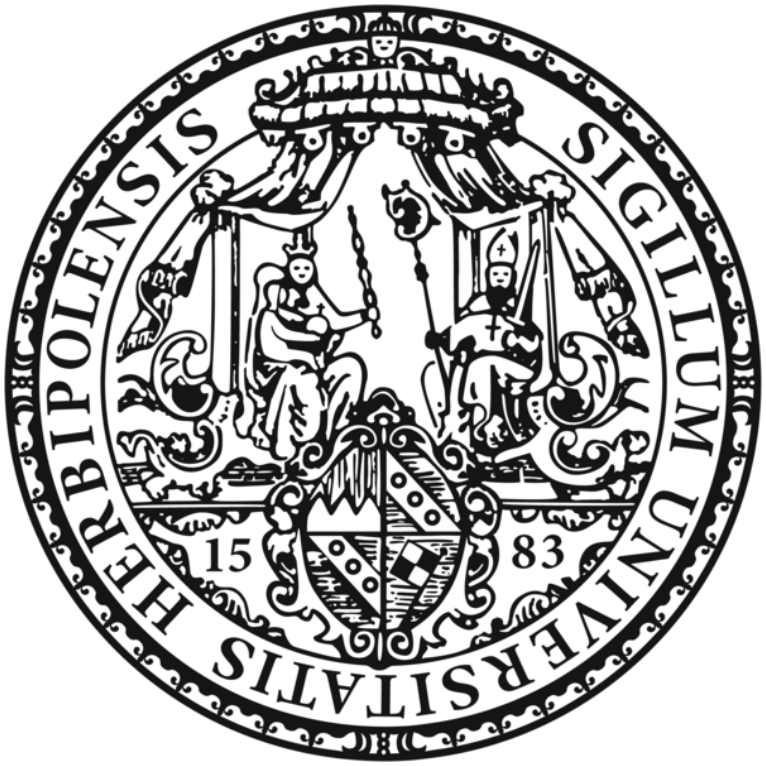
Masterthesis
in Mathematical Physics

Multiscale Analysis and Ill Conditioning in the Diffusive Limit of the Inverse Problem for Chemotaxis in the Bayesian Setting

Kaja Alina Jurak



Julius-Maximilians-University Würzburg

Supervised by
Prof. Christian Klingenberg

December 2022

Acknowledgements

First and foremost I would like to thank Prof. Klingenberg for the work and energy he put into the finding of the topic and the steady supervision in the process of writing the thesis. In particular, he provided an engaging working environment and introduced me to other researchers work within this field.

Also I am grateful to the *Studienstiftung des deutschen Volkes* for their financial and ideal support, not only during this thesis but my whole studies.

Additionally, I would like to thank Kathrin Hellmuth for the numerous productive discussions and for always having an open door. Many ideas on how to approach arising difficulties came out of those conversations.

I also want to give a special thanks to her as well as Marlies Pirner and Judith Lutz for proofreading my thesis and giving lots of useful feedback.

Last but not least, I want to thank my family and friends for supporting me during this time.

Abstract

Within this thesis the goal is to apply Bayesian inference to the inverse problem coming from chemotaxis of bacteria including a reaction term in order to reconstruct the tumbling kernel from the organism population. Here, two mathematical models on different scales - namely a kinetic and a macroscopic model - are considered and their convergence in the asymptotic limit is shown. The work is organized as follows: in section 2 we will introduce Bayesian inference by first linking it to the classical approach. Then, well-posedness in the Bayesian setting will be discussed and the ideas behind selected numerical applications will be presented. This will serve as the foundation for formulating the inverse problem for chemotaxis. In section 3 the forward models on the kinetic and macroscopic level are established and the convergence in the asymptotic limit of those two models is presented. The results of current literature, where the drift-diffusion limit for the scaled chemotaxis equation has already been carried out, are extended to the case where a reaction term occurs in the equations. This will form the basis to connect the related inverse problems in section 4 which can also be seen as an extension of the results in current literature to the chemotaxis equation where a reaction term is added. Here, the inverse problems are formulated and the relation between the two problems is investigated. The well-posedness and convergence of the resulting posterior distributions in the asymptotic limit are shown, which can be seen as an extension of the current results in literature. In section 5 the loss of information in the asymptotic limit is investigated on the local and global level but with the presented approaches in this work, the loss of information cannot be quantified. A summary and an outlook on further research is given in section 6.

