antiderivative of Regular Distributions [Wal74, p.81 (7)]) *Given a regular distribution f and let $g(t) = \int_0^t f(\tau)d\tau$. Then, g is a regular distribution and absolute continuous.*

Proof (of Theorem 59): According to lemma 63, there is $y_\infty \in W^{1,2}$, $y_{n_l} \to y_\infty \in L^2$ and y_∞ solves eq. (3.2) of problem 53 weakly. We show by bootstrapping, that $y_\infty \in C^1$ is a classical solution of eq. (3.2).

Let $y_\infty \in L^\infty$ be fixed, $\forall \phi \in W_0^{1,2} : K(y_\infty, \phi) = 0 \Rightarrow y'_\infty = g(t)$ for $g(t) = f_2(x(t), y_\infty(t)) \in L^\infty$ in distribution.

$$\Rightarrow \text{Lemma 64} \qquad y_\infty(t) = \int_0^t g(\tau)d\tau + \zeta' \in C^0 \qquad (3.8)$$

$$\Rightarrow \qquad g(t) = f_2(x(t), y_\infty(t)) \qquad (3.9)$$

$$\Rightarrow \text{Lemma 64} \qquad y_\infty(t) = \int_0^t g(\tau)d\tau + \zeta' \in C^1 \qquad (3.10)$$

$$\Rightarrow \qquad y'_\infty = g(t) \text{ for } g(t) \in C^1 \text{ in } \mathcal{D}'. \qquad (3.11)$$

In eq. (3.8) the regularity of $y_\infty(t)$ is lifted by integration from being bounded to C^0, which can be used in eq. (3.9) to lift the regularity of $g(t)$ from the original $||g(t)||_{L^\infty} < \zeta$ to $g(t) \in C^0$, because $f_2 \in C_{\text{loc}}^{0,1}(\mathbb{R}^2, \mathbb{R})$.

We use the deduced property $g(t) \in C^0$ in eq. (3.10) to deduce a better regularity for y_∞, i.e., $y_\infty \in C^1$. In turn, that better regularity enables us to calculate the derivative of y_∞ from $g(t)$ in eq. (3.11).   □

3.5 Noise

In the next step, we analyze the behavior of the solution of the dynamical system from problem 53 in the presence of noise. White noise is a common form of noise.
Heuristically, continuous white noise is formally defined as the derivative of the Brownian motion [FK05, p.124]. However, there is no derivative of the Brownian motion, because there is no regularity [KK93, p. 23]. There are approaches that stick close to the definition and circumvent the fact that the Brownian motion can not be derived using a system of first-order Itô stochastic equations [FK05, p.124]. A continuous white noise signal can not exist in reality. Looking at the power spectral density, which would cover all frequencies, there would be an infinite amount of energy within this signal [KK11, p.45], [Mül90, p. 88].
In a more practical approach, white noise is defined by a constant power spectral density on a defined bandwidth [Mül90, p.42 Fig.13]. When analyzing the practical example of an electrical circuit, the observed white noise visualized in fig. 3.5 has a lower and upper bound. The lower bound is determined by red noise, which needs long observations to become visible. On the other hand, the high frequencies are limited by the circuit bandwidth [Gar04, p. 19].

Figure 3.5: Noise contributions in the frequency domain according to [TS11, Fig. 2.17, p.32].

In contrast to time-continuous white noise, discrete white noise can be created by drawing samples from a normal distribution. We could define a continuous noise function as identical to the discrete white noise at the discrete samples, e.g., using sinc interpolation. Because the following considerations do not rely on the independence properties of a hypothetical white noise, nor on the scale of the noise, we decided to not deal with the complications

of defining a suitable noise model fulfilling a particular definition of noise[1] and introduce an arbitrary noise function $n_c \in W_0^{1,2}$.

3.6 Modeling Noise

We introduce a function $n(t) \in W_0^{1,2}$, which plays the role of the observation-noise within the rest of the chapter.

Definition 65 (Continuous Noise) The observed noise is described by $n(t) \in W_0^{1,2}$. The subscript $n_x(t)$ or $n_y(t)$ can denote the corresponding state variable $x(t)$ or $y(t)$.

This continuous noise is added when observing the process. This results in a new definition of the previous problem 53.

Problem 66 (Dynamical System Observed under Gaussian White Noise) *For $\tilde{x}(t) \in C^1$, $f_1, f_2 \in C_{loc}^{0,1}(\mathbb{R}^2, \mathbb{R})$ and $\tau \in \mathbb{R}_+$ find $\tilde{y}(t) \in W^{1,2}$ such that:*

$$\tilde{x}'(t) = f_1(\tilde{x}(t), \tilde{y}(t)), \quad t \in [0, \tau], \tag{3.12}$$

$$\tilde{y}'(t) = f_2(\tilde{x}(t), \tilde{y}(t)), \quad t \in [0, \tau], \tag{3.13}$$

$$\tilde{x}(t) = x(t) + n_x(t). \tag{3.14}$$

---

[1]Nevertheless, a closer analysis of the robustness against Poisson noise could be beneficial for practical implementations.

3.7 Behavior of Solution – Noisy Input

Now we are going to investigate the influence of noise on the optima and prove a continuous influence at the transition from no noise to a noisy situation. The introduction of noise can lead to altered optima. In the following section, we are investigating how the introduction of noise influences an optimum by introducing a parameter λ, which scales the influence of noise starting at zero influence.

3.7.1 Trace the Optimum by the Implicit Function Theorem

We intend to use the implicit function theorem. Therefore, we show three properties of the operator $A(\lambda, u)[\phi] = \frac{\partial}{\partial \epsilon} J_N(\lambda, u + \epsilon \phi)|_{\epsilon=0}$ derived in lemma 68. We derive the loss function to find the minimum. This is accomplished by the Fréchet derivative of the functional from definition 67, which is providing the optimum under the influence of noise, $\frac{\partial}{\partial \epsilon} J_N(\lambda, u + \epsilon \phi)|_{\epsilon=0} = 0$. Then the implicit function theorem is used to find the trajectory of minima under the variation of noise. The invertibility of the Operator will be shown. Therefore, we show the invertibility of the functional in its second argument. We will prove the invertibility using the linearity.

Then, after having shown the coercivity, the Lax-Milgram theorem ensures the invertibility of the operator.

We start by introducing the noise n_x from definition 65, which will be added to the model.

Now we use definition 65 instead of definition 54 to include noise. Previously, $\hat{y}$ was defined in $W^{1,2}$. To add noise and restrict the variation of $\hat{y}$, which we call u, to $W_0^{1,2}$, we define a new functional.

Definition 67

$$J_N : U \subset \mathbb{R} \times W_0^{1,2} \to \mathbb{R}$$
$$(\lambda, u) \mapsto J_N(\lambda, u)$$
$$J_N(\lambda, u) = \frac{1}{2} \int_0^T ((x + \lambda n_x)' - f_1(x + \lambda n_x, \hat{y} + u))^2 \, dt$$
$$+ \frac{1}{2} \int_0^T ((\hat{y} + u)' - f_2(x + \lambda n_x, \hat{y} + u))^2 \, dt.$$

We derive the functional from definition 67 , which fulfills the equation $J_N(0, 0) = 0$ from definition 67 in the case $\lambda = 0$, as $\hat{y}$ was fulfilling $J(\hat{y}) = 0$ from definition 54 in the absence of noise. This is a global minimum, and therefore, the derivative in the direction of y is zero.

Lemma 68 *Derivative of $J_n(\lambda, u)$ at $\lambda = 0$ in direction of y.*

$$A : \mathbb{R} \times W_0^{1,2} \to \mathcal{L}(W_0^{1,2}, \mathbb{R})$$

$$(\lambda, u) \mapsto A(\lambda, u)$$

$$A(\lambda, u)[\phi] = \frac{\partial}{\partial \epsilon} J_N(\lambda, u + \epsilon\phi)|_{\epsilon=0}$$

$$= \int_0^T ((x + \lambda n_x)' - f_1(x + \lambda n_x, \hat{y} + u))(-1)f_{1y}(x + \lambda n_x, \hat{y} + u)\phi \, dt$$

$$+ \int_0^T ((\hat{y} + u)' - f_2(x + \lambda n_x, \hat{y} + u)) \cdot (\phi' - f_{2y}(x + \lambda n_x, \hat{y} + u)\phi) \, dt.$$

Proof:

$$A : \mathbb{R} \times W_0^{1,2} \to \mathcal{L}(W_0^{1,2}, \mathbb{R})$$

$$(\lambda, u) \mapsto A(\lambda, u)$$

$$A(\lambda, u)[\phi] = \frac{\partial}{\partial \epsilon} J_N(\lambda, u + \epsilon\phi)|_{\epsilon=0}$$

$$= \frac{\partial}{\partial \epsilon} \left(\frac{1}{2} \int_0^T ((x + \lambda n_x)' - f_1(x + \lambda n_x, \hat{y} + u + \epsilon\phi))^2 \, dt \right.$$

$$\left. + \frac{1}{2} \int_0^T ((\hat{y} + u + \epsilon\phi)' - f_2(x + \lambda n_x, \hat{y} + u + \epsilon\phi))^2 \, dt \right)\Bigg|_{\epsilon=0}$$

$$= \left(\int_0^T ((x + \lambda n_x)' - f_1(x + \lambda n_x, \hat{y} + u + \epsilon\phi)) \right.$$

$$\cdot (-1)f_{1,y}(x + \lambda n_x, \hat{y} + u + \epsilon\phi)\phi \, dt$$

$$+ \int_0^T ((\hat{y}' + u' + \epsilon\phi' - f_2(x + \lambda n_x, \hat{y} + u + \epsilon\phi))$$

$$\left. \cdot (\phi' - f_{2,y}(x + \lambda n_x, \hat{y} + u + \epsilon\phi)\phi) \, dt \right)\Bigg|_{\epsilon=0}$$

$$= \int_0^T ((x + \lambda n_x)' - f_1(x + \lambda n_x, \hat{y} + u)) \cdot (-1)f_{1,y}(x + \lambda n_x, \hat{y} + u)\phi \, dt$$

$$+ \int_0^T ((\hat{y}' + u' - f_2(x + \lambda n_x, \hat{y} + u)) \cdot (\phi' - f_{2,y}(x + \lambda n_x, \hat{y} + u)\phi) \, dt.$$

$$\square$$

Because we introduced noise, generically the optima are shifted. So instead of an exact solution of the ODE from problem 66 leading to $J_N(\lambda, u) = 0$, there will only be an available approximation $J_N(\lambda, u) \geq 0$. However, this approximation is still a local minimum fulfilling $A(\lambda, u) = 0$.

The introduction of noise by the parameter λ and the resulting shift of the optima $\hat{y}$ to

$\tilde{y} = \hat{y} + u$ now raises the question of the trajectory of this shift of the optimum. The definition of the trajectory by $A(\lambda, u) = 0$ resembles the implicit function theorem [Bag92, p.257, th. 12.8][1].

Theorem 69 (Implicit Function Theorem) *[Bag92, p.257, th. 12.8] Let the following pre-conditions be fulfilled:*

- *E and F are Banach spaces with the max norm.*

- *There is a map f of an open subset O in $E \times F$ into F, which is continuously differentiable at a point $x = (x_1, x_2) \in O$.*

- *There is a bijective, linear transformation $T(w) = df_x(0, w) \subseteq F \to F$.*

Then, there exists the neighborhoods U_1 and U_2 of x_1 in E and x_2 in F. Furthermore, there exists a unique continuous function $g : U_1 \to U_2$. Its graph coincides with the level set $f^{-1}(f(x)) \cap (U_1 \times U_2)$.

The spaces of the transformation $T : E \times F \to F$ as stated in [Bag92, p.257, th. 12.8] are fulfilled by A due to the property of $W_0^{1,2}$ to be a Hilbert space, i.e. $\mathcal{L}(W_0^{1,2}, \mathbb{R})$ can be identified as $W_0^{1,2}$ [Yos68, p.91, Cor. 1]. To ensure the solvability of the equation, the derivative of A in the direction of y has to result in an invertible operator.

Lemma 70 (Derivative of $A(\lambda, u)$) *We define $B_{(\lambda=0)}[u, \phi]$ to be the derivative of $A(\lambda, u)$ at the scale of the noise $\lambda = 0$:*

$$B_{(\lambda=0)}[u, \phi] : \mathbb{R} \times W_0^{1,2} \to \mathcal{L}(W_0^{1,2}, \mathbb{R})$$

$$(\lambda, u) \mapsto \frac{d}{d\epsilon} A(\lambda, \epsilon u)[\phi]\Big|_{\epsilon=0}$$

$$B_{(0)}[u, \phi] = \int_0^T (f_{1,y}(x, y))^2 \phi u \, dt + \int_0^T (\phi' - f_{2y}\phi)(u' - f_{2y}u) \, dt.$$

Proof:

$$B_{(\lambda=0)}[u, \phi] = \frac{d}{d\epsilon} A(\lambda, \epsilon u)[\phi]\Big|_{\epsilon=0}$$

$$= \frac{d}{d\epsilon} \int_0^T (x' - f_1(x, \hat{y} + \epsilon u))(-1) f_{1y}(x, \hat{y} + \epsilon u) \cdot \phi \, dt\Big|_{\epsilon=0}$$

$$+ \frac{d}{d\epsilon} \int_0^T (\hat{y}' + \epsilon u' - f_2(x, \hat{y} + \epsilon u))(\phi' - f_{2,y}(x, \hat{y} + \epsilon u)\phi) \, dt\Big|_{\epsilon=0}$$

---

[1] The theorem is also covered in the chapter 4.8 'The Implicit Function Theorem' of the book 'Applied Functional Analysis - Main Principles and Their Applications' [Zei95, p.250, Ch. 4.8] and [Bag92, p.257, th. 12.8].

$$= \int_0^T (x' - f_1(x, \hat{y} + \epsilon u))(-1)f_{1y}(x, \hat{y} + \epsilon u) \cdot \phi \cdot (-1)f_{1,y}(x, \hat{y} + \epsilon u) \cdot u \ dt \Big|_{\epsilon=0}$$

$$+ \int_0^T (x' - f_1(x, \hat{y} + \epsilon u))(-1)f_{1yy}(x, \hat{y} + \epsilon u)u \cdot \phi \ dt \Big|_{\epsilon=0}$$

$$+ \int_0^T (u' - f_{2,y}(x, \hat{y} + \epsilon u)u)(\phi' - f_{2,y}(x, \hat{y} + \epsilon u)\phi) \ dt \Big|_{\epsilon=0}$$

$$+ \int_0^T (\hat{y}' + \epsilon u' - f_2(x, \hat{y} + \epsilon u))(-1)f_{2,yy}(x, \hat{y} + \epsilon u) \cdot \phi \cdot u \ dt \Big|_{\epsilon=0}$$

After evaluating $\epsilon = 0$, we use the exact property of x and $\hat{y}$ in the absence of noise.

$$= \int_0^T (-1)f_{1,y}(x, \hat{y}) \cdot u(-1)f_{1y}(x, \hat{y})\phi \ dt$$

$$+ \int_0^T \underbrace{(x' - f_1(x, \hat{y}))}_{=0}(-1)f_{1,yy}(x, \hat{y}) \cdot u \cdot \phi \ dt$$

$$+ \int_0^T (u' - f_{2,y}(x, \hat{y})u)(\phi' - f_{2,y}(x, \hat{y})\phi) \ dt$$

$$+ \int_0^T \underbrace{(\hat{y}' - f_2(x, \hat{y}))}_{=0}(-1)(f_{2yy}(x, \hat{y}) \cdot u \cdot \phi) \ dt.$$

Now evaluate B at $\hat{y}$ for $\lambda = 0$.

$$B_{(0)}[u, \phi] = \int_0^T (f_{1,y}(x, y))^2 \phi u dt + \int_0^T (\phi' - f_{2y}\phi)(u' - f_{2y}u) \ dt.$$

$\square$

The invertibility of B can now be shown by applying the Lax-Milgram theorem [RR04, p.290, Theorem 9.14], which is quoted in theorem 75[1]. The theorem requires $B(x, y)$ to be defined on a product Hilbert space $X \times X$, which satisfies bilinearity, boundedness, and positivity.

Definition 71 (Bilinearity, Boundedness, and Positivity) [RR04, p.290, Theorem 9.14]
Let X be a Hilbert space and $B(x, y)$ defined on $X \times X \to \mathbb{R}$.
• Bilinearity, $\forall \alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$:

$$B(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1 B(x_1, y) + \alpha_2 B(x_2, y) \text{ and}$$

$$B(x, \beta_1 y_1 + \beta_2 y_2) = \beta_1 B(x, y_1) + \beta_2 B(x, y_2).$$

• Boundedness, $\exists \, \zeta > 0$:

$$|(x, y)| \leq \|x\| \cdot \|y\| \cdot \zeta.$$

---

[1]A complex version of the Lax-Milgram theorem can be found in [Yos68, p.92, ch.III.7].

• Coercivity , $\exists \delta > 0$:

$$B(x, x) \geq \delta \|x\|^2.$$

Corollary 72 $B_{(0,y)} : W_0^{1,2} \times W_0^{1,2} \to \mathbb{R}$ *is bilinear.*

Now we show the boundedness.

Lemma 73 ($B_{(0)}[u, \phi]$ Is Bounded.) $|B_{(0)}[u, \phi]| \leq \zeta \|u\|_{W^{1,2}} \cdot \|\phi\|_{W^{1,2}}$.

Proof: To achieve a bound for the second part of the functional, it is essential that u and ϕ are estimated in the Sobolev norm to have an upper bound for the derivatives occurring in part two of B. We use a shortened notation for the i-th component of f in direction of x to improve the readability: $f_{ix} = f_{ix}(t) = f_{ix}(x(t), y(t))$.

$$|B_{(0)}[u, \phi]| \leq \underbrace{|\int_0^T (f_{1,y}(x, y))^2 \phi u \, dt|}_{\text{Part 1}} + \underbrace{|\int_0^T (\phi' - f_{2y}\phi)(u' - f_{2y}u)dt|}_{\text{Part 2}}$$

$$\text{Part 1} \leq \|\mathbb{1}_{[1,T]} f_{1,y}^2\|_{L^\infty} \cdot \int_0^T |\phi u| dt \leq \|\mathbb{1}_{[1,T]} f_{1,y}^2\|_{L^\infty} \cdot \|\phi\|_{L^2} \| u\|_{L^2}$$

$$\text{Part 2} \leq \|\phi' - f_{2y}\phi\|_{L^2} \|u' - f_{2y}u\|_{L^2} \leq (\|\phi'\|_{L^2} + \|f_{2y}\phi\|_{L^2}) \cdot (\|u'\|_{L^2} + \|f_{2y}u\|_{L^2})$$

$$\leq (\|\phi'\|_{L^2} + \|\mathbb{1}_{[1,T]} f_{2y}\|_{L^\infty} \|\phi\|_{L^2}) \cdot (\|u'\|_{L^2} + \|\mathbb{1}_{[1,T]} f_{2y}\|_{L^\infty} \|u\|_{L^2})$$

$$= \|\phi'\|_{L^2} \|u'\|_{L^2} + \|\mathbb{1}_{[1,T]} f_{2y}\|_{L^\infty}^2 \|u\|_{L^2} \|\phi\|_{L^2}$$

$$+ \|\mathbb{1}_{[1,T]} f_{2y}\|_{L^\infty} (\|u'\|_{L^2} \|\phi\|_{L^2} + \|u\|_{L^2} \|\phi'\|_{L^2})$$

Now add all the factors.

$$\zeta := \|\mathbb{1}_{[1,T]} f_{1,y}^2\|_{L^\infty} + \|\mathbb{1}_{[1,T]} f_{2,y}\|_{L^\infty} + \|\mathbb{1}_{[1,T]} f_{2,y}\|_{L^\infty}^2$$

Collect the auxiliary computations into one equation.

$$|B_{(0,y)}[u, \phi]| \leq (\|\mathbb{1}_{[1,T]} f_{1,y}^2\|_{L^\infty} + \|\mathbb{1}_{[1,T]} f_{2,y}\|_{L^\infty}^2) \cdot \|\phi\|_{L^2} \| u\|_{L^2} + \|\phi'\|_{L^2} \|u'\|_{L^2}$$

$$+ \|\mathbb{1}_{[1,T]} f_{2y}\|_{L^\infty} (\|u'\|_{L^2} \|\phi\|_{L^2} + \|u\|_{L^2} \|\phi'\|_{L^2})$$

$$\leq \zeta \cdot \|u\|_{W^{1,2}} \cdot \|\phi\|_{W^{1,2}}.$$

$\square$

Lemma 74 (Coercivity) *There exists a constant δ such that $B(x, x) \geq \delta \|x\|_{W^{1,2}}^2$.*
To prove this lemma the precondition $(f_{1,y}(x(t), y(t)))^2 \geq \epsilon > 0$ of hypothesis 55 has to be fulfilled.