Zusammenfassung

Ziel dieser Arbeit ist es, das Kalkül der Bayessischen Statistik auf das Inverse Problem von der Chemotaxis von Bakterien anzuwenden unter Berücksichtigung eines Reaktionterms, um die Zitterbewegung der Bakterien zu rekonstruieren. Hierbei werden zwei mathematische Modelle auf verschiedenen Skalen - kinetisch und makroskopisch - betrachtet und deren Konvergenz im asymptotischen Grenzfall untersucht. Die Thesis ist wie folgt strukturiert: in Abschnitt 2 wird die Bayessche Statistik und ihre Anwendung auf inverse Probleme eingeführt. Dies wird getan, indem der probabilistische Ansatz zunächst mit dem klassischen Ansatz für einfache Probleme verknüpft wird. Anschließend wird erläutert, was es in der Bayesschen Statistik bedeutet, ein korrekt gestelltes Problem zu sein und die Ideen hinter ausgewählten numerischen Verfahren werden dargestellt. Dies wird den Grundstein für die Formulierung des inversen Problems für Chemotaxis bilden. Abschnitt 3 befasst sich mit den Vorwärtsproblemen auf kinetischem und makroskopischen Niveau der Chemotaxis und die asymptotische Äquivalenz der hierbei auftretenden Gleichungen wird gezeigt. Die Resultate aus der Literatur, in welcher der Übergang zum diffusiven System für Chemotaxis bereits gezeigt wurde, werden für den Fall eines auftretenden Reaktionsterms erweitert. Auf Grundlage dessen, werden in Abschnitt 4 die jeweiligen inversen Probleme formuliert und deren Beziehung zueinander untersucht. Die Konvergenz im asymptotischen Grenzfall wird gezeigt und kann als Erweiterung der bisherigen Ergebnisse in der Literatur gesehen werden. In Abschnitt 5 wird dann der Informationsverlust während des Skalierungsprozesses näher betrachtet und ein erster Versuch unternommen, diesen zu quantifizieren. Mit den in dieser Abreit vorgestellten Ansätzen kann der Informationsverlust jedoch nicht quantifiziert werden. Abschließend gibt es eine Zusammenfassung und einen Ausblick auf weitere (Nach-)Forschung in Abschnitt 6.

Contents

1 Introduction

When describing nature, human behaviour or other kinds of real-life processes mathematically, some sort of *model* is needed. A mathematical model always reduces reality in the way that assumptions and simplifications need to be made in order to describe the system of interest. The resulting relations are then translated into the mathematical language. A model consists of a set of equations, relating an input to some output, namely the data. A more detailed introduction to mathematical modelling can be found for instance in [35, 44]. When trying to model complex systems, not every process can be represented within the mathematical description since they are either not fully known yet or the resulting equations become too complex to be dealt with. Quantitative climate models are a great example for models of complex systems and how they may vary in complexity. Here, the ingoing energy from the sun as well as the outgoing energy in form of electromagnetic waves are taken into account to calculate a change in temperature as a result from occurring imbalances. Those kind of models are named *energy-balance-models* and are described in [29]. For instance, the *radiant heat transfer equation* is a very simple model of this process where the earth is approximated to be a single point and one considers an average outgoing energy [29, 42]. A more sophisticated approach is the coupling of *atmosphere-ocean-sea ice models*, which takes many more processes into account [42]. While the second model will yield results that are way more accurate, the first one will have a much shorter runtime, is easier to understand and may still show a trend for the change of temperature. Hence, there is always a trade-off between the accuracy of a model and its computational cost.

The different arising models can be classified in various types. They may include linear or nonlinear operators, be stationary or dynamic, discrete or continuous, deterministic or probabilistic or of many other kinds. Also, they can either be used to predict the behaviour of a system or to deduce the cause of an observed effect from a system. Models that are used with the intention of predicting an outcome are titled *forward models* and consist of a model with fully known parameters (see Figure 1).

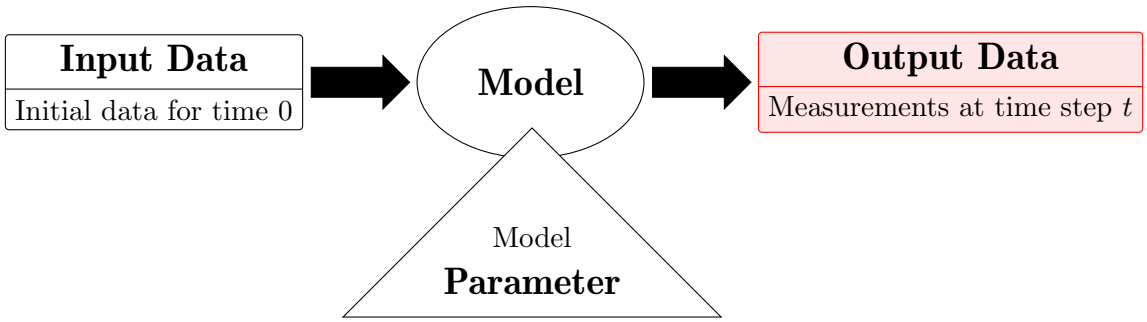


Figure 1: Illustration of a forward model. The input data, the model and its parameters are known, whereas we are interested in the output data.

An example of a forward model would be to predict the shadow cast by an object, while knowing its shape and the position of the light source. The mathematical model then consists of the equations that describe the propagation of light and the parameters put into the model would be the shape of the object blocking the light and the position of the light source. For every forward model one can formulate the *inverse problem* where one is interested in deducing the right parameters of a model in order to be able to reproduce the observed outcome with it (see Figure 2). For the example above, the inverse problem would be to reconstruct the shape of the object with the equations of light propagation by knowing the position of the light source and the shape of the shadow. With this rather simple example it already becomes clear, that even if the forward problem may have a unique solution, this must not hold true for the related inverse problem.

Within each model lies uncertainty, either in the values of the parameter or the accuracy of the

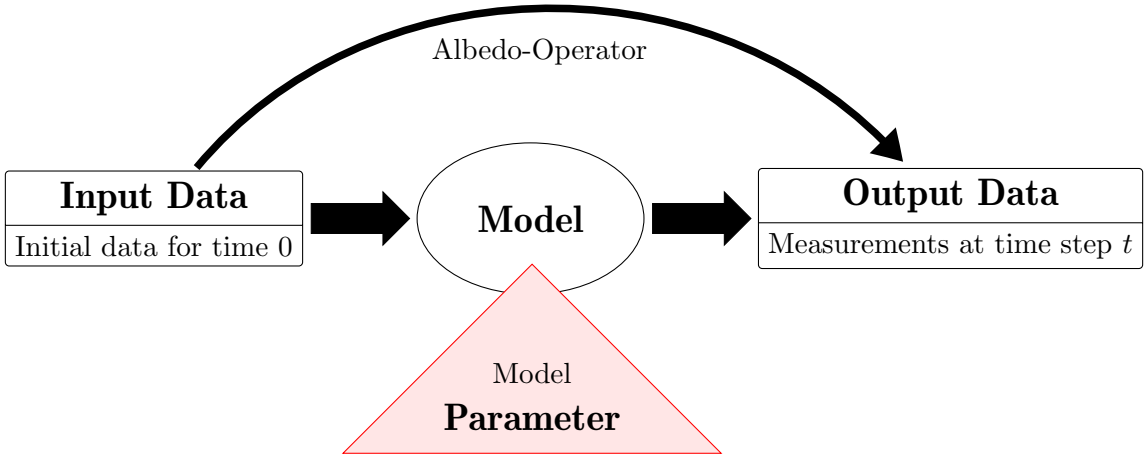


Figure 2: Illustration of an inverse model. The initial data is known as well as the model and the outcome, i.e. the measured data. One is interested in the parameters of the model such that the model would yield the desired output. The Albedo-Operator maps the input data to the output data.

model itself. Additionally, the measured data is equipped with uncertainty as well. Therefore, the field of *uncertainty quantification (UQ)* is of importance, when dealing with inverse problems and experiments.

> "UQ studies all sources of error and uncertainty, including the following: systematic and stochastic measurement error; ignorance; limitations of theoretical models; limitations of numerical representations of those models; limitations of the accuracy and reliability of computations, approximations, and algorithms; and human error. A more precise definition is UQ is the end-to-end study of the reliability of scientific inferences." [40]

While UQ does not answer the questions, whether or not a model is right, it gives information about the consistency of a model. Hence, one can see it as a tool for model verification. Including the mentioned uncertainties into the forward model means that one acknowledges that the parameters are not fully known but are random variables. If the parameters that are put into the model are random variables then this also applies to the output, this process is called *forward propagation of uncertainty*. The same holds true for the inverse problem: the data one can measure will always have some uncertainty, thus the to-be-reconstructed quantities of interest also are not known fully [37, Chapter 1.1]. This type of uncertainty, which comes from the random nature of physical systems, is often referred to as *Aleatoric Uncertainty*, from the Latin word *alea* meaning *gamble* or *a die*. The uncertainty coming from the lack of knowledge about a system is commonly called *Epistemic Uncertainty*, coming from the Greek word for *knowledge* [37]. The distinction between those two types of uncertainty represents two different ways of looking at probability in a mathematical framework. For example, if one rolls a few dice the outcome can said to be random and thus are of aleatoric nature. On the other hand, one could argue that the outcome is only random because of the missing knowledge about the initial condition of the system, such as the material and geometry of the dice, the wind affecting the roll and so on, making the uncertainty epistemic [37]. The second way of looking at this problem would nowadays be labelled as the *Bayesian Perspective*.

Thomas Bayes (* around 1701, †1761), a mathematician, philosopher and the namesake of Bayesian probability, lived in England and is known to only have published two works during his lifetime, one of philosophical and one of mathematical nature:

- "Divine Benevolence, or an Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creatures"(1731)
- "An Introduction to the Doctrine of Fluxions: and a Defence of the Mathematicians Against the Objections of the Author of The Analyst" (published anonymously in 1736)

The second publication mainly focused on defending the logical foundation of Isaac Newtons calculus against a bishop, George Berkeley who was also the author of „The Analyst" and it is speculated that Bayes was accepted as a „Fellow of the Royal Society" because of this work [3]. Even though, Thomas Bayes wrote down his solution to an inverse problem in „An Essay towards solving a Problem in Doctrine of Chances", which was presented to the Royal Society after his death [8], it was Pierre-Simon LaPlace, who made Bayesian probability popular [3].