Proof: Recall that $|a \cdot b| \leq \epsilon a^2 + \frac{1}{\epsilon} b^2$, as $0 \leq (\sqrt{\epsilon}a + \frac{1}{\sqrt{\epsilon}}b)^2$. In the following equation, we will use this auxiliary calculation to split the mixed term of the expansion. The ϵ will be used to distribute the term in a way, which does vanish neither η, nor $\tilde{\eta}$. This is possible due to the precondition $\|f_{1,y}\|^2 > 0$, which can balance the arbitrary small $(1-\epsilon)\|f_{2,y}\|$.

$$B_0(\phi, \phi) = \int_0^T (f_{1,y})^2 \phi^2 dt + \int_0^T (\phi' - f_{2,y}\phi)^2 dt$$

$$= \int_0^T (f_{1,y})^2 \phi^2 dt + \int_0^T (\phi')^2 dt - 2\underbrace{\int_0^T f_{2,y}\phi\phi' dt}_{*} + \int_0^T (f_2 y)^2 \phi^2 dt$$

Now apply Young's inequality at * $\left(-2|\int_0^T f_{2y}\phi \cdot \phi' dt| \geq -\epsilon \int_0^T f_{2y}^2 \phi^2 dt - \frac{1}{\epsilon}\int_0^T \phi'^2 dt \right)$.

$$\geq \int_0^T \underbrace{\left(f_{1,y}^2 + f_{2,y}^2 - \epsilon f_{2,y}^2\right)}_{\eta} \phi^2 dt + \int_0^T \underbrace{\left(1 - \frac{1}{\epsilon}\right)}_{\tilde{\eta}} \phi'^2 dt$$

$$\underbrace{\geq}_{\exists \epsilon > 1} \min(\eta, \tilde{\eta})\|\phi\|_{W^{1,2}}^2.$$

The possibility of having a bound $\epsilon > 1$, while still allowing $\eta > 0$ due to the precondition, ensures the positivity of η and $\tilde{\eta}$. $\qquad\square$

Theorem 75 (Lax-Milgram) *[RR04, p.290, Theorem 9.14] [Yos68, p.92] Let $B : W_0^{1,2} \times W_0^{1,2} \to \mathbb{R}$ be a functional defined on the product Hilbert space $W_0^{1,2} \times W_0^{1,2}$ which satisfies the conditions: Bilinearity, boundedness and coercivity.*
Then there exists a uniquely determined bounded linear operator S with $\|S\| \leq \delta^{-1}$ and a bounded linear inverse S^{-1} with $\|S^{-1}\| \leq \gamma$ such that $(x, y) = B(x, Sy)$.

Corollary 76 *Because bilinearity, boundedness, and positivity are proven in corollary 72 and lemmata 73 and 74, using Lax-Milgram, theorem 75, $f_{1,y} > \delta > 0$ means $B(y, \phi)$ can be inverted.*

3.7.2 Quantify the Influence of Noise and Upper Bound for Condition Number

It is easy to see that δ^{-1} provides an upper bound for the immediate influence of noise. The higher the value of δ^{-1}, the larger the possible initial influence of noise in altering the result at the transition from zero noise to a noisy observation.

Evidently, an upper bound for a condition number like value was calculated as a by-product of proving the boundedness and coercivity, by δ, γ from theorem 75 as it states: $\|S^{-1}\| \leq \gamma$ and $\|S\| \leq \delta^{-1}$.

Remark 77 (Condition Number Like Value - Upper Bound) $\tilde{\kappa} = \|S^{-1}\|\|S\| \geq \delta^{-1}\gamma$.

3.7.3 Example

This subsection shows a practical example of recovering an unobserved variable using the Gradient Matching ansatz by minimizing a loss function $J(y_i)$ on discrete space.

We generate some data by taking the Lotka-Volterra equations modeling a predator-prey system perturbed by poisson noise. We will recover the trajectory of the predator. The data generation restarts using different random initial conditions to show the Gradient Matching to be independent of said initial conditions.

Problem 78 (Lotka-Volterra Example) *There is a prey $x(t)$ and a predator $y(t)$, $t \in [0, 40]$.*

$$\dot{x}(t) = 0.5 \cdot x(t) - 0.1 \cdot x(t)y(t), \tag{3.15}$$

$$\dot{y}(t) = -0.9 \cdot y(t) + 0.03 \cdot x(t)y(t). \tag{3.16}$$

The simulation, the trajectory, is restarted at time points τ_i drawn from a Poisson distribution with parameter $\lambda = 0.01$. These time points define the positions, marking the restarts to simulate Poisson noise. The initial conditions are uniformly drawn $x(\tau_i+) \sim \mathcal{U}([7, 50])$ and $y(\tau_i+) \sim \mathcal{U}([3, 9])$ using fixed-point arithmetic with accuracy 10^-3. The trajectory is generated using the 'ode' function from the 'deSolve' package using 'lsoda' in its default configuration. We take an array of samples on a discrete space $t_{i+1} - t_i = 0.1$, $t \in [0, 40]$. Finally, additive Gaussian white noise is added by drawing values from $\mathcal{N}(0, 1.2)$. The result of this simulation can be seen in fig. 3.6.
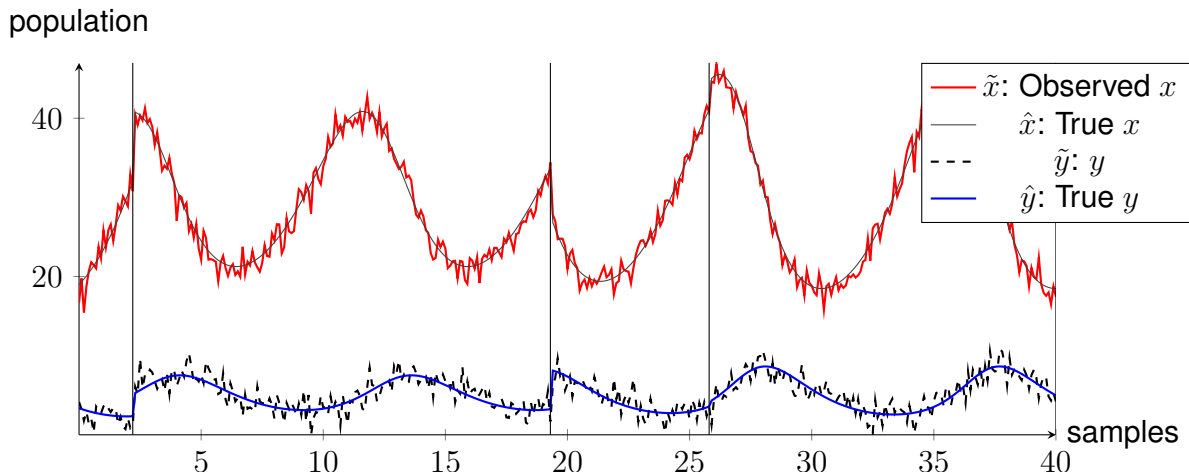
Figure 3.6: The example system showing the true functions $\hat{x}$ and $\hat{y}$ and the observed function $\tilde{x}$ and the function $\tilde{y}$ as it could have been observed.

After preparing the data shown in fig. 3.6, the noisy, observed $\tilde{x}(t_i)$ is handed to the algorithm. We describe the separate steps of the program running the algorithm in more

detail in section 6.2.2. The initial data for $\tilde{y}(t_i)$ are uniformly distributed integer values $y(t) \sim \mathcal{U}([0, 100])$. Before handing the discrete version of the objective function stated in definition 54 to the solver 'rgenoud' using BFGS and genetic optimization, we compute the derivative of $x(t)$ using total variation differentiation [Cha11]. The differentiation within the objective function is done using a five-point stencil. All differentiations are capped to their 99% quantile to further limit the influence of the jumps, which result in exceptionally high derivatives. By this approach, the influence of these derivatives, which do not fulfill the differential equations, is limited.

The result of this optimization compared to the 'true', noise-free data can be seen in fig. 3.7.
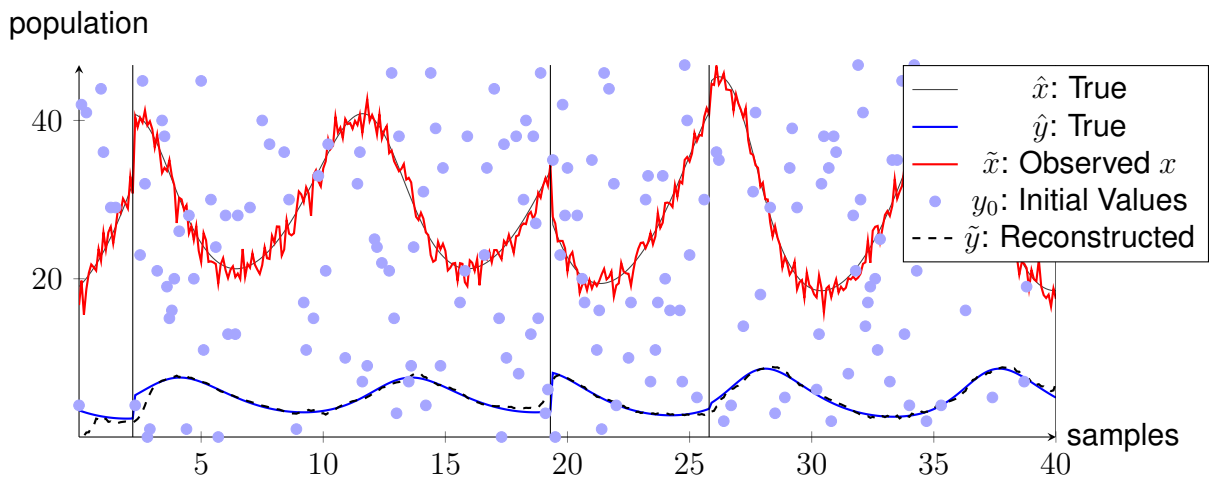
Figure 3.7: The reconstructed $\tilde{y}$ compared to the unobserved original $\hat{y}$ given the observed function $\tilde{x}$ and the starting values y_0. The vertical lines indicate the time points of the Poisson noise.
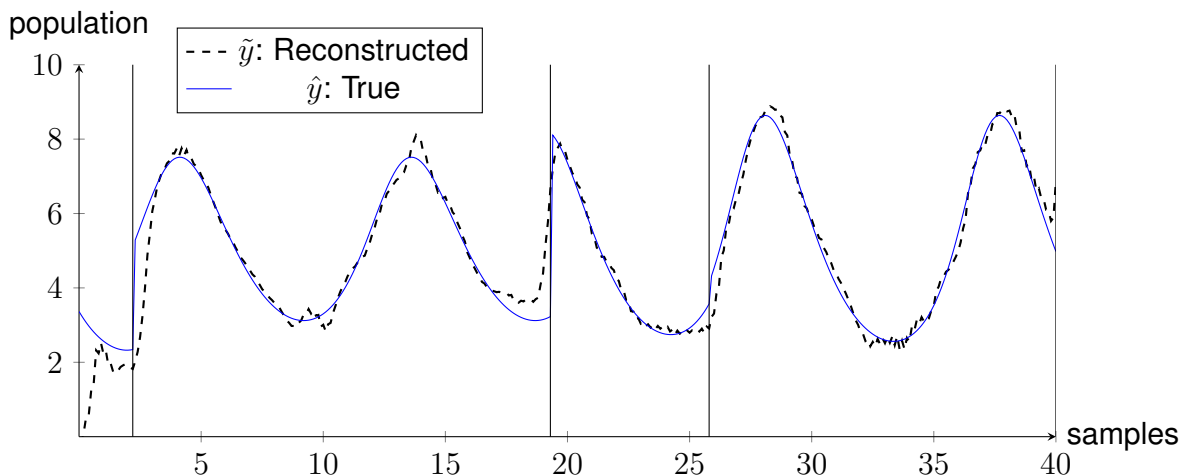
Figure 3.8: Here only the reconstructed $\tilde{y}$ and the true $\hat{y}$ are shown.

As the scale of $\hat{y}(t)$ and $\tilde{y}(t)$ results in a hardly observable difference $\hat{y}(t) - \tilde{y}(t)$ the plot,

fig. 3.8, shows a rescaled version of the previous curves allowing a judgment of the quality of the approximation.

3.7.4 Uniqueness Example

We understand that the addition of noise and the involved optimization can result in local optima. This section will provide an example to illustrate the existence of non-unique solutions in the exact case. Furthermore, we hint at the existence of local optima in the presence of noise not resembling a solution of the exact case.

We start constructing an example, which allows for two initial conditions for the unobserved state variable resulting in two separate unique solutions in the analytical sense.

We start this construction using a dynamical system of two state variables, x and y. The variable y will denote the unobserved state variable. We decided to use a system, which results in a repeating pattern, an oscillation resembling sine and cosine (see fig. 3.9).

Problem 79 (Basic Dynamical System)

$$\dot{x} = y,$$
$$\dot{y} = -x.$$

The basic problem 79 will now serve as a basis to extend it to a dynamical system with two equivalent solutions. As an intermediate step, we will drop the prerequisite of the right-hand side to be polynomial. Nevertheless, it will keep the Lipschitz continuity.

We will shift the levels of the state variables x and y and control the influence of the sign manually.

First, we shift the state variable x upwards. There-

Figure 3.9: Solution of problem 79 using the initial conditions $x(0) = 1$ and $y(0) = 0$.

fore, we need to replace x by $(x - 1)$ to still positively and negatively influence y.

The influence of y on x should be independent of the sign of y. Therefore, we square y and realize the change of the sign of y influencing the first equation by subtracting one. The factor for y in the first equation is empirical.

We allow a second solution for y with an opposite sign by multiplying an approximation of the signum function to the second equation, $\frac{y}{\sqrt{y^2 + \epsilon}}$.

Problem 80 (Non-Unique Problem)

$$\dot{x} = (2.6 \cdot y)^2 - 1,$$

$$\dot{y} = -\frac{y}{\sqrt{y^2 + \epsilon}} \cdot (x - 1),$$

$$\epsilon = 0.001,$$

$$x(0) = 0.5, \; y_1(0) = 0.6, \; y_2(0) = -0.6.$$
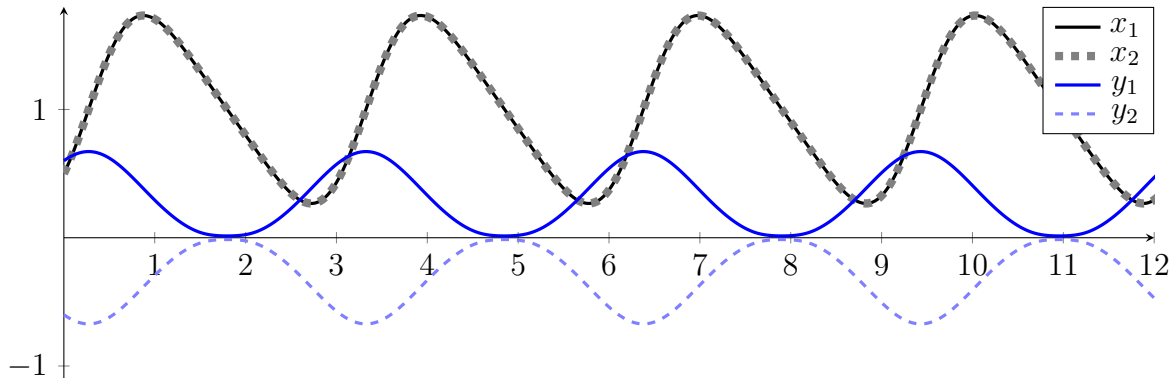
Figure 3.10: Solution of problem 80.

Figure 3.10 shows the two solutions of problem 80, which have an identical solution for x, but two solutions y_1 and y_2 for the state variable y depending on the sign of the initial condition. The figure visualizes that the two solutions y_1 and y_2 become almost identical when approaching zero. As Gradient Matching is a local approach, it seems possible that the algorithm mixes these solutions in the presence of noise. Such a mixed curve combining parts from y_1 and y_2 into a new, seemingly continuous curve does not resemble a solution to the problem. However, it might have a considerable potential to resemble a local optimum when searching for an optimal solution by optimization.

Polynomial Problem

The approximation to the signum function $\frac{y}{\sqrt{y^2+\epsilon}}$ is not a polynomial. Nevertheless, it resembles a smooth function, which can be turned into a polynomial by approximating it using a Taylor series expansion. Obviously, the limit $\epsilon \to 0$ is not a smooth function. Smaller ϵ values quickly increase the effort it takes to achieve a sufficient Taylor series expansion covering the range of the function. However, a relatively conservative choice of a fixed ϵ, which we use below, and a Taylor series expansion up to the order of 17 are sufficient to provide a polynomial right-hand side for this example.

Problem 81 (Polynomial Problem)

$$\dot{x} = -(2.4 \cdot y)^2 + 0.2,$$

$$\dot{y} = \left(\frac{0.1 \cdot Y}{\sqrt{\epsilon}} - \frac{(0.1 \cdot Y)^3}{2 \cdot \epsilon^{3/2}} + \frac{(3 \cdot (0.1 \cdot Y)^5}{8 \cdot \epsilon^{5/2}} - \frac{5 \cdot (0.1 \cdot Y)^7}{16 \cdot \epsilon^{7/2}} + \frac{35 \cdot (0.1 \cdot Y)^9}{128 \cdot \epsilon^{9/2}} \right.$$

$$\left. - \frac{63 \cdot (0.1 \cdot Y)^{11}}{256 \cdot \epsilon^{11/2}} + \frac{231 \cdot (0.1 \cdot Y)^{13}}{1024 \cdot \epsilon^{13/2}} - \frac{429 \cdot (0.1 \cdot Y)^{15}}{2048 \cdot \epsilon^{15/2}} + \frac{6435 \cdot (0.1 \cdot Y)^{17}}{32768 \cdot \epsilon^{17/2}} \right)$$

$$\cdot (X - 0.2),$$

$$\epsilon = 0.001, \ x(0) = 0.45, \ y_1(0) = 0.15, \ y_2(0) = -0.15.$$

Figure 3.11 visualizes two possible solutions for problem 81, having an identical solution for state variable x and two different solutions y_1 and y_2, for the state variable y.
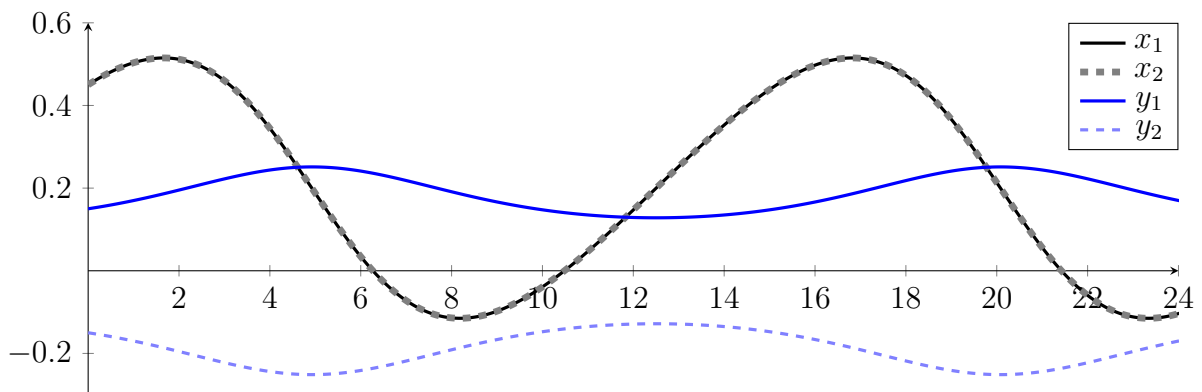
Figure 3.11: Solution of problem 81.

3.8 Summary and Outlook

In summary, this chapter presented an approach to recover an unobserved state variable of a dynamical system using Gradient Matching. We proved the existence of a classic solution without noise regarding the preconditions. In a second step, we added a function resembling noise and hinted at the continuous effect on the solution depending on the scale of this function.

The ansatz leaves many topics for further investigation. A rigorous analysis of the observability based on existing theory like the Luenberger observer would help to describe the performance and limits of the ansatz. It would be interesting to investigate if identifying the parameters and reconstruction of unknown state variables could be combined. E.g., is it possible to observe the number of predators in a predator-prey system and estimate the number of the prey and the parameters of the system?

This combination might be achieved within an expectation-maximization algorithm, which would resemble some similarities to algorithms for blind deconvolution [PF16]. In the case of blind deconvolution.

Furthermore, there are other interesting approaches to parameter identification. Therefore, we would like to locate our Gradient Matching approach using penalized regression and a polynomial right-hand side among some of these alternative approaches to mention possible directions for further research.

Suppose a space possesses an orthogonal basis, and the parameters coincide with the direction of this orthogonal basis. In that case, the identification is a simple task, which a projection can perform on the existing basis.

Polynomials are no orthogonal basis for functions. The monomials can be seen as an overcomplete set spanning the space. We call this spanning set the dictionary. The identification of the parameters within this set is called dictionary learning [Tro+05][Tro06][NT09]. We will use an approach to identify a sparse representation within this dictionary in chapter 4 to identify the parameters. The dictionary allows for sparse parameters, which provide an interpretation, which is less likely, if we are constrained to an orthogonal basis.

Large data sets might enable to learn the dictionary from the observed data [IB20] [DI18, p. 216 Remark 8.4]. If the reconstructed dictionary does not consist of monomials, it might lead to further insights into the observed problem.

Independent of the dictionary learning approach, disentanglement by β-variational autoencoder [Bur+18] might be an interesting approach using neuronal networks. The autoencoder might detect transformations, which are not reproducible by a sparse set of parameters found within the dictionary approach.

4 Gradient Matching Experiment

To show the applicability of the methods used in chapter 3 'Gradient Matching'[1] and section 3.2 'Latent Function Recovery Using Gradient Matching'[2], we perform three experiments. We will show on a first dataset how Gradient Matching will successfully estimate the parameters of a dynamical system. In a second experiment, we will show the robustness of the algorithm. Therefore, we will use a new dataset, which consists of multiple trajectories of a dynamical system, and we adjust the penalty parameter τ to allow for a smaller, a medium, and a larger set of parameters to be estimated. All sets of parameters will cover the most important parameters and characteristics of the dynamical system. However, the local nature of Gradient Matching and invariance to gradual drift might influence the sensitivity towards such global changes like drift and damping. The third experiment adds a new dataset, which contains a state variable of bad quality as an example of unobserved data. In this case, there are two goals. First, we will show that the algorithm successfully reconstructs the unobserved data, and the recovered state variable will be more precise than the observed one. Then, we will repeat the algorithm using the good dataset. This good dataset will support the hypothesis that the algorithm recovers the state variable, which we measured.

The dynamical system providing the data will be a harmonic oscillator consisting of an inductor L, a capacitor C, and a resistor R. We will start the oscillation by a short impulse, which resembles a Dirac impulse. Our oscilloscope can measure the voltage directly. It does not measure the current, but the voltage of a small shunt resistor. The voltage of the shunt resistor allows us to calculate the current by employing Ohm's law $U = I \cdot R$ using the known resistance. However, the shunt resistor increases the energy loss of every oscillation, the damping. The damping stops the visible oscillation after around 10 periods. Retriggering the oscillation and starting a new trajectory by periodic Dirac impulses is a convenient way to record more data. The Gradient Matching can process the data covering multiple trajectories in a single run without separating them, because it is invariant to the change of the trajectories.

---

[1] We describe the corresponding program flow to estimate the parameters in section 6.2.1.
[2] We describe the corresponding program flow to estimate unobserved components in section 6.2.2.

The current we observe by employing the low resistance of the shunt resistor is the hard to observe state variable of the third experiment. The resistance has to be low to not add too much to the internal resistance of the inductor and capacitor. On the other hand, using Ohm's law, a low resistance of the shunt resistor results in a low observable voltage. Because of the damping due to the high resistance of the overall circuit, the voltages will drop close to the minimum sensitivity of the oscilloscope. There are conflicting goals. A high resistance would result in damping, which damps the oscillation close to zero within a very short time. A low resistance of the shunt resistor does not eliminate the resistance, which is present as a parasitic element of real-world capacitors or inductors. If the resistance is very low, it prohibits a good measurement in the first place. Furthermore, there is the 8 bit resolution of the oscilloscope[1]. The 8 bits represent the whole range of observable voltages. We will observe that the high speed of the oscilloscope allows for a surprisingly differentiable signal despite the 8 bit, because the noise works like dithering. Smoothing the signal will result in more than 8 useable bits. We will observe the possible quantization artifacts, if we remove the noise.

We recognize that recording the current without impacting the system too much and without more equipment to sense the current directly is considerably more challenging than measuring the voltage. Therefore, it is a good example to test the algorithm, which recovers unobserved state variables. In this setting, it can be beneficial, because observing good data is challenging. On the other hand, we can record data of sufficient quality to check the result of the recovered state variable. If we do not use the full 8 bit, the quantization and the noise show, and we see the distortion and errors in our state variable. This observed and bad data illustrates the improvement we achieve by recovering the current state variable using our algorithm.

When recording the voltage of the shunt resistor using the full 8 bit, we observe a relatively precise picture of the current. This relatively precise picture illustrates the quality of our recovered state variable when comparing both state variables.

4.1 Estimate the Parameters

Our implementation of Gradient Matching assumes a dynamical system of n ordinary differential equations[2]. These n ODEs have a polynomial right-hand side, and the left-hand side consists of the derivative of the data.

Examples for models of this format are Lotka-Volterra predator-prey systems, chemical

---

[1] There are high resolution modes, but they limit the speed, the samplerate of the scope.
[2] When exceeding more than three equations, noise is becoming an increasing problem.

reactions, spring/mass systems, or when rearranging the parameters in our case an LCR circuit. Compared to the previously mentioned systems, an electric voltage is extremely easy to measure by using oscilloscopes.

In this section, we will build an electric circuit, we will measure its oscillations by measuring the voltages, smooth the data, reduce the number of samples, and we run the Gradient Matching algorithm on the acquired voltage and current data, which is calculated from the voltage of the shunt resistor. The Gradient Matching algorithm will estimate the parameters.

4.1.1 Experimental Setup

This experiment consists of an LCR oscillator and a trigger source, creating a Dirac-like impulse. The LCR oscillator consists of an $18\,\text{nF}$ ceramic capacity[1], an added resistor of $0.6\,\Omega$, an inductivity of unknown resistance, and an impedance between 5 and $600\,\mu\text{H}$. The negative Dirac impulse triggers the oscillator.

The schematic fig. 4.1 shows the circuit and the connections 'Gnd', A ,and B, which we use to measure the voltage of the free oscillation. 'Gnd' is the reference level for the measurement. We use connection A to measure the voltage of the shunt resistor $R_{1.2}$ to calculate the current. We use connection B to measure the voltage of the free oscillation. The voltage level is measured by a Rohde und Schwarz RTO2044 4Ghz oscilloscope (see the screenshot of the measurement in fig. 4.1).
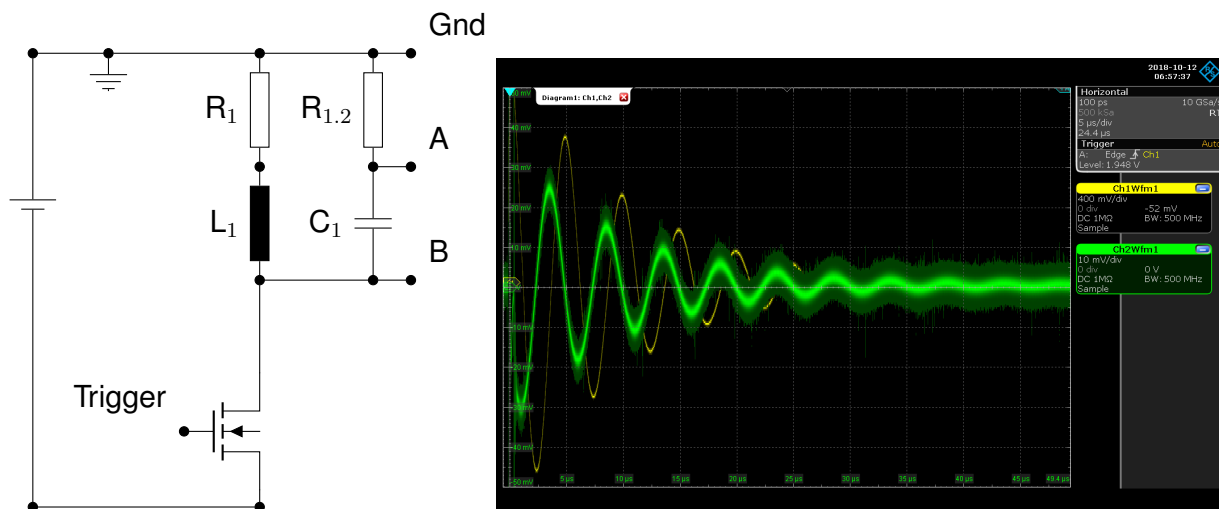
Figure 4.1: Circuit diagram and screenshot of the oscilloscope of the parallel LCR oscillator.

---

[1]This capacity itself consists of two 103M ceramic capacities, which means $2\cdot 10\,\text{nF}$ at 20% tolerance. However, our LCR meter Voltcraft LCR-300 measured it at $18\,\text{nF}$.

4.1.2 Model

The following set of equations models the LCR oscillator:

$$L\frac{dI(t)}{dt} = U(t) + R \cdot I(t),$$
$$C\frac{dU(t)}{dt} = I(t).$$

We reformulate the equation, because the values for L and C are on the left-hand side, which does not follow the required pattern of the algorithm. The algorithm does not expect parameters on the left-hand side. It is a result of the assumption that the left-hand side is the derivative of the data and the right-hand side a polynomial. We use:

$$\frac{dI(t)}{dt} = \frac{1}{L} \cdot U(t) + \frac{R}{L} \cdot I(t),$$
$$\frac{dU(t)}{dt} = \frac{1}{C} \cdot I(t).$$

Furthermore, the resistance R in the equation will cover all resistive influences. That will be a sum of the known resistor $R_{1.2}$ and losses of the wires, resistor, capacity, and additional high-frequency losses. The circuit diagram fig. 4.1 summarizes these additional losses by R_1. We use the resistor $R_{1.2}$ to calculate a current value from a measured voltage.

4.1.3 Measurements and Calculation

Because the realization of every Dirac impulse introduces unwanted nonlinearities, we start measuring a single impulse (see fig. 4.2).
We use the Gradient Matching algorithm to calculate the capacity, inductivity, and resistance values. These values become visible after rearranging the terms.

$$\dot{U} = 0.061\frac{1}{\text{nF}} \cdot I \qquad \Rightarrow \qquad 16.4\,\text{nF} \cdot \dot{U} = I,$$
$$\dot{I} = -26.1\frac{1}{\text{mH}} \cdot U - 284\frac{\Omega}{\text{mH}} \cdot I \qquad \Rightarrow \qquad 38.3\,\mu\text{H} \cdot \dot{I} = -U - 10.9\,\Omega \cdot I.$$

We set the factors of I and U to one by dividing the equations, the values for $L = 38.3\,\mu\text{H}$, $C = 16.4\,\text{nF}$ and $R = 10.9\,\Omega$ show up. The value of the capacity is within 15% of the capacity used. We measured that capacity at $18\,\text{nF}$, and it is labeled as $2 \cdot 103\text{M} = 2 \cdot 10\,\text{nF}$ at 20% tolerance.
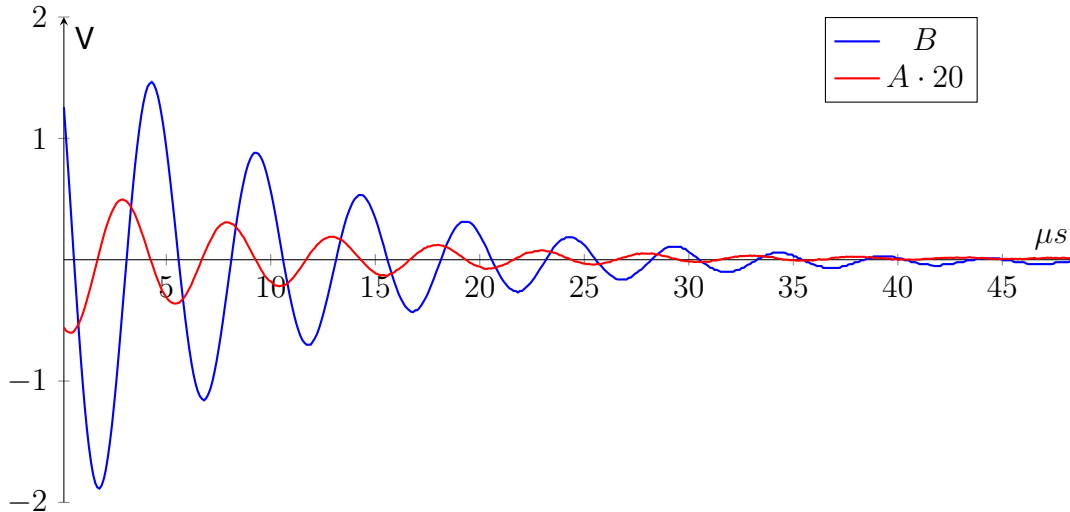
Figure 4.2: The raw data measured at the connectors A and B. The voltage measured at A is tiny because of the low resistance and therefore scaled by a factor of 20. Figure 4.1 shows a screenshot of the scope recording the data.

4.1.4 Discussion

We consider the values to be rather precise. The value of the resistor is acceptable considering the unknown resistance of the capacity, the solder joints, the coil, and the losses induced by employing a magnetic core of unknown material properties.

We can check the validity of the calculated values, especially of the inductor, by comparing the theoretical frequency of the oscillator and the measured frequency. Calculating the frequency results in $f = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{16.4\,\text{nF}\cdot38.3\,\text{µH}}} = 201\,\text{kHz}$. A quick check on the data shows a period of just below $5\,\text{µs}$, which would result in a frequency of a little over $200\,\text{kHz}$. This result shows no extraordinary performance, but it indicates that this ansatz can result in reasonably accurate results. The next two experiments will show the application of this algorithm in the presence of Dirac impulses.

4.2 Robustness – Dirac Impulse Noise

The previous experiment demonstrated that the concept can extract the model and its parameters from data. However, especially when the form of the model equation is actually known, other established methods are available, which are less sensitive to noise as they do not depend on the quality of the derivative. They are usually based on integrating the differential equation. This integration requires initial conditions. When measuring a single impulse and when starting the analysis after the Dirac impulse, the initial values can be estimated from the first samples of the signal. Integral methods can be applied without knowing the initial values beforehand.

In opposite to those methods, the Gradient Matching ansatz is a much more local ansatz. There are no initial values. The present experiment shows that it is not even necessary to stay on one trajectory to calculate the parameters[1]. That means we can record data spanning multiple events without caring about the local perturbations by Dirac impulses and changed trajectories for every new oscillation. This robustness is an advantage compared to classical methods.

4.2.1 Experimental Setup

The setup is identical to the one of the previous experiment. This time an RTO1024 was used to record the data, and environmental parameters changed the inductivity.
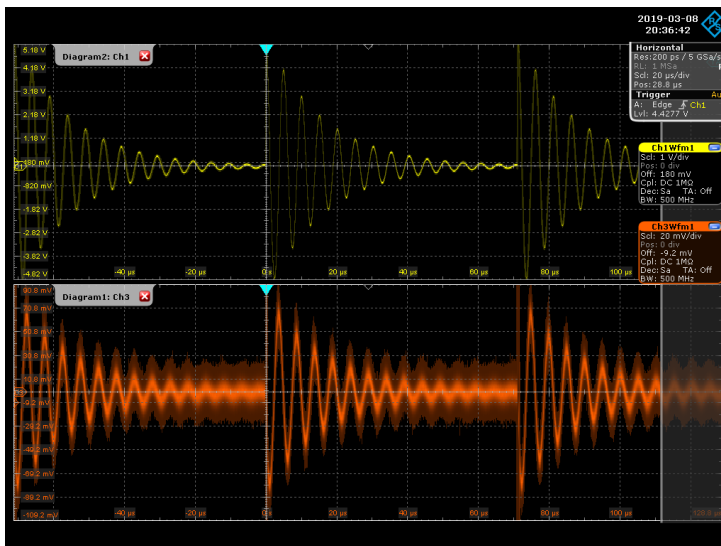
Figure 4.3: Screenshot of the oscilloscope recording the parallel LCR oscillator.

---

[1]This independence of the initial values is virtually possible with the classical approaches, too, if they detect the restart and estimate the initial values. However, in the case of Gradient Matching, the algorithm does not have to detect the impulses to deal with the effects. As this algorithm is very local, the effect stays local, too.

4.2.2 Measurements and Calculation

This experiment includes Dirac-like impulses to trigger the free oscillation, which introduces noise spanning more than a single sample in practice. This impulse is not a part of the model, so the resulting noise is observable as outliers. As in the previous experiment, there are Gaussian noise and quantization artifacts introduced by the 8 bit resolution of the oscilloscope, which influences the derivative, but the high throughput of the oscilloscope compensates these challenges. The noise still allows a useful measurement.

The Dirac-impulses are not as insignificant as the theory could be. The impulses have to transfer enough energy to start the oscillation and are not of infinitesimal short length. Furthermore, the temporal derivative of Gradient Matching has to process the impulses. In the case of evaluating data from theoretical models, we can use the model equations to generate a temporal derivative per data sample. In the practice of Gradient Matching[1], we use a numerical derivative, the five-point stencil, which needs a relatively smooth function. The recorded data is not smooth because there is an 8 bit resolution. Obtaining a smooth function is not a problem due to the high throughput data. The data is recorded at a sample rate of $5\,\mathrm{GS\,s^{-1}}$, and the frequency of our signals apart from the Dirac impulses is below $0.5\,\mathrm{MHz}$. We choose to use a running mean filter[2] of a sample length up to $4001\,\mathrm{S}$ and to reduce the amount of data by only using every 25th sample for all the calculations after employing the running mean. The high throughput, the Gaussian noise, and the running mean result in high resolution and high throughput data. Figure 4.4 shows the resulting data.

We use penalized regression to perform the parameter estimation of the Gradient Matching, because the polynomial right-hand sides of the ODEs include collinearity. Penalized regression introduces a penalty value τ into the regression. Every new parameter, hence every new monomial, reduces the residual value of the regression, but a penalty value is added to the residual. Thus, the additional monomial is only accepted if it is more beneficial than the penalty.

When analyzing this data spanning a time interval that includes three Dirac impulses, we initialized the penalty value τ of the penalized regression to calculate three sets of parameters representing an ODE. Each penalty value results in a different set of parameters.