The Bayesian interpretation of probability focuses on the strength of beliefs and hypotheses. In that sense, Bayes' theorem gives a method to update probabilities based on prior knowledge and data. This means, that the data is assumed to be fixed and rather the hypotheses (for example the values of parameters of a model) are assumed to be distributed according to some probability distribution. In contrast to that, there exists the frequentist view on probability, where one assumes a null hypothesis to be true and the data to be random, meaning a data set is seen as one possibility out of many. Then, it can be investigated how likely the data may appear given the true null hypothesis. If this likelihood is very small, then the null hypothesis may be rejected. Thus, opposing to the Bayesian framework where one takes data and a prior belief into account, the frequentist analysis only relies on data. Still, the results can be manipulated by the choice of the acceptance-probability, by which a hypothesis is accepted or rejected. Now, it could also be argued that one manipulates the result in the Bayesian setting, by putting in certain prior believes in order to get the desired answer. Especially, when the outcome relies heavily on the prior, the choice of the prior, which not always comes that easy, is important. This can be avoided by choosing an „uninformative" prior or by having multiple priors with different beliefs. But the prior knowledge may also come from other experiments, history or intuition on how the solution may or may not look like, thus incorporating many different kinds of information, which could not be done with the frequentist approach. Overall, both approaches have strengths and may be useful for different kinds of problems. Nevertheless, it is always important to make clear how the used method works in order to not manipulate the results. [14]

Overview. Within this thesis the goal is to apply Bayesian inference to the inverse problem coming from chemotaxis of bacteria including a reaction term in order to reconstruct the tumbling kernel from the organism population. Here, two mathematical models on different scales - namely a kinetic and a macroscopic model - are considered and their convergence in the asymptotic limit is shown. The work is organized as follows: in section 2 we will introduce Bayesian inference by first linking it to the classical approach. Then, well-posedness in the Bayesian setting will be discussed and the ideas behind selected numerical applications will be explained. This will serve as the foundation for formulating the inverse problem for chemotaxis. In section 3 the forward models on the kinetic and macroscopic level are established and the asymptotic limit between those two models is presented. The results of [10], where the drift-diffusion limit for the scaled chemotaxis equation is carried out, are extended to the case where a reaction term occurs in the equations. This will form the basis to connect the related inverse problems in section 4. Here, the inverse problems are formulated and the relation between them is investigated. The well-posedness and convergence of the resulting posterior distributions in the asymptotic limit are shown, which can be seen as an extension of the results in [17]. In section 5 the loss of information in the asymptotic limit is investigated on the local and global level. This can be viewed as an attempt to adapt the results in [28] to the non-stationary setting. But the presented approaches do not suffice to quantify the loss of information for the inverse chemotaxis problem in the diffusive limit. A summary and an outlook on further research is given in section 6.

2 Bayesian Inference

In this section, the concept behind Bayesian Inference will be explained. The Bayesian approach to regularization of ill-posed problems will be outlined and its advantages and disadvantages are described. Additionally, some algorithms will be discussed in order to give an idea on how to extract information from Bayesian inversion numerically. We will mainly follow the elaborations in [36].

2.1 Linking the Bayesian and Classical Approach

When conducting an experiment the goal is to derive the input parameters of a model from the measured data. Such *inverse problems* are in general ill-posed in the sense of Hadamard, meaning the solution either lacks existence or uniqueness or does not depend continuously on the data. Mathematically, one can describe an experiment by a function relating an input to an output, as explained for example in [37, Chapter 6.1]. For two Banach spaces $\mathcal{U}, \mathcal{Y}$ we define $H : \mathcal{U} \to \mathcal{Y}$ to be the model operator and call $u \in \mathcal{U}$ the input and $y \in \mathcal{Y}$ the data. Having an inverse problem then corresponds to be interested in u for some given y and H. In reality, the measured data is disturbed by some noise such as

$$y = H(u) + \eta$$

where η represents the uncertainty. If the problem is ill-posed, which is typical for inverse problems, an ordinary approach would be to seek a least-square solution by solving

$$\underset{u \in \mathcal{U}}{\operatorname{argmin}} \|y - H(u)\|_{\mathcal{Y}}^2.$$