The free oscillation of an LCR-circuit can be described by the following formula, which we

---

[1]The program flow of the algorithm is described in section 6.2.1

[2]We tested the effect of a Gaussian kernel, but the runtime of the tested packages did not justify the neglectable benefits due to the high throughput data available.
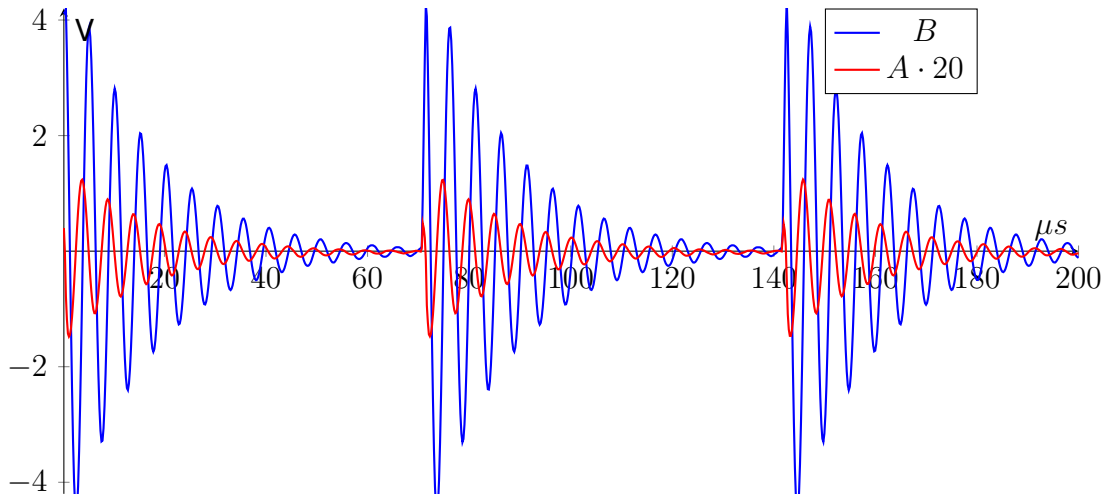
Figure 4.4: The smoothed recorded data, which we used to run the algorithm. The data was recorded at fig. 4.3.

already covered in the previous experiment:

$$\frac{dI(t)}{dt} = \frac{1}{L} \cdot U(t) + \frac{R}{L} \cdot I(t) \qquad \Rightarrow \qquad L\frac{dI(t)}{dt} = U(t) + R \cdot I(t),$$

$$\frac{dU(t)}{dt} = \frac{1}{C} \cdot I(t) \qquad \Rightarrow \qquad C\frac{dU(t)}{dt} = I(t).$$

We compare the three results of the penalties $\tau = 0.015$, $\tau = 0.018$, and $\tau = 0.20$ with this 'true' formula. All penalties resulted in relatively similar and realistic values for the capacitor and inductor, and therefore in similar frequencies of the solutions corresponding to the parameters, which are shown in fig. 4.5.

$$\tau = 0.015,$$

$$\dot{U} = 0.183\frac{1}{\mu F\,\Omega} \cdot U + 0.0631\frac{1}{nF} \cdot I \qquad \Rightarrow \qquad \dot{U} \cdot 15.8\,\text{nF} = 2.9\frac{1}{k\Omega} \cdot U + I,$$

$$\dot{I} = -28.8\frac{1}{\mu H} \cdot U - 200.32\frac{\Omega}{\mu H} \cdot I \qquad \Rightarrow \qquad \dot{I} \cdot 34.7\,\mu H = -U - 6.95\,\Omega \cdot I.$$

The low penalty of $\tau = 0.015$ introduced an additional parameter. Although the estimated value corresponding to the resistance is close to the resistance estimated for $\tau = 0.018$, the damping of the resulting system of ODEs does not match. Figure 4.5 shows the solution using estimated initial values.

$$\tau = 0.018,$$

$$\dot{U} = 0.0618\frac{1}{\text{nF}} \cdot I \qquad\qquad \Rightarrow \qquad\qquad \dot{U} \cdot 16.2\,\text{nF} = I,$$

$$\dot{I} = -28.8\frac{1}{\mu\text{H}} \cdot U - 200.32\frac{\Omega}{\text{mH}} \cdot I \qquad \Rightarrow \qquad \dot{I} \cdot 34.7\,\mu\text{H} = -U - 6.95\,\Omega \cdot I.$$

The resulting system of ODEs corresponding to $\tau = 0.018$ matches the system of the LCR-circuit. Figure 4.5 shows a solution, which experiences similar damping compared to the measured waveform.

$$\tau = 0.020,$$

$$\dot{U} = 0.061868\frac{1}{\text{nF}} \cdot I \qquad\qquad \Rightarrow \qquad\qquad \dot{U} \cdot 16.16\,\text{nF} = I,$$

$$\dot{I} = -28.174\frac{1}{\mu\text{H}} \cdot U \qquad\qquad \Rightarrow \qquad\qquad \dot{I} \cdot 35.5\,\mu\text{H} = -U.$$

The high penalty of $\tau = 0.020$ results in a smaller set of parameters. Gradient Matching estimated the inductor and capacity and, therefore, the correct frequency. However, the penalty suppressed the monomial corresponding to the resistor and did not include damping into the model. Figure 4.5 shows the solution of this ODE, an oscillation without damping.

4.2.3 Discussion

We observe: The small penalty results in an additional parameter. The medium one recovers the correct number of parameters, and the largest penalty value results in a dynamical system, which misses the resistance value and, therefore, the damping. Independent of the correct right-hand side of the differential equations, i.e., the correct number of parameters, the experiment demonstrated how a range of penalty values resulted in the correct capacity and inductivity values. Combined, these values define the frequency of the oscillation. Locally they have a strong effect on the shape of the signal and a major effect on the derivative, impacting the regression. Even the resistance was similar, when available. On the other hand, the damping of the oscillation seems to be sensitive to noise and the value of the penalty value τ. Figure 4.5 shows the three dynamical systems resulting from the three penalty values.

This effect seems comprehensible. On the one hand, there are global effects, like the damping of the oscillator. These global effects are easy to identify when looking at the
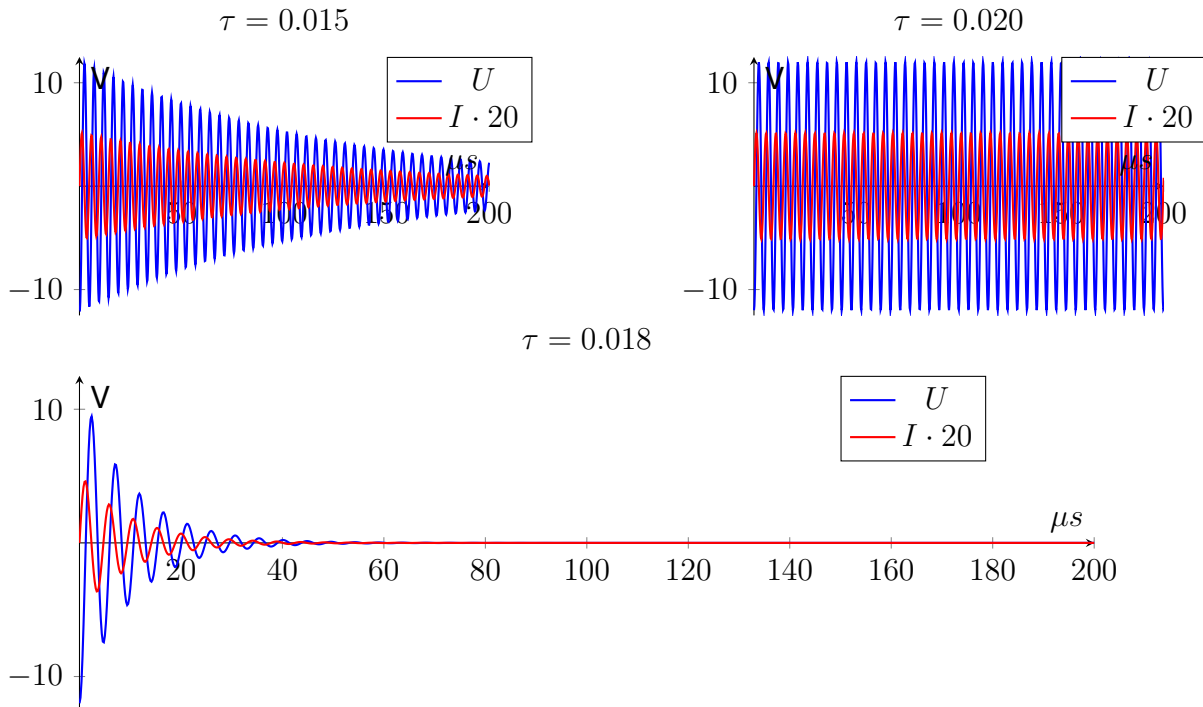
Figure 4.5: Plot of the dynamical systems built from the parameters calculated by using the parameters $\tau = 0.015$, $\tau = 0.020$ and $\tau = 0.018$ and using the same initial values.

data globally. On the other hand, there is the algorithm evaluating the derivatives of the waveform per sample. That is a very local property. Locally, damping only has a very small effect. If this effect is perturbed by noise and other influences not covered by the model, the damping is an effect, which is hard to detect.

The Dirac impulses introduce some of the noise. They span over a noticeable time. They are neither symmetric like Gaussian noise nor covered by the model of the oscillator. Of course, it is possible to preprocess the data and remove the impulses. In this case, writing a simple preprocessor deleting the start impulses would be an easy task. On the other hand, we expect classic methods to have very little problems to detect the damping, and the classical methods can be extended to detect the impulses. These impulses mean a jump into a different trajectory of the dynamical system.

Preprocessing of this kind is not performed to show the advantage of Gradient Matching. Gradient Matching does not have to deal with the jumps to different trajectories because it does not matter to the ansatz and only introduces a little noise, which can be suppressed by the noise reduction within the algorithm. This reduction can consist of smoothing or neglecting derivatives exceeding a certain value.

4.3 Gradient Matching – Recover Latent Function Experiment

This time we use the algorithm to recover unobserved data[1]. In the case of an LCR oscillator, it is often easier to observe voltage than current, especially if the current is observed by measuring the voltage over a tiny resistor, the shunt. In this section, we will use two recorded datasets. One will use most of the available 8 bit range of the oscilloscope to record the current indicating voltage. However, it will still be noisy data due to the small resistor, resulting in small voltages close to the noise-floor. The other dataset will only use a fraction of the available bit range to record the data. That will result in high quantization effects, especially when the recorded voltage becomes small during the dampened oscillation. We will take the dataset, which shows the quantization effects first. It will neglect the poorly recorded current and consider the current as an unobserved state variable. Instead, we will use the algorithm to recover a state variable from the voltage data and the known parameter of the dynamical system, the LCR-oscillator.

After comparing the poor data and the recovered data to show the improvement, we will use the better dataset. The better dataset already provides an acceptable current indicating voltage for comparison. We use the acceptable data to demonstrate that the data did not only improve visually, but the correct shape of the data can be recovered. This time we neglect the good data and expect the recovered data to resemble a similar shape.

4.3.1 Experimental Setup

The setup will be identical to the previous section using an RTO1024.

4.3.2 Measurements and Calculation

Low Resolution

The oscilloscope records the data at 8 bit resolution. We measure the current by using a shunt resistor. The voltage resulting from the current is relatively small, and it is not scaled to cover the maximal range. Because it does not cover the maximal range, it is recorded using just a fraction of the 8 bits. Figure 4.6 shows the whole function, which looks relatively smooth, especially after smoothing. We apply a running mean filter in a preprocessing

---

[1]Section 6.2.2 shows the program flow of the algorithm. Section 3.4 introduces the functional, which is optimized to recover an unobserved state variable.

step. The running mean applied to the noise interpolates between the 8 bit quantization levels and enables smooth gradients.
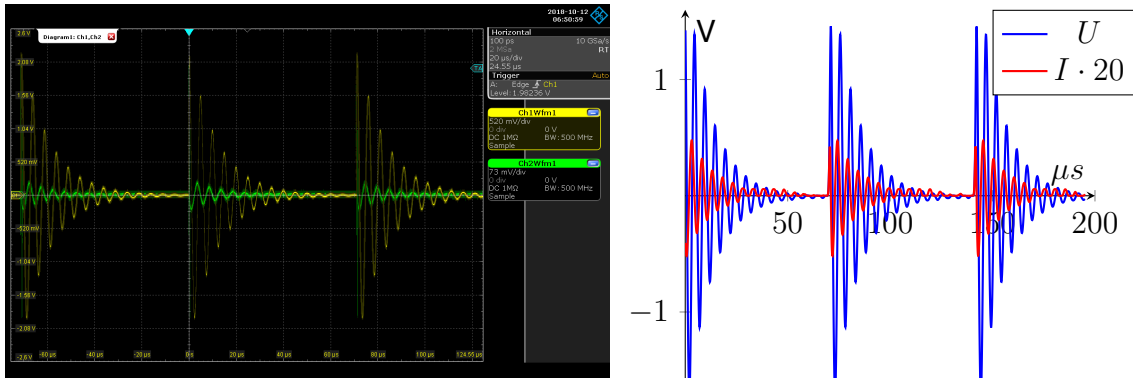
Figure 4.6: Screenshot of the oscilloscope of the parallel LCR oscillator. At this scale, the recorded data looks good after smoothing. Figure 4.7 zooms the tail of the oscillation and shows the raw observation at a larger scale. Figure 4.9 shows the smoothed values of the current on a larger scale compared to the reconstructed values.

The quantization due to the low resolution becomes very visible when the amplitude converges to zero. That becomes especially apparent when zooming into the raw data, as shown in fig. 4.7.
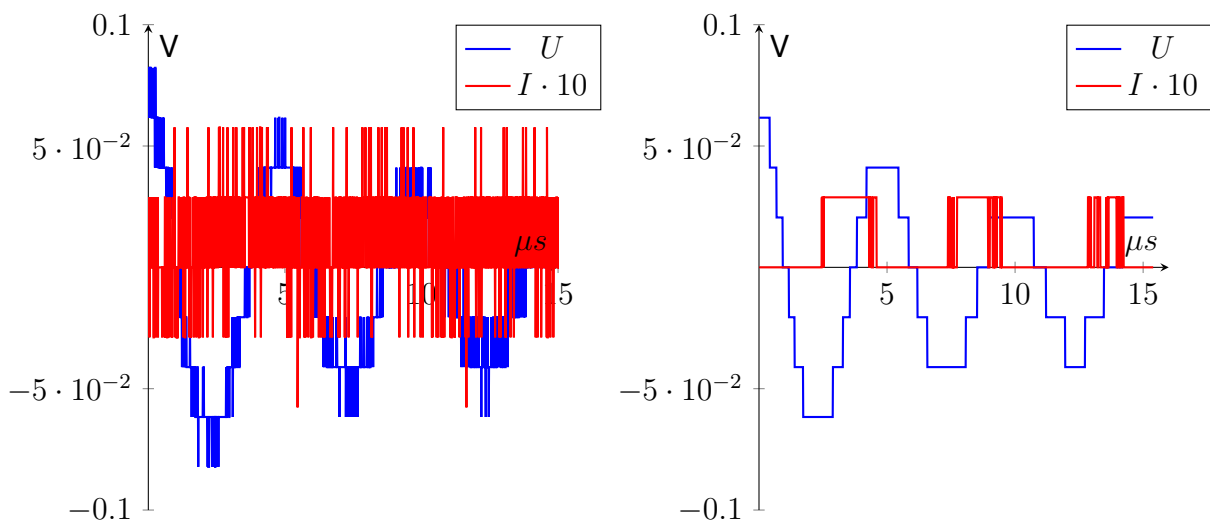
Figure 4.7: When zooming into the oscillation of the originally recorded data while converging to zero, massive quantization artifacts become visible. Median filtering removes the noise, but erases a lot of information left in the noise. Nevertheless, it is evident that even filtering will not repair the positive offset in the recorded current values.

Using a median filter shows that most of the dynamic of the current is represented by one bit. I.e., after using the median filter, only two voltage level describing the current are left. Two bits cover almost the whole dynamic of the unfiltered data except for rare

exceptions. Smoothing can dampen the quantization artifacts, but there seems to be an offset. Smoothing does not remove the offset. Strong smoothing would dampen the signal, too, but it would be necessary to generate a smooth function[1]. On the other hand, there are almost 3 bits describing the dynamic of the voltage.

The voltage can be smoothed and used to find the correct current using the known differential equation and nonlinear optimization.
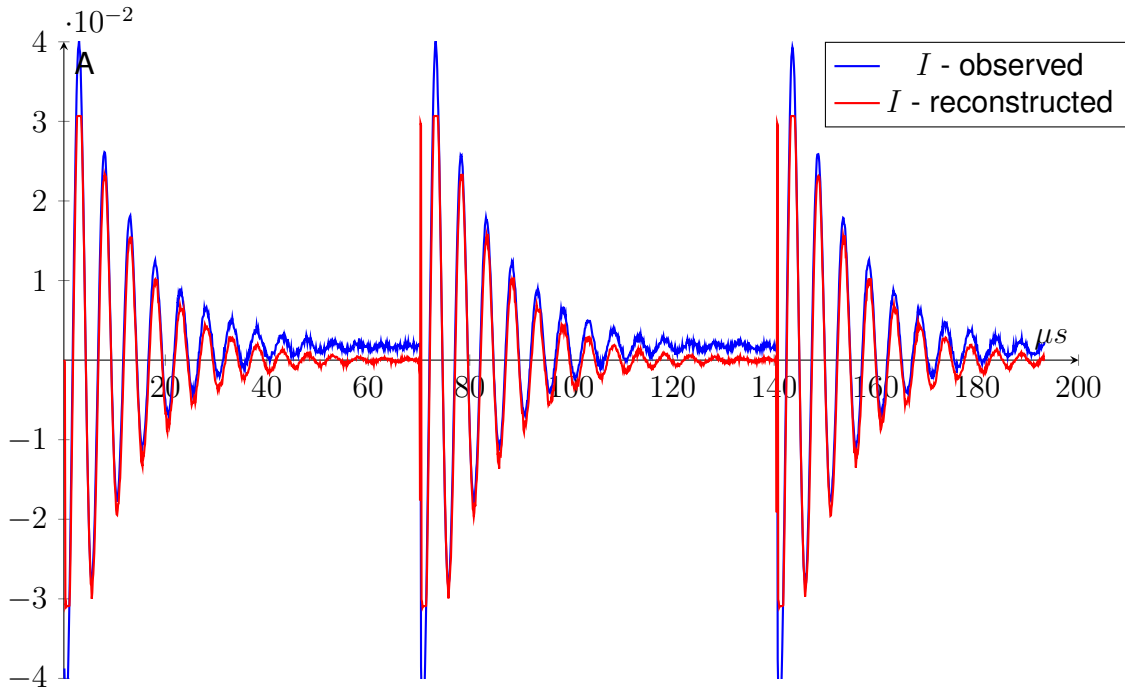
Figure 4.8: The current calculated from the observed low-resolution voltage of the shunt resistor compared to the calculated, reconstructed voltage.

Figure 4.8 shows the result of calculating the current of the shunt resistor compared to the poorly observed data filtered by a running mean.

Zooming into the first tail of the oscillation (fig. 4.9) emphasizes the artifacts from the low-resolution current data compared to the current values reconstructed from the voltage data. Offset errors, noise, and distortion become more visible at low current levels. The reconstructed signal does not show the strong offset error of the current signal calculated from the observed voltage of the shunt resistor. Furthermore, the reconstructed signal suffers from less noise. The last period of the oscillation becomes much more visible in the reconstructed signal; while almost invisible within the current data calculated from the observed voltage of the shunt resistor.

The next part will describe the other dataset, the good data, which uses most of the 8bit range to record the voltage of the shunt resistor. Repeating the experiment and recording

---

[1]An alternative could be filters optimized to remove quantization artifacts.

the current sensing voltage of the shunt resistor at a higher resolution allows us to compare the quality of the recovered signal to an observed signal of better resolution.
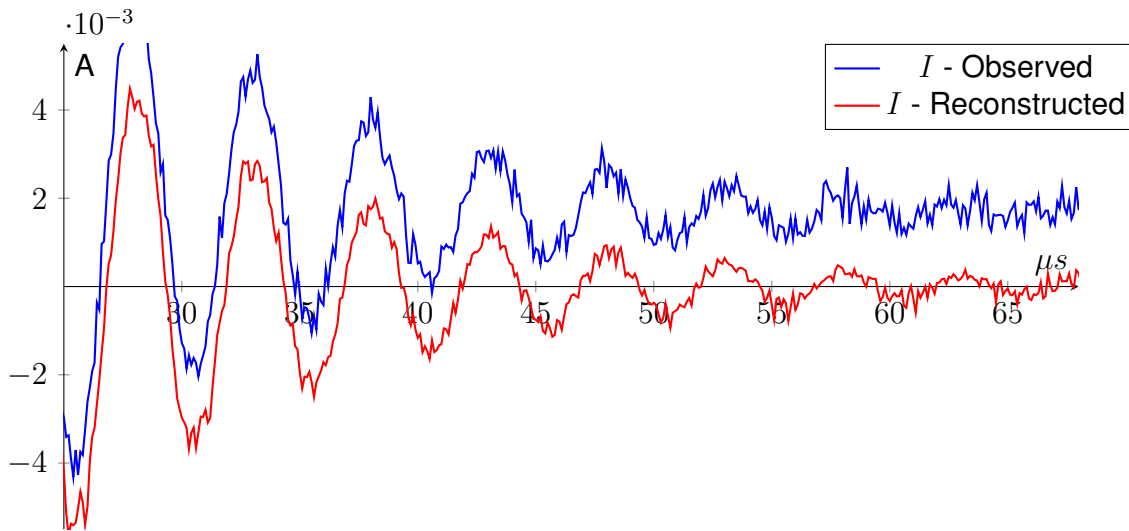
Figure 4.9: The current from the observed and smoothed low-resolution voltage of the shunt resistor compared to the reconstructed current. We zoomed the current values to show the differences, which were less prominent in the full data fig. 4.8.

Higher Resolution
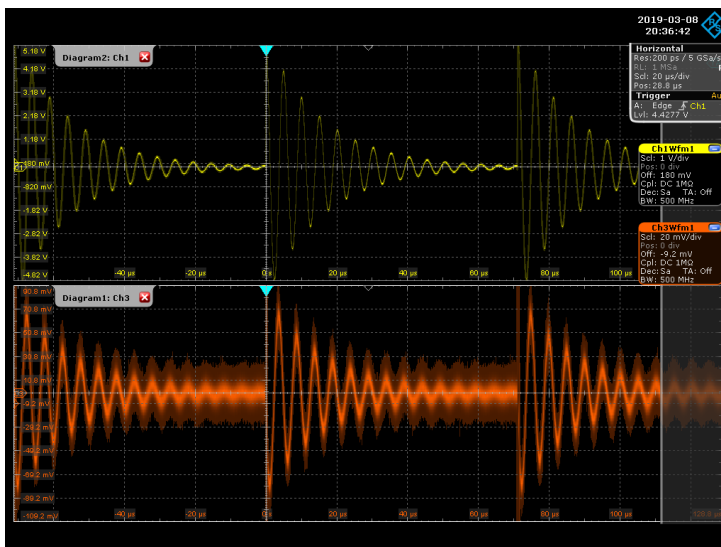
Figure 4.10: Screenshot of the oscilloscope of the parallel LCR oscillator. The first channel records the voltage in yellow. The second channel records the voltage of the shunt resistor in orange.

This time, the voltage measurements of the shunt resistor span most of the 8 bit range.
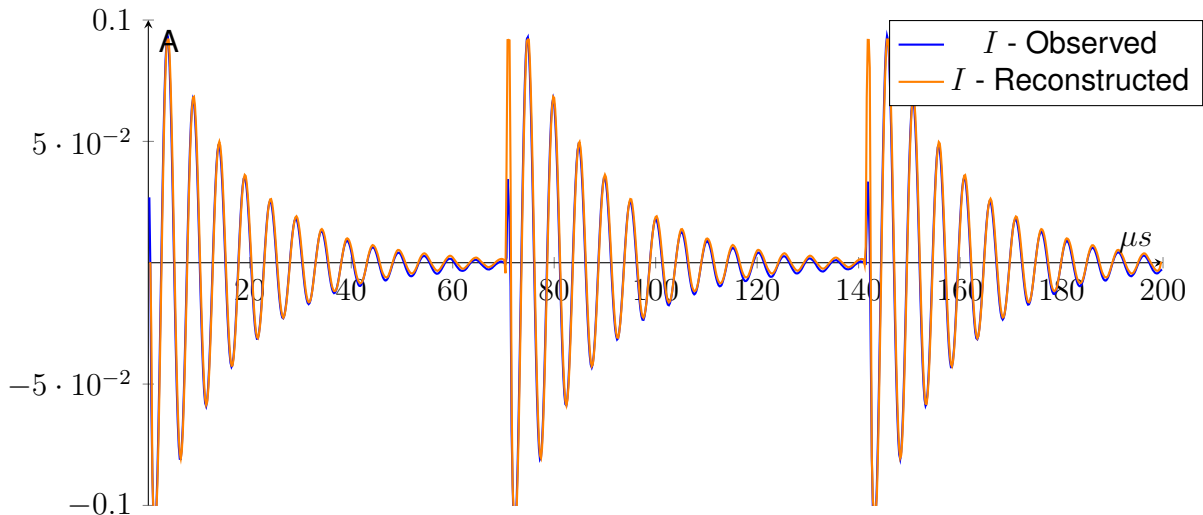
Figure 4.11: The current calculated from the observed voltage of the shunt resistor compared to the reconstructed current.

Figure 4.10 shows a screenshot of the oscilloscope recording the data. The first channel records the voltage, the second channel the voltage of the shunt resistor. This time, the voltage of the shunt resistor uses the whole 8 bit. Nevertheless, the higher noise level is visible compared to the first channel. Figure 4.11 shows the current calculated from the observed shunt resistor voltage compared to the recovered current by the algorithm based on Gradient Matching. When zooming into the function (fig. 4.12), it is visible that the reconstruction resembles the observed shape of the waveform, but its mean value is closer to zero. It suffers less from an offset error.
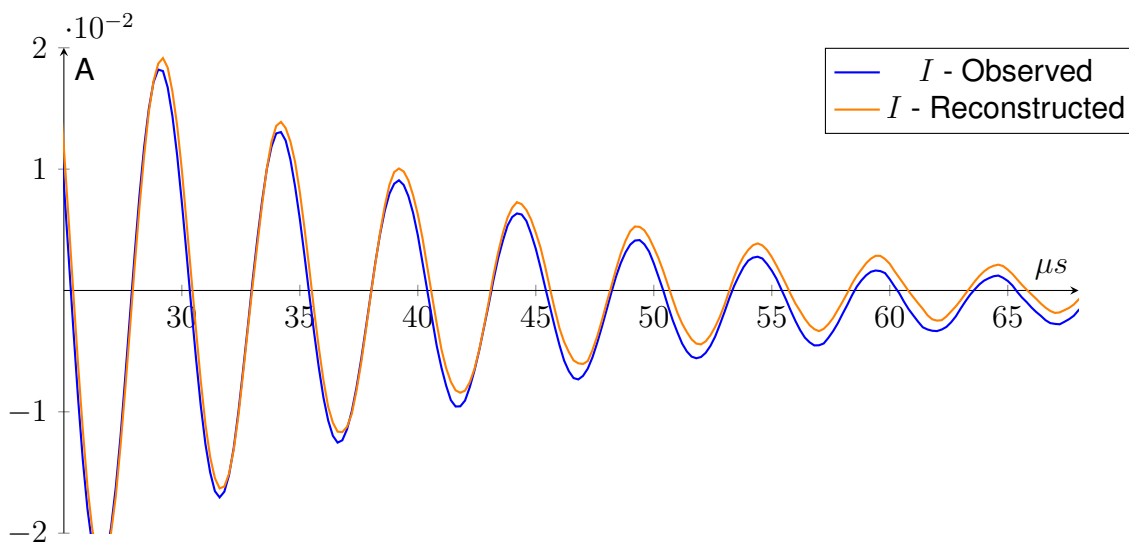
Figure 4.12: The current calculated from the observed voltage of the shunt resistor compared to the reconstructed current. We zoomed the current values to show the differences.

Result of the Algorithm

We observe that the algorithm calculated an useable waveform describing the current despite the bad data quality. Figure 4.9 visualizes this improvement of the recovered values compared to the observed waveform. In the case of available 8-bit data of the current, the calculated waveform was remarkably close to the observed waveform but lacked an offset error. Figure 4.12 shows the comparison in the case of good data.

4.3.3 Discussion

All in all, we find that it is possible to recover unobserved components in case of bad data. In the case of acceptable data, the waveform reconstructs the expected shape of the observed data. This indicates that we recover the 'correct' solution. Furthermore, the algorithm can recover improved data even in cases of available data of acceptable quality, which was recorded at non-optimal circumstances like measuring a current using a small shunt resistor without having a high voltage resolution. Figure 4.12 shows that despite the smooth values of the recorded waveform, the algorithm was able to recover a component lacking the visible offset error. Although the observed data is of acceptable quality, the quality of the reconstruction is slightly better while preserving the same shape.

This recovery is possible without having to estimate initial values or having to detect the Dirac impulses, which act as Poisson noise and alter the trajectory of the dynamical system.

5 Appendix

The appendix provides additional information on auxiliary calculations and additional graphs.

5.1 Contact Tracing

The appendix covering the contact tracing collects several additional calculations and graphs. It starts calculating theoretical equilibria in the case of no tracing, one-step backward tracing, and forward tracing using the corresponding approximations. We calculate those equilibria using the approximation of the effective reproduction number R_{eff}. In the case of forward tracing, we use the same forward tracing approximation within the endemic model as we use within R_{eff}, which leads to a very good fit between the model and the approximation in that case.

We show the accuracy of the theoretical limits using the example of forward tracing in section 5.1.2.

When evaluating the recursion formula of the forward tracing and using a fixed delay numerically, it can be further simplified. We show this calculation in section 5.1.3.

In the end, section 5.1.4 shows why using $t = 5$ is a viable time to sample an age distribution, although it is at the end of the onset case.

5.1.1 Calculating Theoretical Equlibria

The equilibrium of the epidemic is characterized by a reproduction number $R_{\text{eff}} = 1$, i.e., on average, every individual infects another individual. Higher or lower reproduction numbers would result in raising or falling numbers of infected individuals, which contradicts the equilibrium.

Thus, we can take the calculations and approximations of the R values without tracing, with fixed delay backward tracing, fixed delay forward tracing, and fixed delay full tracing from the onset case. We adapt the contact rate of susceptible individuals for the endemic case. We use the adapted reproduction numbers and use them in an equation, which we can solve for the number of infected individuals. Note that the integration of κ accumulates the approximation errors, and we do not respect cluster contacts. Without tracing, the reproduction number is exact. In the case of forward tracing, we achieve good results because we use the forward tracing approximation and therefore use a relatively low p. The low p results in smaller approximation errors, and there are no cluster contacts to respect.

Lemma 82 (Equilibrium Without Tracing) *The number of susceptible individuals in the equilibrium case without tracing is defined by* $S = \frac{N(\alpha+\sigma)}{\beta}$.

Proof: No contact tracing results in a probability of being infective of $\hat{\kappa}(a) = e^{-(\sigma+\alpha)a}$. We take the reproduction number without tracing from corollary 6, $R = \int_0^\infty \beta \hat{\kappa}(a) da$. We adapt the contact rate to the endemic case. Thus, there is only a fraction $\frac{S}{N}$ of susceptible individuals to infect, which results in $R = \int_0^\infty \beta \frac{S}{N} \hat{\kappa}(a) da$. We define $1 \overset{!}{=} R$ and solve for the number of susceptible individuals S.

$$1 \overset{!}{=} R = \int_0^\infty \beta \frac{S}{N} \hat{\kappa}(a) da = \beta \frac{S}{N} \left[\frac{-e^{-(\alpha+\sigma)a}}{\sigma+\alpha} \right]_0^\infty da = \frac{\beta \cdot S}{N(\alpha+\sigma)}$$

$$\Rightarrow S = \frac{N(\alpha+\sigma)}{\beta}$$

$\square$

Using the population size of $4 \cdot 10^5$ and $\beta = 2$, which was often used during the thesis, we get the following number of infected individuals in the equilibrium case.

$$S = \frac{N(\alpha+\sigma)}{\beta} = \frac{4 \cdot 10^5}{2} = 2 \cdot 10^5.$$

Lemma 83 (Equilibrium with Fixed Delay Forward Tracing) *The number of susceptible individuals in the equilibrium case with forward tracing is approximated by*

$$S = \frac{N}{\beta \left(\frac{1}{\alpha+\sigma} - p\, p_{obs} \frac{\hat{\kappa}(T)}{2(\alpha+\sigma)} \right)} + \mathcal{O}(p^2).$$

Proof: We recall the reproduction number of fixed delay forward tracing proposition 29 and add the adaption of the contact rate to the endemic case by multiplying the contact rate β by $\frac{S}{N}$.

$$1 \stackrel{!}{=} R = \beta \frac{S}{N} \left(\frac{1}{\alpha+\sigma} - p\, p_{\text{obs}} \frac{\hat{\kappa}(T)}{2(\alpha+\sigma)} \right) + \mathcal{O}(p^2)$$

$$\Rightarrow S = \frac{N}{\beta \left(\frac{1}{\alpha+\sigma} - p\, p_{\text{obs}} \frac{\hat{\kappa}(T)}{2(\alpha+\sigma)} \right)} + \mathcal{O}(p^2).$$

$\square$

To calculate the number of infected individuals, we use $I = N - S$. The formula

$$I = N \left(1 - \frac{1}{\beta \left(\frac{1}{\alpha+\sigma} - p\, p_{\text{obs}} \frac{\hat{\kappa}(T)}{2(\alpha+\sigma)} \right)} \right)$$

evaluates to $I = 1{,}688 \cdot 10^5$ using the population size of $N = 4 \cdot 10^5$, $\beta = 2$, $T = 0$, and $p = 0.3$. Evaluating the formula for $T = 0.5$ results in $I = 1{,}82 \cdot 10^5$. Figure 2.7 shows the equilibrium for $T = 0.5$ and the results of the simulation and endemic model for comparison. Section 5.1.2 focuses on showing some theoretical equilibria of forward tracing.

Lemma 84 (Equilibrium with Fixed Delay Backward Tracing) *If $p \ll 1$, the number of susceptible individuals in the equilibrium case is approximated by*

$$S = \frac{(\alpha+\sigma)N \left(1 - \sqrt{1 - 2\,\kappa(T)\, p\, p_{obs}} \right)}{\beta\, \kappa(T)\, p\, p_{obs}} + \mathcal{O}(p^2).$$

Proof: Recall the reproduction number of fixed delay backward tracing proposition 17

$$R = \frac{\beta}{\alpha+\sigma} - p\beta^2 p_{\text{obs}} \frac{\hat{\kappa}(T)}{2(\alpha+\sigma)^2} + \mathcal{O}(p^2).$$

Add the adaption of β to the endemic case.

$$1 \stackrel{!}{=} R = \frac{\frac{S}{N}\beta}{\alpha + \sigma} - p\frac{S^2}{N^2}\beta^2 p_{\text{obs}}\frac{\hat{\kappa}(T)}{2(\alpha + \sigma)^2} + \mathcal{O}(p^2)$$

$$\Rightarrow S = -\frac{-N^2(\alpha + \sigma)^2 \left(-\frac{\beta}{N(\alpha+\sigma)} - \sqrt{\frac{\beta^2}{N^2(\alpha+\sigma)^2} - \frac{2\beta^2\kappa(T)p_{\text{obs}}p}{N^2(\alpha+\sigma)^2}}\right)}{\beta^2\kappa(T)p\ p_{\text{obs}}} + \mathcal{O}(p^2)$$

$$= \frac{N(\alpha + \sigma)\left(\beta + N\sqrt{\frac{\beta^2(1-2\kappa(T)p\ p_{\text{obs}})}{N^2(\alpha+\sigma)^2}}(\alpha + \sigma)\right)}{\beta^2\kappa(T)p\ p_{\text{obs}}} + \mathcal{O}(p^2)$$

$$= \frac{N(\alpha + \sigma)\left(1 + \sqrt{(1 - 2\kappa(T)p\ p_{\text{obs}})}\right)}{\beta\kappa(T)p\ p_{\text{obs}}} + \mathcal{O}(p^2)$$

When evaluating the term for a population of $N = 4 \cdot 10^5$ and the parameters $\alpha = 0.1$, $\sigma = 0.9$, $\beta = 2$, $p = 0.3$, $p_{\text{obs}} = 0.9$ and a probability of still being infective after the delay T of $\kappa(T) = 1$, we calculate a susceptible fraction of the population bigger than the total population. Therefore, we concentrate on the other result of the quadratic equation and note that in general a low value of p is needed. The low value of p not only ensures more accurate approximations of the backward tracing, but is also necessary for a real result of the square root.

$$\vee S = -\frac{-N^2(\alpha + \sigma)^2 \left(-\frac{\beta}{N(\alpha+\sigma)} + \sqrt{\frac{\beta^2}{N^2(\alpha+\sigma)^2} - \frac{2\beta^2\kappa(T)p_{\text{obs}}p}{N^2(\alpha+\sigma)^2}}\right)}{\beta^2\kappa(T)p\ p_{\text{obs}}} + \mathcal{O}(p^2)$$

$$= \frac{N(\alpha + \sigma)\left(\beta - N\sqrt{\frac{\beta^2(1-2\kappa(T)p\ p_{\text{obs}})}{N^2(\alpha+\sigma)^2}}(\alpha + \sigma)\right)}{\beta^2\kappa(T)p\ p_{\text{obs}}} + \mathcal{O}(p^2)$$

$$= \frac{N(\alpha + \sigma)\left(1 - \sqrt{(1 - 2\kappa(T)p\ p_{\text{obs}})}\right)}{\beta\kappa(T)p\ p_{\text{obs}}} + \mathcal{O}(p^2)$$

$\square$

The approximation bases on a power series around p. Lower values of p result in more accurate approximations. If we evaluate the formula using $N = 4 \cdot 10^5$, $\beta = 2$, $T = 0$, and $p = 0.1$ and calculate $I = N - S = 1.9 \cdot 10^5$, we achieve a good fit. Figure 2.7 shows the theoretical equilibrium and a simulation reaching the theoretical equilibrium at $1.802 \cdot 10^5$ individuals for $N = 4 \cdot 10^5$, $\beta = 2$, $T = 0.5$, and $p = 0.3$.

Equilibrium with Fixed Delay Full Tracing   The reproduction number of fixed delay full tracing can be calculated similar to backward tracing and forward tracing using the corresponding reproduction number from proposition 39. Nevertheless, it combines the approxi-

mation errors from both approximations.

5.1.2 Example Equilibria in the Forward Tracing Case

We use the forward tracing approximation to integrate forward tracing into the endemic model. Therefore, forward tracing is a good candidate to show the theoretical limits from section 5.1.1 and how they represent the simulated values.
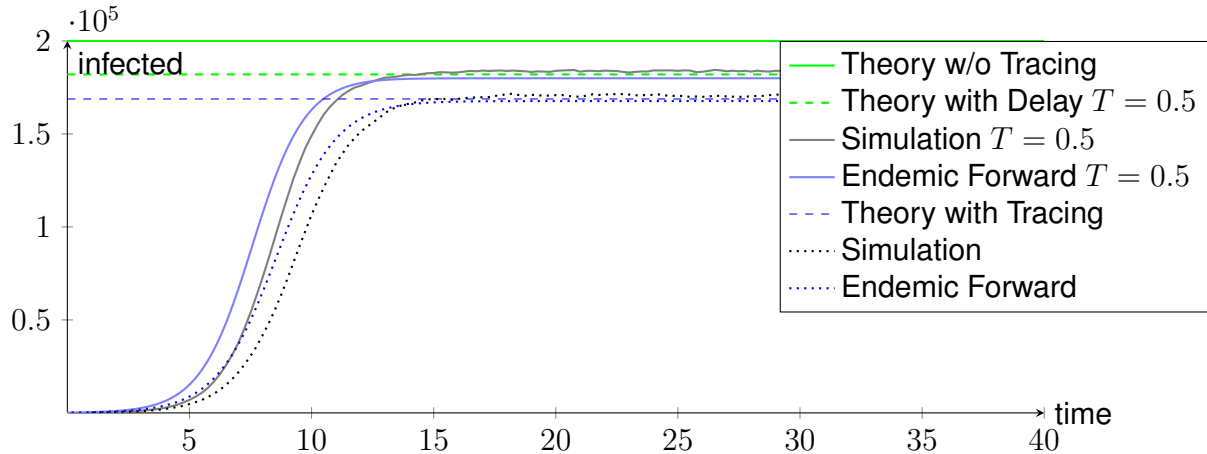
Figure 5.1: The count of infected individuals among a population of $4 \cdot 10^5$ with forward tracing with a delay of $T = 0.5$, a recall probability $p = 0.3$ and a contact rate $\beta = 2$, spontaneous recovery $\alpha = 0.1$ and supervised recovery $\sigma = 0.9$. Black shows the simulation of the spread, blue the partial differential equation and green the theoretical equilibrium.

Figure 5.1 shows forward tracing with and without a tracing delay of $T = 0.5$. We observe a good correspondence between the numeric evaluation of the approximation within the endemic model, the theoretical limit, and the simulation. Overall, there is a slight overestimation of the tracing effect by the approximation.

All calculations base on the probability of being infective $\kappa(a)$. Figure 5.2 shows this probability of being infective. The theoretical probability of being infective from the endemic model, the probability from the onset model, and the probability sampled from the simulation do not show divergence. Furthermore, fig. 5.2 shows an obvious impact of the tracing delay.

There is the green and dashed graph $\hat{\kappa}(a)$ showing the highest probability in case of no tracing. The three graphs for the onset model, endemic model, and simulation are similar to $\hat{\kappa}(a)$ for $a < T = 0.5$. After $a \geq T = 0.5$, they branch away from $\hat{\kappa}(a)$ converge to the three graphs, which originate from tracing without a tracing delay. The graphs without a tracing delay show the smallest probability to be infective after a time a, because of the more effective tracing. The logarithmic plot fig. 5.3 magnifies the differences between the graphs when converging to zero. It shows a shrinking effect of the tracing delay over time.
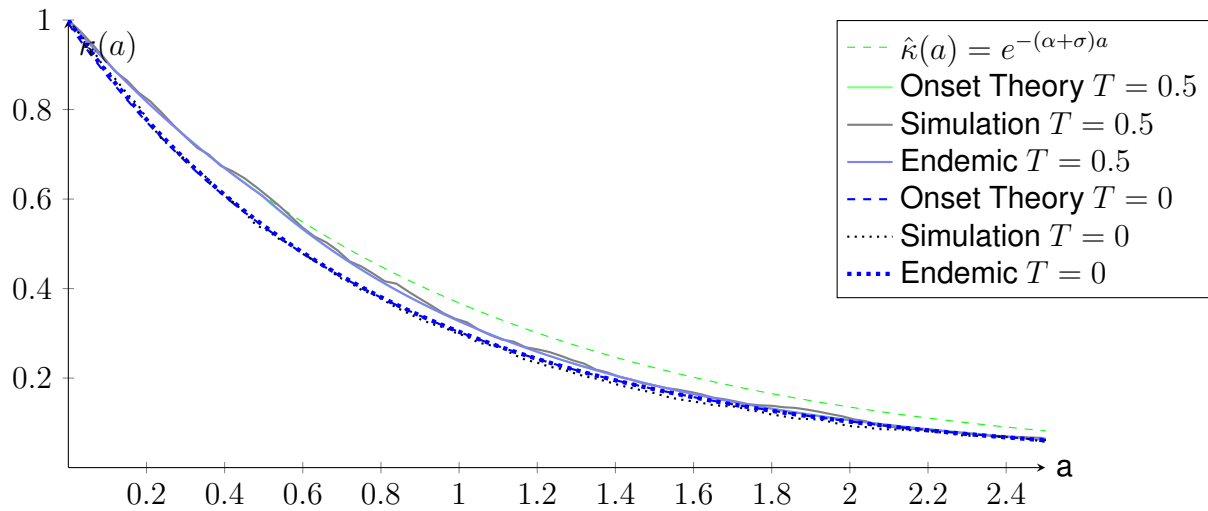
Figure 5.2: The probability of being infective $\kappa(a)$ per individual plotted using a linear scale corresponding to fig. 5.1.