Finding the minimum is often difficult because there may not be a limit in $\mathcal{U}$ of the minimizing sequence, it could have multiple minima or it depends sensitively on the data [36]. If the minimum depends sensitively on the data, this means that it does not depend continuously on the data. In this case, one can try to weaken the influence of the data by looking for a solution that does not fit the data too closely. This can be done via regularizing the problem by seeking the minimizer $\hat{u}$ of

$$J(u) := \frac{1}{2}\|y - H(u)\|_{Q^{-1}}^2 + \frac{1}{2}\|u - \bar{u}\|_{R^{-1}}$$

for a chosen self-adjoint operator Q, a point $\bar{u} \in \mathbb{R}^n$ and an operator R describing the structure of the regularization, as shown in [36]. By the choice of Q, different components of the data can be weighted. The choice of $\bar{u}$ and R encodes the prior belief one has about the solution of the problem. For example, if H is a linear function, R could be a positive-definite *Tikhonov matrix*. Tikhonov-Regularization, also known as *Ridge-Regression*, is a common approach for solving ill-posed inverse problems, as explained in e.g. [23]. In the case, where R is chosen as a scalar multiple of the identity-matrix, the regularization is called L^2-regularization [32]. Now, the minimizer $\hat{u}$ either lies close to $\bar{u}$ and is influenced more by the regularization term or it lies closer to the minimizer of the quadratic cost function and is thus influenced more by the data y. Still, without further assumptions about the model, the choice of regularization and weight operators as well as the point $\bar{u}$ are rather arbitrary [36, Chapter 2.2].

When introducing a statistical framework, those assumptions can be made more explicit. In Bayesian statistics the data and quantities of interest are considered to be random variables and are distributed according to some probability distribution The knowledge one has about the quantity of interest before some evidence is taken into account, is represented by a probability distribution for all values of the to-be-reconstructed parameter we believe are possible. It is called the *prior distribution* and corresponds to the regularization term discussed before. For example, one could assume the random variable to be normally distributed. Further, the *data likelihood distribution* encodes the likelihood of measuring the data for some given input and model. Thus, it gives an estimation of

how likely it would be to measure this data under the chosen model and input. Additionally, one has a probability distribution, often referred to as the *normalization constant* which estimates the likelihood of measuring the data without any given conditions. It is called a constant, since it does not depend on the to-be-reconstructed parameter. Solving the inverse problem in the Bayesian setting means to reconstruct the *posterior distribution*, which is the distribution of possible values for the quantity of interest after taking additional information into account. It updates our beliefs about the input based on the data, using Bayes theorem [36, Chapter 6].

Theorem 2.1 (Bayes' rule) *Let $(\Theta, \mathcal{F}, \mu)$ be a probability space, $u, y \in \mathcal{F}$ with $\mu(u), \mu(y) > 0$. Let $\mu(u)$ be the prior distribution, $\mu(y)$ the normalization constant, $\mu(u \mid y)$ the posterior distribution and $\mu(y \mid u)$ the likelihood. Then it holds that*

$$\mu(u \mid y) = \frac{\mu(y \mid u)\mu(u)}{\mu(y)}.$$

In other words, the theorem states that the posterior distribution is proportional to the likelihood of measuring the data y for given inputs u. Now, finding the minimizer of $u \mapsto \frac{1}{2}\|H(u) - y\|^2_{Q^{-1}}$ corresponds to finding the maximum likelihood estimator of u given the data y. Additionally, the Bayesian approach not only yields the most probable value but the distribution of all possible values for u.

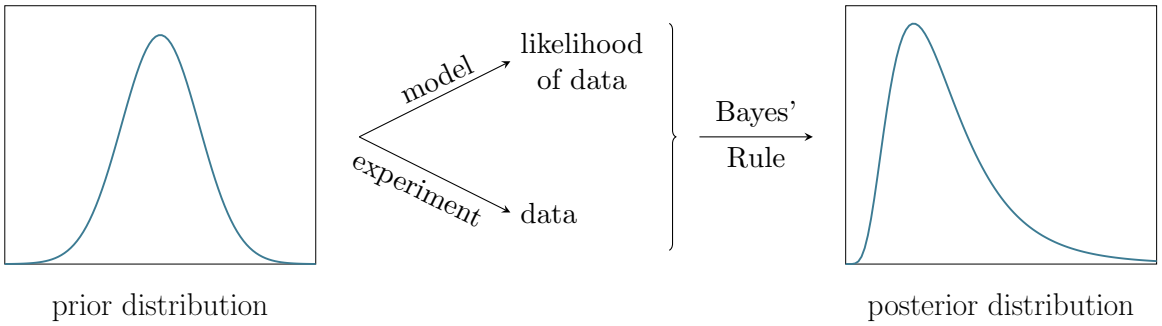


Figure 3: Illustration of Bayes Theorem. The prior distribution encodes the knowledge of the quantity of interest before any evidence is taken into account. Here, a normal distribution is chosen. From the experiment we then get the measurements and can calculate the probability of measuring this set of data without any given conditions and we get the likelihood of the measured data under the given model. With Bayes' rule the posterior distribution can be calculated.