The tracing effect stays prominent, and there is no difference visible between the models apart from the stochastic variation of the simulation.
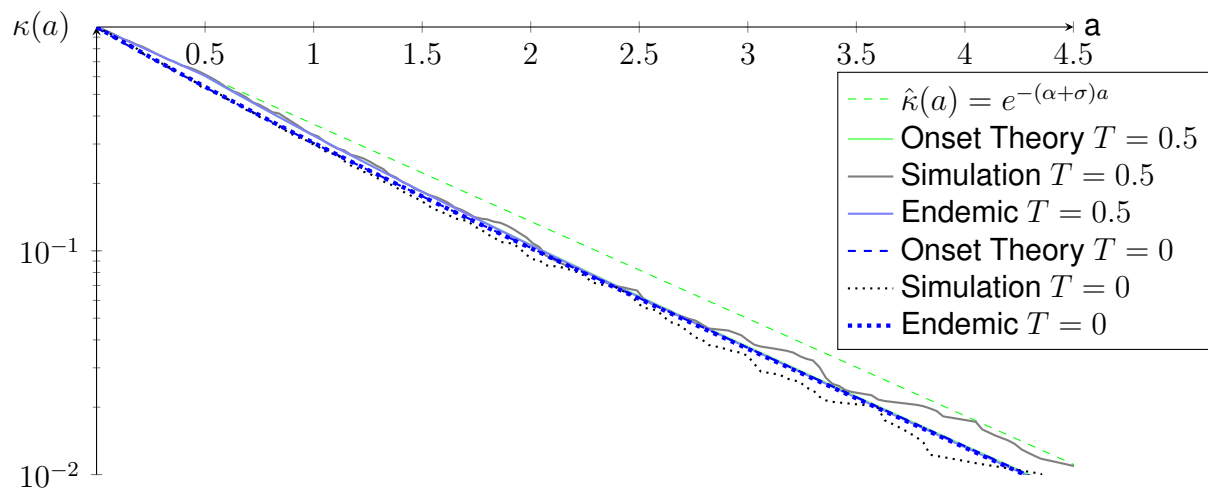
Figure 5.3: The probability of being infective $\kappa(a)$ per individual plotted using a logarithmic scale corresponding to fig. 5.1.

168

5.1.3 Simplifying the Fixed Delay Forward Tracing Recursion

The forward tracing model, which is a part of the full tracing model, employs an outer un-
bounded integral. We can alter the order of the integrals by employing Fubini. Employing
Fubini allows to move the unbounded integral inside, and it can be precalculated. Further-
more, the convolution is transformed into a shift operation using the fixed delay.

Model

Recall the forward tracing from proposition 22 and its application within full tracing from
proposition 33. The replacement of $\kappa_0^+(a) = \hat{\kappa}(a)$ by $\kappa_0^+(a) = \kappa_-(a)$ defines the difference
between forward- and full- tracing, because $\kappa_0^+(a) = \kappa_-(a)$ includes backward tracing

$$\kappa_0'(a) = -\kappa_0(a) \left\{ \alpha + \sigma + p\beta \left[(\phi * (1 - \kappa_0))(a) - \alpha(\phi * \kappa_0^{\#})(a) \right] \right\}, \tag{5.1}$$

$$\kappa_0(0) = 1.$$

In case of $i > 1$ the recursion formula for $\kappa_i(a)$ is given by

$$\kappa_{i-1}(b)\kappa_i(a|b) = \kappa_0(a) \left\{ \kappa_{i-1}(b) - p \int_0^a (-\kappa_{i-1}'(b+c) - \alpha\kappa_{i-1}(b+c)) \int_c^a \phi(a'-c)da'dc \right\} \tag{5.2}$$

$$\kappa_i(a) = \frac{\int_0^\infty \kappa_i^+(a|b)\kappa_{i-1}(b)db}{\int_0^\infty \kappa_{i-1}(\tau)d\tau} \tag{5.3}$$

The three nested integrals are computationally demanding. They result in two nested inte-
grals within an unbounded integral.

Simplification

Lemma 85 (Fixed Delay Simplification of the Forward Tracing Recursion) *When restrict-
ing the delay distribution to a Dirac distribution, the formula can be further simplified.*

$$\kappa_i(a) = \frac{\int_0^\infty \kappa_0(a) \left\{ \kappa_{i-1}(b) - p \int_0^a (-\kappa_{i-1}'(b+c) - \alpha\kappa_{i-1}(b+c)) \int_c^a \phi(a'-c)da'dc \right\} db}{\int_0^\infty \kappa_{i-1}(\tau)d\tau}$$

$$= \kappa_0(a) \left\{ 1 - p \frac{\int_0^{a-T} -\kappa_{i-1}^+(a')da' - \alpha \int_0^{a-T} K(a')da'}{\int_0^\infty \kappa_{i-1}(\tau)d\tau} \right\},$$

$$K(a') = \int_{a'}^\infty \kappa_{i-1}^+(b)db$$

Proof:

Remember how Fubini can be used to change the order of finite integrals in $\mathbb{R}^2$ of an integrable function $f(x, y)$ by swapping the integrals.

$$\int_X \left(\int_Y f(x,y) dy \right) dx = \int_{X \times Y} f(x,y) d(x,y) = \int_Y \left(\int_X f(x,y) dx \right) dy$$

The simplification consists of two steps. First, simplify by integrating and respecting $\kappa(0) = 1$:

$$\int_0^{a-T} \int_0^{\infty} -\kappa_{i-1}^{+}{}'(b + a') db = \int_0^{a-T} -\kappa_{i-1}^{+}(b + a') db.$$

Then by separating an inner one dimensional integral, which can be precalculated:

$$\int_0^{a-T} \int_0^{\infty} \alpha \kappa_{i-1}^{+}(b + a') db da' = \alpha \int_0^{a-T} \int_{a'}^{\infty} \kappa_{i-1}^{+}(b) db da'$$

In this case, the shift of the function by a' is replaced by a shift of the array bounds. Both by employing Fubini to exchange inner and outer integral.

Then convolving with the distribution $\phi(\tau)$ is now simplified to a shift: $\int_c^a f(\tau)\phi(\tau - c) d\tau = \int_c^a f(\tau)\delta_T(\tau - c) d\tau$. As in our case $f(a) = \kappa(a) = 0, \quad \forall a < 0$, this is similar to shifting the integration bounds $\int_0^{a-T} f(\tau) d\tau$.

$$\int_0^a \left(\frac{-\kappa_{i-1}^{+}{}'(b + c)}{\kappa_{i-1}^{+}(b)} - \frac{\alpha \kappa_{i-1}^{+}(b + c)}{\kappa_{i-1}^{+}(b)} \right) \int_c^a \phi(a' - c) da' dc$$

Swap the inner and outer integral.

$$= \int_0^a \int_0^{a'} \left(\frac{-\kappa_{i-1}^{+}{}'(b + c)}{\kappa_{i-1}^{+}(b)} - \frac{\alpha \kappa_{i-1}^{+}(b + c)}{\kappa_{i-1}^{+}(b)} \right) \delta_T(a' - c) dc da'$$

$$= \frac{1}{\kappa_{i-1}^{+}(b)} \int_0^a \left(-\kappa_{i-1}^{+}{}'(b + a' - T) - \alpha \kappa_{i-1}^{+}(b + a' - T) \right) da'$$

$$= \frac{1}{\kappa_{i-1}^{+}(b)} \cdot \int_0^{a-T} \left(-\kappa_{i-1}^{+}{}'(b + a') - \alpha \kappa_{i-1}^{+}(b + a') \right) da'$$

Multiply by $\kappa_{i-1}^+(b)$ and integrate

$$\Rightarrow \int_0^\infty \int_0^{a-T} \left(-\kappa_{i-1}^+{}'(b+a') - \alpha\kappa_{i-1}^+(b+a') \right) da'db$$

$$= \int_0^{a-T} \int_0^\infty -\kappa_{i-1}^+{}'(b+a')db - \int_0^\infty \alpha\kappa_{i-1}^+(b+a')dbda'$$

$$= \int_0^{a-T} -\kappa_{i-1}^+(b+a')da' - \int_0^{a-T} \int_0^\infty \alpha\kappa_{i-1}^+(b+a')dbda'$$

$$= \int_0^{a-T} -\kappa_{i-1}^+(b+a')da' - \alpha \int_0^{a-T} \int_{a'}^\infty \kappa_{i-1}^+(b)dbda'$$

$\int_{a'}^\infty \kappa_{i-1}^+(b)db$ can now be precalculated to save computation time

Inserted into the whole equation, we get:

$$\kappa_i(a)$$
$$= \frac{\int_0^\infty \kappa_0(a) \left\{ \kappa_{i-1}(b) - p\int_0^a (-\kappa_{i-1}'(b+c) - \alpha\kappa_{i-1}(b+c)) \int_c^a \phi(a'-c)da'dc \right\} db}{\int_0^\infty \kappa_{i-1}(\tau)d\tau}$$

$$= \kappa_0(a) \frac{\int_0^\infty \kappa_{i-1}(b)db}{\int_0^\infty \kappa_{i-1}(\tau)d\tau}$$

$$- \kappa_0(a) \frac{\int_0^\infty \left\{ p\int_0^a (-\kappa_{i-1}'(b+c) - \alpha\kappa_{i-1}(b+c)) \int_c^a \phi(a'-c)da'dc \right\} db}{\int_0^\infty \kappa_{i-1}(\tau)d\tau}$$

$$= \kappa_0(a) \left\{ 1 - p\frac{\int_0^{a-T} -\kappa_{i-1}^+(a')da' - \alpha \int_0^{a-T} K(a')da'}{\int_0^\infty \kappa_{i-1}(\tau)d\tau} \right\},$$

$$K(a') = \int_{a'}^\infty \kappa_{i-1}^+(b)db$$

$\square$

5.1.4 Samples of the Onset Age Structure

In section 2.4 'The Age Structure', we take an age distribution at time $t = 5$ to represent the onset age distribution. However, fig. 5.4 shows $t = 5$ to be about to leave the onset phase of exponential growth. We could not choose a much earlier time, because the cohort starting at time $t = 0$ does not reach the maximum time of numeric evaluation we choose to be at $a = 5$. Furthermore, the probability of being infective, shown in fig. 5.5, visualizes that there is still a significant probability to be infective at earlier ages, like the age of 3. The following graphs show that $t = 5$ is a valid choice to visualize the age distribution of the onset case.
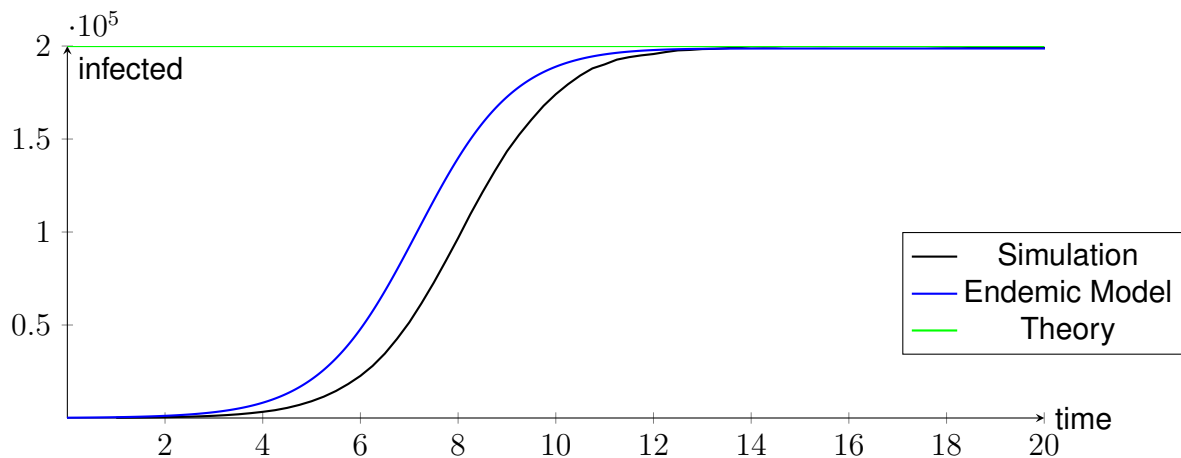
Figure 5.4: The count of infected individuals among a population of $4 \cdot 10^5$ and a contact rate $\beta = 2$, spontaneous recovery $\alpha = 0.1$ and supervised recovery $\sigma = 0.9$. Black shows the simulation of the spread, blue the partial differential equation and green the theoretical equilibrium.
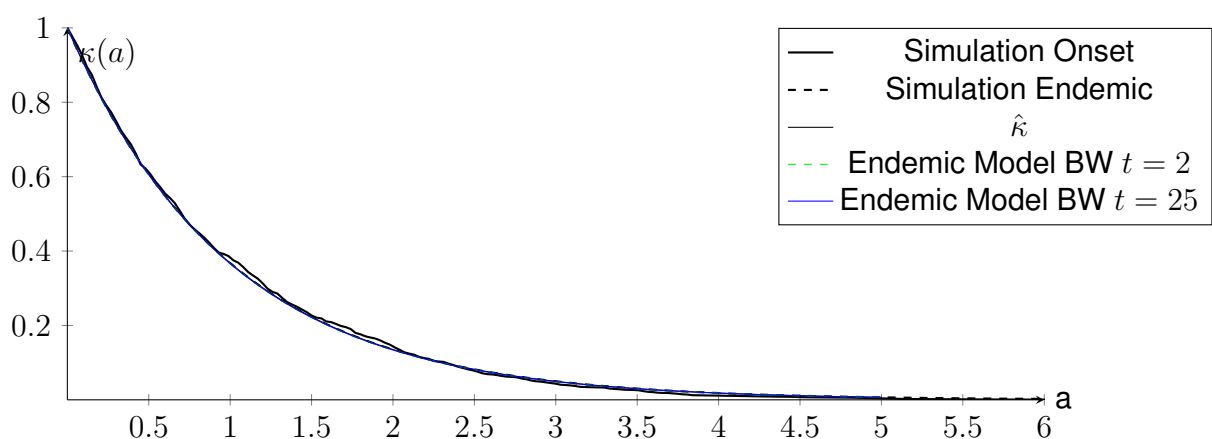
Figure 5.5: This $\kappa(a)$ corresponds to fig. 5.4. In opposite to the age distribution, the probability to stay infective after a period of time a of the onset is similar to the endemic case, if no tracing is performed.

Figure 5.5 shows the probability of being infective after a period of time a to be independent of the onset and endemic caste, which is expected without tracing. Figure 5.6 shows the onset and endemic age distributions to be different. The age distribution sampled at $t = 2$ shows the cohort starting at $t = 0$ as a spike at $a = 2$, because the individuals of the initial Dirac impulse cohort have a significant probability of being still infective at age $a = 2$. However, there is just very little difference between the cohort at $t = 2$ and $t = 5$ apart from the artifacts from the initial population creating the spike in the age distribution at $t = 2$. The logarithmic plot fig. 5.7 support this observation. Therefore, we choose the age distribution at $t = 5$ to omit the starting Dirac age group[1] artifacts. We do not distract the reader from the message, which is the impact of exponential growth on the age distribution.
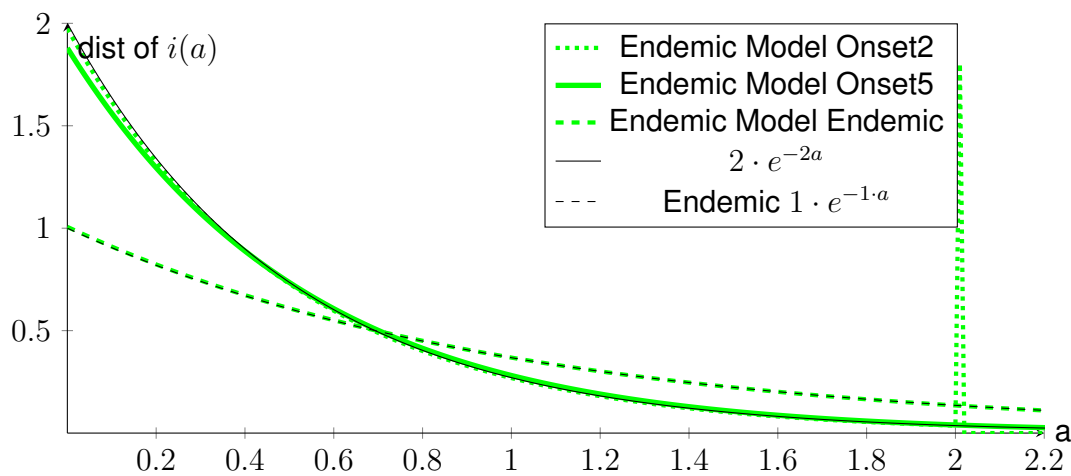
Figure 5.6: Linear plot of the age distribution of $i(a)$.
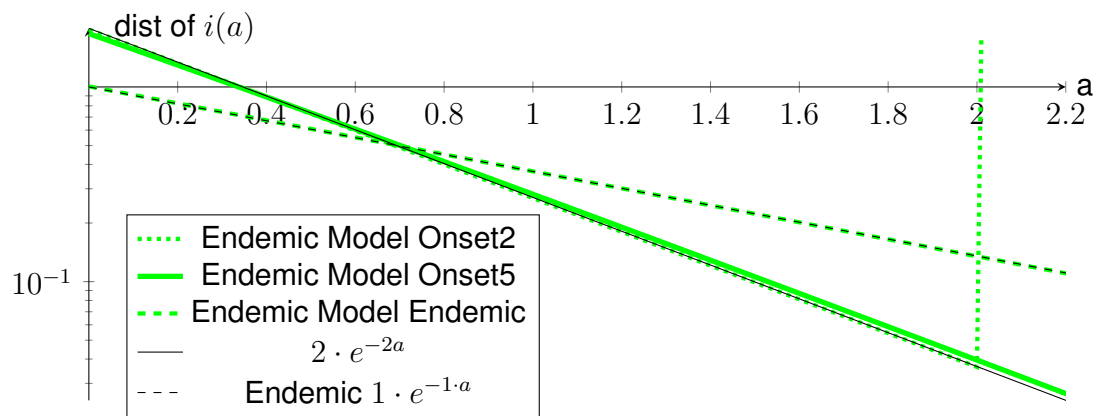
Figure 5.7: Logarithmic plot of the age distribution of $i(a)$.

---

[1]An age group consists of individuals of a cohort of a specified age.

6 Appendix 2 – Programs

This chapter describes the concepts and functionality of the larger programs used within the thesis.

6.1 Contact Tracing

The iterative models and delay equations of the onset case of contact tracing were solved using the methods of R [R C18] and the package deSolve [SPS10], which solves ordinary differential equations, delay differential equations, and partial differential equations.

6.1.1 Simulation

The simulation takes the size of a starting population, the initial percentage of infected individuals, the time to end the simulation, and the rates defining the spread of the infection. The output consists of two tables. One table notes the number of infected individuals every time an infection or recovery took place. The other table stores the time of the infection and the duration of the infection for every healed individual. That table will be used to sample κ.

Concept

The time between contacts can be modeled by an exponential distribution. Therefore, we can use an exponential distribution to predict the next contact per individual; the contact is seen as an event. These individuals are then stored in a queue sorted by the next event to be processed. Furthermore, the queue contains supervised and unsupervised recovery events.

When using this architecture, there is no time grid, which dictates a resolution. There is no iteration taking place without an event being processed. These two properties ensure a high resolution in time and a high speed, which only depends on the number of events being processed.

Implementation
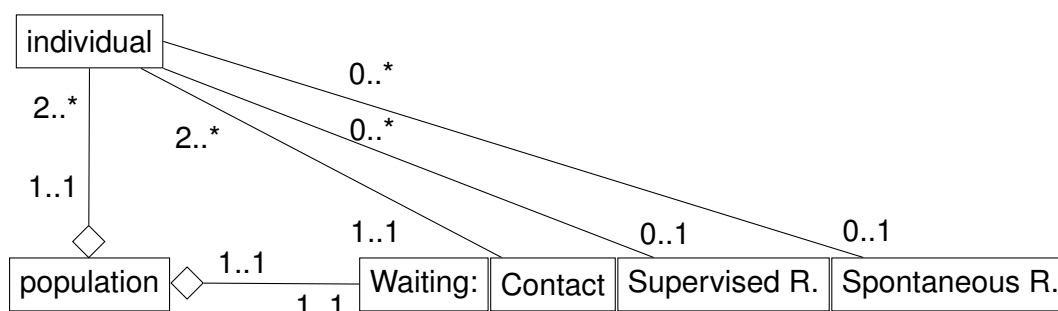
Figure 6.1: Simplified core class structure of the C++ Simulation.

175

The simulation of the contact tracing bases on two classes, the class 'population' and its corresponding individuals, modeled in the class 'individual'. The class 'population' plans the simulation. It creates the individuals in its constructor, simulates, e.g., plans, the contact, and keeps track of all contact and recovery processes by a queue. The contacts are distributed exponentially. Therefore, the time until contact occurs can be drawn from an exponential distribution and memorized by storing a pointer in the contact queue. When executing a contact, a pointer of the 'individual' is stored in the queue if an infection occurs and new contacts are planned. The probability p_{obs} defines if the individual will be enqueued in the queue as supervised (medic) or spontaneous recovery. Supervised recovery means a medic treats the individual, and the contact tracing is triggered. The relation of the classes and queues is outlined in fig. 6.1.

6.1.2 Evaluate the Endemic Model

Given a population size N, a number of initially infected individuals I_0, the contact rate β, and the parameters of supervised- and unsupervised recovery σ and α, it calculates the number of infected individuals over time. Like the IBM simulation, this program takes a maximum time to stop the numeric evaluation.

In addition, it takes a maximum age per individual a_{max} and a resolution h, which was not necessary to simulate the population using the IBM simulation from section 6.1.1. A matrix, which saves $\kappa(t, a)$, is generated using that information.

The recall likelihood of contacts p, the infectivity likelihood p_{is}, and the tracing delay T are needed to calculate the tracing. The forward tracing, backward tracing (with and without adaption to the endemic case), and cluster contact tracing can be activated.

The results of the probabilities and rates are stored in matrices of the sample size $(0 : n_{a_{\mathsf{max}}}, -a_{\mathsf{max}} - 1.1\,T : n_{t_{\mathsf{max}}}))$.

We use the forward Euler algorithm due to the simplicity of its implementation. The negative side effects of this algorithm, like underestimating exponential decay, and the stability issues do not influence the result at a significant level when evaluating short periods at the chosen levels of α and σ. Therefore, we estimate the decay and evaluate the Dahlquist equation [Dah56]. During the thesis, we choose an overall supervised and unsupervised recovery rate of the magnitude of 1. When searching for an upper limit for the tracing effect, we can employ a rough estimation by taking the contact rate, which we choose to be 2. To leave some buffer, we estimate $\lambda = -10$. Now, we evaluate the equation $|1 + h\,\lambda| < 1$ to calculate the stability area. As long we ensure $h < 0.1$, the equation hints stability even after these generous estimations. The forward Euler only achieves linear convergence, so we choose $h = 0.01$. On the other hand, the forward Euler algorithm does not spread the influence of the discontinuous first derivative, which is introduced by the impact of the tracing. There is only a piecewise linear approximation, no interpolation by polynomials. So it does not spread the errors, which polynomials can not interpolate, but it keeps this discontinuity localized.

Program Flow

Initialization

- All vectors and matrices are filled with zeros after initialization.

- Initialize a time axis vector for the time t and age of the cohort a, which map a time t or age a for each sample n. The time contains a history up to $-a_{\mathsf{max}} - 1.1\,T$ to include

the history for every age group at time 0. An age group is a group of individuals of a cohort at a specified time. The history of an age group consists of the birth of the cohort until the specified age at that time. The value h saves the time difference between two samples.

- Initialize matrices for the probability to still be infective after aging a timesteps $\kappa_-(n_a, n_t)$ and its derivative $\dot{\kappa}_-(t, a)$.

- Use similar matrices to approximate the forward tracing κ_+.

- We use a matrix of similar size $i(n_a, n_t)$ to store the number of infected individuals per age group sample and time sample, which are calculated using a hazard rate derived from κ_+.

- Initialize similar matrices for the rate of backward tracing by IS contacts and for the rate of tracing by cluster contacts.

- Initialize the vectors $I(n_t)$ and $S(n_t)$, storing the number of infected and susceptible individuals at the specific time samples.

- Save the sample index of time $t = 0$. The matrix storing the probability of being infective is initialized with a Dirac start population $i(0, n_{t_0}) = I_0$. Before time $t = 0$, all samples of age zero are initialized to have an initial probability of being infective of $k_-(0, 0 : n_{t_0}) = 1$.

Now iterate the main loop from n_{t_0} till reaching the second to last sample.

- Sum over all age groups of the matrix, saving the infected individuals per time and age group to update $I(n_t)$ and $S(n_t)$. This corresponds to the sum of the column of $i(n_a, n_t)$.

- Calculate the IS tracing rate for every age group sample n_a.
  - Calculate the adaption of the tracing rate to the higher rate of infected individuals. Thus, save a vector S_{fraction} of length $n_a - n_T + 1$ storing the fraction of susceptible individuals starting at the current time minus the number of tracing delay samples, $n_t - n_T$, till the current time minus the maximum age $n_t - n_{a_{\max}}$, we consider for cohorts.
  - Calculate the IS effect by $\text{IS}(n_a, n_t) = p\,\beta\,\sigma\sum S_{\text{fraction}} \cdot \kappa_-(0 : n_a - n_T, n_t - n_T) \cdot h$.

- Calculate the cluster contact tracing rate by evaluating the integral

$$\mathsf{CC}(n_a, n_t) = p \int_0^{a-T} \beta \int_0^\infty \frac{i(\tilde{a}, t - T - c)}{N} \frac{\kappa_-(c + \tilde{a}, t - T)}{\kappa_-(\tilde{a}, t - c - T)} \sigma d\tilde{a} dc$$

  for every n_a at the current time n_t. This is implemented by three loops.
  The outer loop iterates the age group $n_a = n_{a_{\max}}$ to be able to work on scalar results.

  - Then we integrate the history of the cohort by iterating the variable c. Iterate n_c from 0 till $n_a - n_T$.

    * The innermost loop then respects the age distribution of the contacted individual at the time of contact. Therefore, $\tilde{a}$ integrates all possible age groups, and we iterate $n_{\tilde{a}} = 0 : n_{a_{\max}}$.

  - We describe the series of nested loops by employing sums. The h is introduced by the integral over time.

$$\mathsf{CC}(n_a, n_t) = p \, \beta \, \sigma \, h \sum_{n_c=0}^{n_a - n_T} \frac{1}{N} \sum_{n_{\tilde{a}}=0}^{n_c + n_{\tilde{a}} < n_{a_{\max}}} i(n_{\tilde{a}}, n_t - n_T - n_c) \frac{\kappa_-\left(n_{\tilde{a}}, n_t - n_c - n_T\right)}{\kappa(n_c + n_{\tilde{a}}, n_t - n_T)}.$$

- Apply backward and cluster contact tracing to the $\dot{\kappa}_-$ matrix column by adding the backward tracing and cluster contact tracing correction to the healing rates:

$$\dot{\kappa}_-(:, n_t) = -k_-(:, n_t) \cdot (\alpha + \sigma + \mathsf{IS}(:, n_t) + \mathsf{CC}(:, n_t)).$$

- Calculate a forward Euler step.

$$\kappa_-(1 : n_{a_{\max}} - 1, n_t + 1) = \kappa_-(0, n_{a_{\max}} - 2, n_t + \dot{\kappa}_-(0, n_{a_{\max}} - 2, n_t) \cdot h$$

- Apply the forward tracing correction.

  - In the case of no tracing or an age of the cohort n_a smaller than the tracing delay n_T, this is identical to the κ_- and $\dot{\kappa}_-$ matrices.

  - In the case of $n_a \geq n_T$, apply the forward tracing correction $\kappa_+(n_a : n_{a_{\max}}, n_t) = \kappa_-(n_a : n_{a_{\max}}, n_t) - p \, p_{\mathsf{obs}} \left(k_-(n_a : n_{a_{\max}}, n_t) \cdot (1 - \hat{\kappa}((n_a : n_{a_{\max}} - n_T)))\right).$

  - Update the derivative $\dot{\kappa}_+$ in the direction of the cohort.

- Progress the existing cohorts at the time sample n_t to $n_t + 1$ and from n_a to $n_a + 1$ by

employing an Euler step using the hazard rate derived from κ_+:

$$i(1 : n_{a_{\max}}, n_t + 1) = i(0 : n_{a_{\max}} - 1) + i(0 : n_{a_{\max}} - 1)\frac{\dot{\kappa}_+(0 : n_{a_{\max}} - 1, n_t)}{\kappa_+(0 : n_{a_{\max}} - 1, n_t)}h.$$

- Produce the infected individuals of the next iteration using the number of infected individuals $I(n_t)$ and the contact rate β multiplied by the time till the next step h. The number of new infected individuals is adapted by the fraction of infections per contact p_{IS} and the probability of meeting a susceptible individual $\frac{S(n_t)}{N}$. Thus, $i(0, n_t + 1) = \beta\, p_{\mathsf{IS}}\, \frac{S(n_t)}{N}\, I(n_t)\, h$.

6.2 Gradient Matching

6.2.1 Estimate Parameters

This program takes the data of a dynamical system of a known number of observed components.

It assumes the dynamical system to be a system of ordinary differential equations having a polynomial right-hand side.

It takes a penalty parameter λ. The size of λ determines if the penalty regression favors larger or smaller sets of parameters to describe the dynamical system.

The result is a set of parameters resembling the observed system as close as possible while the set is kept as small as possible.

Variations Many variations of this program can perform similar tasks. Instead of using the data of a dynamical system, we can use a system of ordinary differential equations to generate the data. These equations allow generating an exact derivative. Because this eliminates possible errors, the algorithm itself can be benchmarked, the robustness against certain types of noise, the influence of smoothing, and the performance of the numerical derivative.

We decided not to use polynomials or splines to smooth the data and by using the five-point stencil as few polynomials as possible to derive the data. This decision was taken with the intention to create as few polynomial artifacts as possible, which might be picked up by the regression step.

Program Flow

Before describing the body of the program, we introduce some subroutines:

The Five-Point Stencil

$$f'(x(n)) \approx \frac{-f(x(n+2n) + 8f(x(n+1)) - 8f(x(n-1) + f(x(n-2))}{12h}.$$

The first and last two samples of the returned vector are set to zero.

The constant h defines the distance between two samples. The five-point stencil is derived by taking the Taylor series for $f(x \pm h)$ and $f(x \pm 2h)$ and solving it for $f'(x)$.

Alternative methods used to estimate a derivative were fitting and evaluating local polynomials or using the exact derivative, in case the data was evaluated from a system of ordinary differential equations, which then provide an exact derivative.

181

Penalized Regression Possible algorithms to perform the penalized regression are the orthogonal matching pursuit [TW10] and iterative thresholding [FR08]. These algorithms perform an l_0 regression.

As a very shortened and rough description, they resemble an incomplete depth-first search and a breath-first search of monomials of the right-hand side. To concentrate on the feasibility of the approach to gain the true right-hand side of the dynamical system, which allows for physical interpretation, we choose a brute force search and iterate through the whole space and not chose the approaches, which are much more feasible from a performance point of view. This brute force search allowed for a ranking to compare the best result with the alternatives to estimate the robustness of the value and the robustness of the approach. The size of the space is manageable because the approach starts to suffer from ambiguous solutions when exceeding more than three components in the presence of realistic noise levels. Furthermore, we do not need degrees greater than two within the dynamical system to describe most practical situations. In exchange, we know to find the best approximation instead of approximating it, which adds a new layer of uncertainty. In addition, this uncertainty depends on the quality of the algorithm, its extensions to ensure good approximations, and the quality of the implementation, which allows for programming errors. We avoid this uncertainty by choosing the brute force algorithm.

In practice, the faster algorithms can approximate the outcome of the brute force algorithm in many cases while vastly outperforming it, especially in the case of higher degrees.

From a performance point of view, the possibilities to avoid l_0 optimization should be evaluated. While the sparsity of the model is an advantage, the lack of convexity results in a noncontinuous behavior of parameters. Algorithms like lasso employing the l_1-norm do not reach the sparsity, which can be achieved using l_0. The problems resulting from the noncontinuous behavior might be fixed when gradually combining different thresholding methods [LF20].

Brute Force Penalized Regression The brute force penalized regression iterates through all possible right-hand sides up to a limited number of monomials, limited maximum exponent, and a limited number of components per monomial, which does not affect the case of only two components without prohibiting any interaction.

The brute force algorithm calculates a residual for each right-hand side. However, it adds a penalty of the size of λ, the regularization parameter, for each monomial within the right-hand side. This penalized regression searches for the best fit while trying to utilize as few monomials as possible.

In many cases, normalization helps. In the case of an unnormalized dynamical system, the

algorithm uses too many monomials to limit the residual of large components and neglects smaller ones, which might be as essential to reproduce the overall dynamic of the system. On the other hand, employing an orthonormalization results in a sparse solution of the orthonormal system. This result does not necessarily offer a sparse solution in the original space and might not translate into a physical explanation.

Initialization

- Load the data of the dynamical system.

- Define: Do not evaluate the first and last samples, defined by the variable 'cropmargin', of a function to avoid introducing more errors by a broken derivative.

- Define: Limit the exponent and the number of components to two.

- Define: Use the five-point stencil to derive.

- Define: Normalize.

The Body

- Initialize

- Load data

- Preprocessing

    - Add noise (if defined)

    - Smooth (if defined)

    - Calculate the derivative (using the defined algorithm)

    - Crop a margin of n samples (if defined)

    - Normalize

- Penalized regression (using the defined algorithm)

- Postprocessing

    - Denormalize

    - Plot the results and solve the estimated dynamical system for comparison.

6.2.2 Estimate Component

This program estimates an unobserved component of an ordinary differential equation of two components.

Therefore, it takes the data of an observed component, the first component, and the parameters of the ODE, which define its structure.

Program flow

First, apply a median filter and perform a running mean filter of length k. Then, perform a data reduction by subsampling and take every n-th element.

Before starting the actual program, define some subroutines.

The Five-Point Stencil

$$f'(x(n)) \approx \frac{-f(x(n+2n)) + 8f(x(n+1)) - 8f(x(n-1)) + f(x(n-2))}{12h}.$$

The first and last two samples of the returned vector are set to zero.

The constant h defines the distance between two samples. The five-point stencil is derived by taking the Taylor series for $f(x \pm h)$ and $f(x \pm 2h)$ and solving it for $f'(x)$.

Limit the Values   The theory limited the space to a maximum value. To automate this process this function limits the values to exist between the 0.01 and 0.99 quantiles.

The Functional J Describing the Error

- Get a numeric derivative $y'(n)$ of $y(n)$ using the five-point stencil.

- Calculate the returned error
  $e = \sum_0^N \left[x'(n) - f_1(x(n), y(n)) \right]^2 + \sum_0^N \left[y'(n) - f_2(x(n), y(n)) \right]^2.$

The Body

- Initialize the second component by random values.

- Calculate the derivative of the first component using the five point stencil and cap the resulting values to exist between the 0.01 and 0.99 quantiles.

- Start the optimization of the second component using the functional J. The optimization is performed by the optimizer Rgenoud [MS11]. Rgenoud is based on the

Genoud algorithm, which is based on BFGS, but it also uses genetic algorithms to increase the probability of finding the global optimum [SM98].

• Plot the result.

Alternative Implementations   In the Lotka-Volterra example, we used total variation differentiation [Cha11] to gain the initial derive of f_1. The total variation is the sum of differences between adjacent samples. The idea of using total variation is the correlation: The lower the total variation, the lower the noise level. This way of differentiating the function chooses the derivative to match the observed data in the 2-norm after being integrated. Furthermore, the noise is controlled in the image space by minimizing the total variation. This is a remarkably robust way of calculating a derivative, but it can not be used within the optimization without having to implement more adaptions to deal with the behavior of the total variation term.

Furthermore, the Lotka-Volterra example caps the values of the derivative at every evaluation of J.

List of Figures

List of Tables

Bibliography

[AB00]   Håkan Andersson and Tom Britton. *Stochastic epidemic models and their sta-tistical analysis*. Vol. 151. Lecture notes in statistics. New York, N.Y.: Springer, 2000. ISBN: 978-0-387-95050-1. DOI: 10.1007/978-1-4612-1158-7.

[And82]   Roy M. Anderson. *The Population Dynamics of Infectious Diseases: Theory and Applications*. Population and Community Biology. Boston, MA: Springer US, 1982. ISBN: 978-1-4899-2901-3. DOI: 10.1007/978-1-4899-2901-3.

[Bag92]   Lawrence W. Baggett. *Functional analysis: A primer*. Vol. 153. Monographs and textbooks in pure and applied mathematics. New York [etc.]: Marcel Dekker, op. 1992. ISBN: 0824785983.

[Bai64]   Norman Thomas John Bailey. *The elements of stochastic processes: With ap-plications to the natural sciences*. A Wiley publication in applied statistics. New York: Wiley, 1964. ISBN: 0471041653.

[Bak+63]   Kate Bak et al. "Numerical Differentiation and Integration in Chemical Kinetics". In: *Acta Chemica Scandinavica* 17 (1963), pp. 985–991. ISSN: 0904-213X. DOI: 10.3891/acta.chem.scand.17-0985.

[BCd14]   Nicolas J-B Brunel, Quentin Clairon, and Florence d'Alché-Buc. "Parametric Estimation of Ordinary Differential Equations With Orthogonality Conditions". In: *Journal of the American Statistical Association* 109.505 (2014), pp. 173–185. ISSN: 0162-1459. DOI: 10.1080/01621459.2013.841583.

[BD95]   Frank Ball and Peter Donnelly. "Strong approximations for epidemic models". In: *Stochastic Processes and their Applications* 55.1 (1995), pp. 1–21. ISSN: 03044149. DOI: 10.1016/0304-4149(94)00034-Q.

[BG96]   George A. Baker and Peter Graves-Morris. *Padé Approximants Second Edi-tion*. Cambridge: Cambridge University Press, 1996. ISBN: 9780511530074. DOI: 10.1017/CBO9780511530074.

[Bil99]   Patrick Billingsley. *Convergence of probability measures*. 2. ed. Wiley series in probability and statistics Probability and statistics section. New York, NY: Wiley, 1999. ISBN: 0-471-19745-9.

[BKO15]   Frank G. Ball, Edward S. Knock, and Philip D. O'Neill. "Stochastic epidemic models featuring contact tracing with delays". In: *Mathematical Biosciences* 266 (2015), pp. 23–35. ISSN: 00255564. DOI: 10.1016/j.mbs.2015.05.007.

[BP19]   Tom Britton and Etienne Pardoux. *Stochastic Epidemic Models with Inference*. Vol. 2255. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-30899-5. DOI: 10.1007/978-3-030-30900-8.

[BPK16]   Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.15 (2016), pp. 3932–3937. ISSN: 0027-8424. DOI: 10.1073/pnas.1517384113.

[Bré11]   Haim. Brézis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. New York and London: Springer, 2011. ISBN: 978-0-387-70913-0.

[Bru08]   Nicolas J-B Brunel. "Parameter estimation of ODE's via nonparametric estimators". In: *Electronic Journal of Statistics* 2.0 (2008), pp. 1242–1267. ISSN: 1935-7524. DOI: 10.1214/07-EJS132.

[Bur+18]   Christopher P. Burgess et al. *Understanding disentangling in β-VAE*. Apr. 2018. URL: http://arxiv.org/pdf/1804.03599v1.

[BvW08]   Fred Brauer, Pauline van den Driessche, and Jianhong Wu. *Mathematical epidemiology*. Vol. 1945. Mathematical biosciences subseries. Berlin: Springer, 2008. ISBN: 978-3-540-78910-9. DOI: 10.1007/978-3-540-78911-6.