Example 2.2 Following [36, Chapter 2.2] the example of $\mathcal{U}, \mathcal{Y}$ being finite-dimensional spaces will be introduced. Let $u \in \mathbb{R}^p = \mathcal{U}$ and $y \in \mathbb{R}^q = \mathcal{Y}$. Further, we assume the noise η to be a random variable with distribution ν and denote the prior by μ_0. Recalling the equation

$$y = \mathcal{H}(u) + \eta, \tag{2.1}$$

the probability of y given u - the likelihood of the data for a given model - can then be described by the distribution

$$\mu(y \mid u) := \nu(y - H(u)). \tag{2.2}$$

By Bayes' rule it follows that the posterior distribution is proportional to the probability of y given u

$$\mu(u \mid y) \propto \nu(y - H(u))\mu_0 \tag{2.3}$$

where the constant of proportionality depends only on the data y. We still refer to it as a constant, since it does not depend on u. The posterior distribution will be denoted by μ^y in the sense that

$\mu^y(u) = \mu(u \mid y)$. Then, the relation above can also be expressed by the Radon-Nikodym derivative via

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \nu(y - H(u)) \tag{2.4}$$

as stated in [36, Theorem 6.2], which is a more abstract interpretation of Bayes' Theorem. Since ν is a density and thus a non-negative function, one can write without loss of information

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi(u; y)) \tag{2.5}$$

for a potential $\Phi(u; y)$. With this formulation, the generalization to $\mathcal{U}, \mathcal{Y}$ being infinite dimensional can be done in a natural way [36]. From now, we will denote a variable v that is normally distributed with mean m and variance γ by $v = \mathcal{N}(m, \gamma)$. If the noise is Gaussian $\nu = \mathcal{N}(0, \gamma)$ as well as the prior distribution $\mu_0 = \mathcal{N}(\bar{u}, C)$, relation (2.3) yields

$$\mu^y(u) \propto \exp\left(-\frac{1}{2}|y - H(u)|_\gamma^2 - \frac{1}{2}|u - \bar{u}|_C^2\right) \tag{2.6}$$

and the maximum a posteriori estimator, i.e. the estimator at which the posterior distribution has its maximum, then corresponds to

$$\mathrm{argmin}_{u \in \mathbb{R}^p}\left(\frac{1}{2}|y - H(u)|_\gamma^2 + \frac{1}{2}|u - \bar{u}|_C^2\right), \tag{2.7}$$

the solution of a weighted least square problem [36]. Here, the choice of C and $\bar{u}$ encode the prior information and thus it is known where they come from. The posterior distribution is in general not Gaussian, except when the Operator H is linear. How to identify its mean and covariance is shown, for example in [36, Theorem 6.20, Lemma 6.21].

Remark 2.3 In high dimensions it may become difficult to get information from a probability distribution, thus one has to come up with different methods in order to obtain information. One idea is, to look at the *maximum a posteriori (MAP) estimator*. At this point, the posterior distribution has its maximum and thus its the value with the highest probability. Another approach from the field of numerics is to use *sampling methods*, such has *Markov chain Monte Carlo (MCMC)* methods. The idea behind those methods is to generate a set of points $\{u_i\}_{i \in I}$ that are (approximately) distributed according to $\mu^y(u)$, as discussed in subsection 2.3. Here, Bayes' theorem is useful, since it provides cases where the posterior function is known up to a constant, as illustrated in the example above in (2.5).

Remark 2.4 (Influence of Prior Distribution on Uncertainty of the Solution) A natural question is, if Bayesian inference yields the correct result, regardless of the prior, when exposed to enough sample data. To investigate the influence of the prior distribution on our solution, we are interested in the small observational noise limit. For example, in the case of a linear model where the posterior distribution is Gaussian and its mean and covariance can be explicitly determined, one can look at the underlying noise-free problem. If the problem is underdetermined, the mean of the posterior is determined by the prior and thus there remains uncertainty, even if the observational noise disappears. This implies, that the prior plays a huge role, even if we have small observational noise. Especially, the uncertainty is dependent on the choice of prior. In the case of an overdetermined system, the posterior distribution converges to a Dirac measure, meaning that the uncertainty disappears. In other words, the prior plays no role in this limit. Examples of such problems can be found in [36, Example 2.1-2.2]. In reality, many of the inverse problems are underdetermined, meaning that the assumptions we put into the model by choosing a prior are important. One advantage of the Bayesian approach is, that those assumptions are at least made clear and precise. [36, Chapter 2.2-2.3].

2.2 Bayesian Well-Posedness

This section will discuss the question of stability of distributions to perturbations. In order to use Bayes' Rule for an inverse problem, we need to ensure that the posterior distribution is well-defined, the normalization constant is indeed bounded and that the distributions depend continuously on the data. We allow the data to be in form of a function, the case of discrete data will be discussed later on. In order to get the desired results, the potential $\Phi(u;y)$ introduced in Example 2.2 is assumed to have the following properties [36, Assumptions 2.6]:

(A1) $\forall \varepsilon > 0,\ r > 0\ \exists M = M(\varepsilon, r) \in \mathbb{R}\colon\ \forall u \in \mathcal{U},\ y \in \mathcal{Y} \mid \|y\|_{\mathcal{Y}} < r$:

$$\Phi(u;y) \geq M - \varepsilon \|u\|_{\mathcal{U}}^2$$

(A2) $\forall r > 0\ \exists K = K(r) > 0\colon\ \forall u \in \mathcal{U},\ y \in \mathcal{Y} \mid \|u\|_{\mathcal{U}}, \|y\|_{\mathcal{Y}} < r$:

$$\Phi(u;y) \leq K$$

(A3) $\forall r > 0\ \exists L = L(r) > 0\colon\ \forall u_1, u_2 \in \mathcal{U},\ y \in \mathcal{Y} \mid \|u_1\|_{\mathcal{U}}, \|u_2\|_{\mathcal{U}}, \|y\|_{\mathcal{Y}} < r$:

$$|\Phi(u_1;y) - \Phi(u_2;y)| \leq L\|u_1 - u_2\|_{\mathcal{U}}$$

(A4) $\forall \varepsilon > 0,\ r > 0\ \exists C = C(\varepsilon, r) > 0\colon\ \forall u \in \mathcal{U},\ y_1, y_2 \in \mathcal{Y} \mid \|y_1\|_{\mathcal{Y}}, \|y_2\|_{\mathcal{Y}} < r$:

$$|\Phi(u;y_1) - \Phi(u;y_2)| \leq \exp(\varepsilon\|u\|_{\mathcal{U}}^2 + C)\|y_1 - y_2\|_{\mathcal{Y}}$$

The first assumption guarantees that the potential does not decay at $-\infty$ too quickly, the second one gives the boundedness on bounded sets and the third and fourth encode the Lipschitz-continuity in both arguments.

From [36, Chapter 4.2] we cite the following two theorems.

Theorem 2.5 *Let Φ satisfy our standard assumptions A1, A2 and A3. Let μ_0 be a Gaussian probability distribution on $\mathcal{U}$. Then, for each $y \in \mathcal{Y}$ the measure μ^y is given by*

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{\exp(-\Phi(u;y))}{Z(y)}$$
$$Z(y) = \int_{\mathcal{U}} \exp(-\Phi(u;y))\,\mathrm{d}\mu_0(u)$$

and is a well-defined probability distribution on $\mathcal{U}$.

For a Gaussian prior, assumptions A1-A3 yield a well-defined posterior distribution for each possible instance of the observed data. To quantify the behaviour of the posterior distribution with respect to changes in data, we introduce the *Hellinger metric*

$$d_{\mathrm{Hell}}(\mu_1, \mu_2) := \frac{1}{2}\int_{\mathcal{A}} \left(\sqrt{\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}(u)} - \sqrt{\frac{\mathrm{d}\mu_2}{\mathrm{d}\mu_0}(u)} \right)^2 \mathrm{d}\mu_0(\mathrm{u}). \tag{2.8}$$

The metric is discussed in more detail in subsection 4.2.

Theorem 2.6 *Let Φ satisfy assumptions A1, A2, A4 and suppose that μ_0 is a Gaussian probability distribution on $\mathcal{U}$. Let μ^y be absolutely continuous with respect to μ_0 with density given by Bayes' rule for each $y \in \mathcal{Y}$. Then there exists a constant $C > 0$ such that for all $y, y' \in \mathcal{Y}$*

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y'}) \leq C\|y - y'\|_{\mathcal{Y}}.$$

This theorem establishes the stability of the posterior distribution with respect to the data under the Hellinger metric.

2.3 Numerical Applications

In this section, the main ideas behind different kind of algorithms will be introduced with which information can be drawn from Bayes' formula for the posterior distribution. The distribution can be computed in a closed form only in rare cases, as discussed in the preceding sections. One example of such a case is the inverse problem for the heat equation, where the prior is Gaussian and the potential quadratic, leading to a Gaussian posterior distribution [36, Chapter 3.5]. Since the posterior distribution can not be described that explicitly in general, approximations are required to extract information from Bayes' rule [36, Chapter 5.1]. One sampling algorithm, which enjoys a high reputation among scientists, is the *Markov chain Monte Carlo* method. While this is a very powerful tool, it is also costly to implement which is why *variational* and *filtering* methods will be discussed as well.

2.3.1 Markov Chain Monte Carlo Algorithm

The key idea behind the class of Monte Carlo algorithms is to sample from the distribution one wants to reconstruct. This distribution, in our case the posterior distribution, will be called target distribution. Instead of deriving for instance the mean of a distribution from its equation, the Monte Carlo approach is to draw a large number of samples of the distribution and calculate the sample mean. This is especially useful if random samples are easy to draw and the equations are difficult to deal with [41]. It remains to explain how the samples are generated and that is where the Markov Chain comes in. The algorithm starts by guessing a value which may be appropriate for the target distribution. The next sample will be proposed depending on the previous one by a special process, for example by *Metropolis Hasting* methods, where the next sample is taken from a normal distribution. The algorithm has the Markov property, since generating a new sample only depends on the previous step and no step before that. The function of proposing the new sample is also called the Markov transition kernel. The proposed value will then be accepted or neglected by some probability which depends on the target distribution. Analytically, the posterior distribution is often difficult to calculate due to the normalization constant in the denominator in Bayes' rule. The concept behind MCMC-algorithms is to look at relative probabilities during the process of accepting a new sample, such that the normalization constant gets cancels out. For example, if u_c would be the current sample and u_p the proposed one, one can calculate the acceptance ratio at one position via Bayes' rule

$$\frac{\frac{\mu(y|u_p)\mu(u_p)}{\mu(y)}}{\frac{\mu(y|u_c)\mu(u_c)}{\mu(y)}} = \frac{\mu(y \mid u_p)\mu(u_p)}{\mu(y \mid u_c)\mu(u_c)}$$

which equals the acceptance ratio for the non-normalized posterior distributions [43]. A more detailed construction of the algorithm can be found in [36, Chapter 5.2].