[Cav78]   James A. Cavender. "Quasi-stationary distributions of birth-and-death processes". In: *Advances in Applied Probability* 10.3 (1978), pp. 570–586. ISSN: 0001-8678. DOI: 10.2307/1426635.

[Cha11]   Rick Chartrand. "Numerical Differentiation of Noisy, Nonsmooth Data". In: *ISRN Applied Mathematics* 2011.1–4 (2011), pp. 1–11. ISSN: 2090-5564. DOI: 10.5402/2011/164564.

[CWT12]   James Clarke, K. A. Jane White, and Katy Turner. "Exploring short-term responses to changes in the control strategy for Chlamydia trachomatis". In: *Computational and mathematical methods in medicine* 2012 (2012), p. 803097. DOI: 10.1155/2012/803097.

[Dah56]   Germund Dahlquist. "Convergence and stability in the numerical integration of ordinary differential equations". In: *MATHEMATICA SCANDINAVICA* 4 (1956), p. 33. ISSN: 0025-5521. DOI: `10.7146/math.scand.a-10454`.

[DHB13]   O. Diekmann, Hans Heesterbeek, and Tom Britton. *Mathematical tools for understanding infectious diseases dynamics*. Princeton series in theoretical and computational biology. Princeton: Princeton University Press, 2013. ISBN: 9780691155395.

[DHM90]   O. Diekmann, J. A. Heesterbeek, and J. A. Metz. "On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations". In: *Journal of mathematical biology* 28.4 (1990), pp. 365–382. DOI: `10.1007/BF00178324`.

[DI18]   Bogdan Dumitrescu and Paul Irofti. *Dictionary Learning Algorithms and Applications*. Cham: Springer International Publishing, 2018. ISBN: 978-3-319-78673-5. DOI: `10.1007/978-3-319-78674-2`.

[DV98]   D. J. Daley and D. Vere-Jones. "An Introduction to the Theory of Point Processes". In: (1998). DOI: `10.1007/978-1-4757-2001-3`.

[EK02]   Ken T. D. Eames and Matt J. Keeling. "Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.20 (2002), pp. 13330–13335. ISSN: 0027-8424. DOI: `10.1073/pnas.202244299`.

[EK86]   Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: Characterization and convergence*. New York: Wiley, ©1986. ISBN: 9780471769866.

[ESS02]   Stephen P. Ellner, Yodit Seifu, and Robert H. Smith. "FITTING POPULATION DYNAMIC MODELS TO TIME-SERIES DATA BY GRADIENT MATCHING". In: *Ecology* 83.8 (2002), pp. 2256–2270. ISSN: 0012-9658. DOI: `10.1890/0012-9658(2002)083[2256:FPDMTT]2.0.CO;2`.

[Eva10]   Lawrence C. Evans. *Partial Differential Equations: Second Edition (Graduate Studies in Mathematics)*. American Mathematical Society, 2010. ISBN: 0821849743.

[F G11]   F. G. Ball, E. S. Knock, P. D. O'Neill. "Threshold behaviour of emerging epidemics featuring contact tracing". In: *Adv. in Appl. Probab.* 43 (2011), pp. 1048–1065.

[Fel68]   William Feller. *An introduction to probability theory and its applications: Volume I.* 3d ed., rev. printing. Wiley series in probability and mathematical statistics. New York: Wiley, 1970 or 1, 1968. ISBN: 9780471257080.

[Fer+20]   Luca Ferretti et al. "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing". In: *Science (New York, N.Y.)* 368.6491 (2020). ISSN: 0036-8075. DOI: 10.1126/science.abb6936.

[FK05]   Fima C Klebaner and Fima C. Klebaner. *Introduction to stochastic calculus with applications.* 2nd ed. London and Singapore: Imperial College Press and Distributed by World Scientific Pub., 2005. ISBN: 1-86094-555-4.

[FR08]   Massimo Fornasier and Holger Rauhut. "Iterative thresholding algorithms". In: *Applied and Computational Harmonic Analysis* 25.2 (2008), pp. 187–208. ISSN: 10635203. DOI: 10.1016/j.acha.2007.10.005.

[Fra+04]   Christophe Fraser et al. "Factors that make an infectious disease outbreak controllable". In: *Proceedings of the National Academy of Sciences* 101.16 (2004), pp. 6146–6151. ISSN: 0027-8424. DOI: 10.1073/pnas.0307506101.

[Fra48]   V. L. Frampton. "A Simple Tangent Meter". In: *Science* 107.2778 (1948), pp. 323–324. ISSN: 0036-8075. DOI: 10.1126/science.107.2778.323.

[Gar04]   C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry, and the natural sciences.* 3rd ed. Berlin, New York: Springer-Verlag, ©2004. ISBN: 3540208828.

[Gra+20]   Nicholas C. Grassly et al. "Comparison of molecular testing strategies for COVID-19 control: a mathematical modelling study". In: *The Lancet Infectious Diseases* 20.12 (2020), pp. 1381–1389. ISSN: 14733099. DOI: 10.1016/S1473-3099(20)30630-7.

[Het89]   Herbert W. Hethcote. "Three Basic Epidemiological Models". In: *Applied Mathematical Ecology* 18 (1989), pp. 119–144. DOI: 10.1007/978-3-642-61317-3{\textunderscore}5.

[HJB67]   D. M. Himmelblau, C. R. Jones, and K. B. Bischoff. "Determination of Rate Constants for Complex Kinetics Models". In: *Industrial & Engineering Chemistry Fundamentals* 6.4 (1967), pp. 539–543. ISSN: 0196-4313. DOI: 10.1021/i160024a008.

[HK10]   Thomas House and Matt J. Keeling. "The impact of contact tracing in clustered populations". In: *PLoS computational biology* 6.3 (2010), e1000721. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1000721.

[Hsi+05] Ying-Hen Hsieh et al. "A Class of Methods for HIV Contact Tracing in Cuba: Implications for Intervention and Treatment". In: *World Scientific* (2005), pp. 77–92. DOI: `10.1142/9789812569264{\textunderscore}0004`.

[HY84] Herbert W. Hethcote and James A. Yorke. *Gonorrhea Transmission Dynamics and Control*. Vol. 56. Springer Berlin Heidelberg, 1984. DOI: `10.1007/978-3-662-07544-9`.

[IB20] Paul Irofti and Andra Băltoiu. *Unsupervised Dictionary Learning for Anomaly Detection*. 2020. URL: `http://arxiv.org/pdf/2003.00293v2`.

[IM17] Mimmo Iannelli and Fabio Milner. *The Basic Approach to Age-Structured Population Dynamics*. Dordrecht: Springer Netherlands, 2017. ISBN: 978-94-024-1145-4. DOI: `10.1007/978-94-024-1146-1`.

[KE05] Matt J. Keeling and Ken T. D. Eames. "Networks and epidemic models". In: *Journal of the Royal Society, Interface* 2.4 (2005), pp. 295–307. ISSN: 1742-5662. DOI: `10.1098/rsif.2005.0051`.

[Kee99] M. J. Keeling. "Correlation equations for endemic diseases: Externally imposed and internally generated heterogeneity". In: *Proceedings of the Royal Society B: Biological Sciences* 266.1422 (1999), pp. 953–960. ISSN: 0962-8452. DOI: `10.1098/rspb.1999.0729`.

[KFH06] Don Klinkenberg, Christophe Fraser, and Hans Heesterbeek. "The effectiveness of contact tracing in emerging epidemics". In: *PLoS ONE* 1 (2006), e12. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0000012`.

[KHR20] Matt J. Keeling, T. Deirdre Hollingsworth, and Jonathan M. Read. "Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19)". In: *Journal of epidemiology and community health* 74.10 (2020), pp. 861–866. DOI: `10.1136/jech-2020-214051`.

[KK11] Karel J. Keesman and K. J. Keesman. *System identification: An introduction*. Advanced textbooks in control and signal processing. London and New York: Springer, 2011. ISBN: 978-0-85729-521-7.

[KK93] Gopinath Kallianpur and Rajeeva L. Karandikar. *White noise theory of prediction, filtering and smoothing*. 2. printing. Vol. 3. Stochastics monographs. New York, NY [u.a.]: Gordon and Breach Science Publ, 1993. ISBN: 2-88124-685-0.

[KM27]   William Ogilvy Kermack and Anderson Gray McKendrick. "A Contribution to the Mathematical Theory of Epidemics". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (1927), pp. 700–721. ISSN: 09501207. DOI: `10.1098/rspa.1927.0118`. URL: `http://www.jstor.org/stable/94815`.

[KM91a]   W. O. Kermack and A. G. McKendrick. "Contributions to the mathematical theory of epidemics–I. 1927". In: *Bulletin of mathematical biology* 53.1-2 (1991), pp. 33–55. ISSN: 0092-8240. DOI: `10.1007/BF02464423`.

[KM91b]   W. O. Kermack and A. G. McKendrick. "Contributions to the mathematical theory of epidemics–II. The problem of endemicity.1932". In: *Bulletin of mathematical biology* 53.1-2 (1991), pp. 57–87. ISSN: 0092-8240. DOI: `10.1007/BF02464424`.

[KM91c]   W. O. Kermack and A. G. McKendrick. "Contributions to the mathematical theory of epidemics–III. Further studies of the problem of endemicity. 1933". In: *Bulletin of mathematical biology* 53.1-2 (1991), pp. 89–118. ISSN: 0092-8240. DOI: `10.1007/BF02464425`.

[KN10]   Brian Karrer and M. E. J. Newman. "A message passing approach for general epidemic models". In: *Physical Review E* 82.1 (2010), p. 161. ISSN: 1539-3755. DOI: `10.1103/PhysRevE.82.016101`. URL: `http://arxiv.org/pdf/1003.5673v2`.

[Koj+21]   Sadamori Kojaku et al. "The effectiveness of backward contact tracing in networks". In: *Nature Physics* (2021). ISSN: 1745-2473. DOI: `10.1038/s41567-021-01187-2`.

[Kre+20]   Mirjam E Kretzschmar et al. "Time is of the essence: impact of delays on effectiveness of contact tracing for COVID-19, a modelling study". In: *medRxiv* (2020). DOI: `10.1101/2020.05.09.20096289`. eprint: `https://www.medrxiv.org/content/early/2020/06/29/2020.05.09.20096289.full.pdf`.

[Kur80]   Thomas G. Kurtz. "Relationships between stochastic and deterministic population models". In: *Lecture Notes in Biomathematics* 38 (1980), pp. 449–467. DOI: `10.1007/978-3-642-61850-5_39`.

[LF20]   Haoyang Liu and Rina Foygel Barber. "Between hard and soft thresholding: optimal iterative thresholding algorithms". In: *Information and Inference: A Journal of the IMA* 9.4 (2020), pp. 899–933. ISSN: 2049-8764. DOI: `10.1093/imaiai/iaz027`.

[Liu+15]   Fengchen Liu et al. "The role of vaccination coverage, individual behaviors, and the public health response in the control of measles epidemics: an agent-based simulation for California". In: *BMC public health* 15 (2015), p. 447. DOI: `10.1186/s12889-015-1766-6`.

[Man+16]   Niall M. Mangan et al. "Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics". In: *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* 2.1 (2016), pp. 52–63. ISSN: 2372-2061. DOI: `10.1109/TMBMC.2016.2633265`. URL: `http://arxiv.org/pdf/1605.08368v1`.

[Mey+05]   Lauren Ancel Meyers et al. "Network theory and SARS: predicting outbreak diversity". In: *Journal of Theoretical Biology* 232.1 (2005), pp. 71–81. ISSN: 00225193. DOI: `10.1016/j.jtbi.2004.07.026`.

[MH07]   Johannes Müller and Volker Hösel. "Estimating the Tracing Probability from Contact History at the Onset of an Epidemic". In: *Mathematical Population Studies* 14.4 (2007), pp. 211–236. ISSN: 0889-8480. DOI: `10.1080/08898480701612857`.

[MK16]   Johannes Müller and Bendix Koopmann. "The effect of delay on contact tracing". In: *Mathematical Biosciences* 282 (Dec. 2016), pp. 204–214. ISSN: 0025-5564. DOI: `10.1016/j.mbs.2016.10.010`. URL: `http://www.sciencedirect.com/science/article/pii/S0025556416302693`.

[MK20]   Johannes Müller and Mirjam Kretzschmar. "Contact tracing – Old models and new challenges". In: *Infectious Disease Modelling* (2020). ISSN: 24680427. DOI: `10.1016/j.idm.2020.12.005`.

[MKD00]   Johannes Müller, Mirjam Kretzschmar, and Klaus Dietz. "Contact tracing in stochastic and deterministic epidemic models". In: *Mathematical Biosciences* 164.1 (2000), pp. 39–64. ISSN: 00255564. DOI: `10.1016/S0025-5564(99)00061-9`.

[MS11]   Walter R. Mebane, Jr. and Jasjeet S. Sekhon. "Genetic Optimization Using Derivatives: The rgenoud Package for R". In: *Journal of Statistical Software* 42.11 (2011), pp. 1–26. URL: `http://www.jstatsoft.org/v42/i11/`.

[Mül90]   Müller. *Rauschen: Zweite, überarbeitete und erweiterte Auflage*. Halbleiter-Elektronik. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990. ISBN: 978-3-642-61501-6.

[Nås11]   Ingemar Nåsell. *Extinction and Quasi-Stationarity in the Stochastic Logistic SIS Model*. Vol. 2022. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. ISBN: 978-3-642-20529-3. DOI: 10.1007/978-3-642-20530-9.

[NT09]   D. Needell and J. A. Tropp. "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples". In: *Applied and Computational Harmonic Analysis* 26.3 (2009), pp. 301–321. ISSN: 10635203. DOI: 10.1016/j.acha.2008.07.002.

[OM20]   Augustine Okolie and Johannes Müller. "Exact and approximate formulas for contact tracing on random trees". In: *Mathematical Biosciences* 321 (2020), p. 108320. ISSN: 00255564. DOI: 10.1016/j.mbs.2020.108320. URL: http://arxiv.org/pdf/1910.06860v3.

[Pad92]   H. Padé. "Sur la représentation approchée d'une fonction par des fractions rationnelles". fre. In: *Annales scientifiques de l'École Normale Supérieure* 9 (1892), pp. 3–93. URL: http://eudml.org/doc/81047.

[PF16]   Daniele Perrone and Paolo Favaro. "A Clearer Picture of Total Variation Blind Deconvolution". In: *IEEE transactions on pattern analysis and machine intelligence* 38.6 (2016), pp. 1041–1055. DOI: 10.1109/TPAMI.2015.2477819.

[PT07]   M. Peifer and J. Timmer. "Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting". In: *IET systems biology* 1.2 (2007), pp. 78–88. ISSN: 1751-8849. DOI: 10.1049/iet-syb:20060067.

[R C18]   R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2018. URL: https://www.R-project.org/.

[Ren12]   Eric Renshaw. *Modelling Biological Populations in Space and Time*. Cambridge University Press, 2012. ISBN: 9780521303880. DOI: 10.1017/CBO9780511624094.

[Ros96]   Sheldon M. Ross. *Stochastic processes*. 2. ed. Wiley series in probability and statistics. New York, NY: Wiley, 1996. ISBN: 0-471-12062-6.

[RR04]   Michael Renardy and Robert C. Rogers. *An introduction to partial differential equations*. Second Edition // Second edition. Vol. 13. Texts in Applied Mathematics. New York, NY: Springer-Verlag New York Inc and Springer, 2004. ISBN: 9780387004440. DOI: 10.1007/b97427.

[SB75]   J. Swartz and H. Bremermann. "Discussion of parameter estimation in biological modelling: Algorithms for estimation and evaluation of the estimates". In: *Journal of mathematical biology* 1.3 (1975), pp. 241–257. DOI: 10.1007/BF01273746.

[SB96]   Josef Stoer and Roland Bulirsch. *Introduction to numerical analysis*. 2nd ed. Vol. 12. Texts in Applied Mathematics. New York: Springer, 1996. ISBN: 0-387-97878-X.

[Sca+21]   Francesca Scarabel et al. "A renewal equation model to assess roles and limitations of contact tracing for disease outbreak control". In: (2021). DOI: 10.1101/2020.12.27.20232934.

[SL09]   Michael Schmidt and Hod Lipson. "Distilling free-form natural laws from experimental data". In: *Science (New York, N.Y.)* 324.5923 (2009), pp. 81–85. ISSN: 0036-8075. DOI: 10.1126/science.1165893.

[SM98]   Jasjeet S. Sekhon and Walter R. Mebane, Jr. "Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models". In: *Political Analysis* 7 (1998), pp. 189–213.

[SMS94]   K. Sato, H. Matsuda, and A. Sasaki. "Pathogen invasion and host extinction in lattice structured populations". In: *Journal of mathematical biology* 32.3 (1994), pp. 251–268. DOI: 10.1007/BF00163881.

[SPS10]   Karline Soetaert, Thomas Petzoldt, and R. Woodrow Setzer. "Solving Differential Equations in R: Package deSolve". In: *Journal of Statistical Software* 33.9 (2010), pp. 1–25. ISSN: 1548-7660. DOI: 10.18637/jss.v033.i09. URL: http://www.jstatsoft.org/v33/i09.

[Squ71]   William Squire. "A simple integral method for system identification". In: *Mathematical Biosciences* 10.1-2 (1971), pp. 145–148. ISSN: 00255564. DOI: 10.1016/0025-5564(71)90055-1.

[Tro+05]   J. A. Tropp et al. "Designing structured tight frames via an alternating projection method". In: *IEEE Transactions on Information Theory* 51.1 (2005), pp. 188–209. ISSN: 0018-9448. DOI: 10.1109/TIT.2004.839492.

[Tro06]   J. A. Tropp. "Just relax: Convex programming methods for identifying sparse signals in noise". In: *IEEE Transactions on Information Theory* 52.3 (2006), pp. 1030–1051. ISSN: 0018-9448. DOI: 10.1109/TIT.2005.864420.

[TS11]   Filip Tavernier and Michiel Steyaert. "Optical communication—A high-level perspective". In: *High-Speed Optical Receivers with Integrated Photodiode in Nanoscale CMOS*. Ed. by Filip Tavernier and Michiel Steyaert. Analog Circuits and Signal Processing. New York, NY: Springer Science+Business Media LLC, 2011, pp. 15–39. ISBN: 978-1-4419-9925-2. DOI: `10.1007/978-1-4419-9925-2_2`.

[TW10]   Joel A. Tropp and Stephen J. Wright. "Computational Methods for Sparse Solution of Linear Inverse Problems". In: *Proceedings of the IEEE* 98.6 (2010), pp. 948–958. ISSN: 0018-9219. DOI: `10.1109/JPROC.2010.2044010`.

[Var82] J. M. Varah. "A Spline Least Squares Method for Numerical Parameter Estimation in Differential Equations". In: *SIAM Journal on Scientific and Statistical Computing* 3.1 (1982), pp. 28–46. ISSN: 0196-5204. DOI: `10.1137/0903003`.

[Wal74] Wolfgang Walter. *Einführung in die Theorie der distributionen*. [Neuaufl.] Zürich: Bibliographisches Institut, 1974. ISBN: 3411014563.

[WBS17] Robert R. Wilkinson, Frank G. Ball, and Kieran J. Sharkey. "The relationships between message passing, pairwise, Kermack-McKendrick and stochastic SIR epidemic models". In: *Journal of mathematical biology* 75.6-7 (2017), pp. 1563–1590. DOI: `10.1007/s00285-017-1123-8`.

[WL07]   J. Wallinga and M. Lipsitch. "How generation intervals shape the relationship between growth rates and reproductive numbers". In: *Proceedings. Biological sciences* 274.1609 (2007), pp. 599–604. ISSN: 0962-8452. DOI: `10.1098/rspb.2006.3754`.

[YHN78] J. A. Yorke, H. W. Hethcote, and A. Nold. "Dynamics and control of the transmission of gonorrhea". In: *Sexually transmitted diseases* 5.2 (1978), pp. 51–56. ISSN: 0148-5717. DOI: `10.1097/00007435-197804000-00003`.

[Yos68] Kôsaku Yosida. *Functional Analysis*. Second edition. Vol. 123. Die Grundlehren der mathematischen Wissenschaften, in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. Berlin, Heidelberg: Springer Berlin Heidelberg, 1968. ISBN: 978-3-662-11793-4.

[Yul24] George Udny Yule. "A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S". In: *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* 213.402-410 (1924), pp. 21–87. ISSN: 0264-3960. DOI: `10.1098/rstb.1925.0002`.

[Zei95]   Eberhard Zeidler. *Applied Functional Analysis: Main Principles and Their Applications*. Vol. 109. Applied Mathematical Sciences, 0066-5452. New York, NY: Springer New York, Imprint, and Springer, 1995. ISBN: 978-1-4612-6913-7.