2.3.2 Variational Methods

The goal of variational Bayes methods is to approximate the posterior distribution. Following [25], the accuracy of the distribution is measured via the Kullback-Leibler divergence

$$d_{\mathrm{KL}}(\mu_1, \mu_2) := \int_{\mathcal{A}} \left(\log \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}(u) \right) \mathrm{d}\mu_2(\mathrm{u}) \tag{2.9}$$

which will be discussed in more detail in subsection 4.2. Let ν be the approximate distribution and μ^y the target distribution. Then it follows from Bayes' rule that

$$d_{\mathrm{KL}}(\nu, \mu^y) = \int_{\mathcal{A}} \left(\log \frac{\mathrm{d}\nu}{\mathrm{d}\mu^y}(u) \right) \mathrm{d}\mu^y(u)$$

$$= \int_{\mathcal{A}} \left(\log \frac{\mathrm{d}\nu(u)}{\mathrm{d}(\mu(y \mid u)\mu(u))} \right) \mathrm{d}\mu^y(u) + \int_{\mathcal{A}} (\log \mathrm{d}\mu(y)) \mathrm{d}\mu^y(u)$$
$$= \int_{\mathcal{A}} \left(\log \frac{\mathrm{d}\nu(u)}{\mathrm{d}(\mu(y \mid u)\mu(u))} \right) \mathrm{d}\mu^y(u) + \log \mathrm{d}\mu(y)$$

or in other words

$$\log \mathrm{d}\mu(y) = d_{\mathrm{KL}}(\nu, \mu^y) + \mathcal{V}(\nu).$$

Since the left hand side of this expression is constant, maximizing $\mathcal{V}(\nu)$ means minimizing the Kullback-Leibler divergence of the approximation and the posterior distribution, hence yielding a better approximation [25]. Overall, neither the posterior distribution nor the normalization constant needs to be calculated. Since now the goal is to find a function that maximizes an operator, namely $\mathcal{V}$, we are in the field of variational analysis leading to the name of this class of methods. In contrast to the MCMC-methods, variational methods are faster but a little less accurate.

2.3.3 Filtering Methods

Similarly to the variational methods, the goal of the filtering methods is to approximate the posterior distribution. This is done by sequentially updating a sequence of probability distributions in time. This type of updating arises naturally in the sense that during an experiment one acquires more and more data and thus wishes to be able to update the solution accordingly. Also, it is efficient to use sequential updating since it reduces the dimension of the target distribution. When assuming an underlying Markov structure, a correlated distribution at several time steps can be split up into several, conditionally independent distributions at each time step [36, Chapter 5.1]. One key idea is to break up the unknown parameter as well and deduce information about it sequentially and conclude a recursive form of Bayes' rule [36, Chapter 5.4].

Following [16, Chapter 3.3] we denote the vector of observations by $y_{1:n} = \{y_1, \ldots, y_n\}$, the state of the system at time t_n by u_n and write Bayes' rule as

$$\mu(u_n \mid y_{1:n}) = \frac{\mu(y_{1:n} \mid u_n)\mu(u_n)}{\mu(y_{1:n})}.$$

Since it holds for the set of observations that $\mu(y_{1:n}) = \mu(y_n \mid y_{1:n-1})\mu(y_{1:n-1})$ [16, Equation (3.3)], Bayes' formula reduces to

$$\mu(u_n \mid y_{1:n}) = \frac{\mu(y_n \mid u_n)\mu(u_n \mid y_{1:n-1})}{\mu(y_n \mid y_{1:n-1})}$$

where we also used the Markov property [16, Equation (3.19)-(3.23)]. In order to get the recursive formula, one then applies the *Chapman-Kolmogorov-equation*, that links the prior distribution and the previous posterior distribution

$$\mu(u_n \mid y_{1:n-1}) = \int \mu(u_n \mid u_{n-1})\mu(u_{n-1} \mid y_{1:n-1}) \mathrm{d}u_{n-1}.$$

With this recursive process, the algorithm approximates the posterior distribution. Further explanations about the processes can be found in [16,36]. Overall, the algorithm can update a new prediction when more data becomes available, but it cannot update old predictions. Thus, the algorithm is powerful, when interested in forecasts, but not when one wants to reanalyse the system, for example when the goal is to improve the parameters of a model.

3 Chemotaxis: Forward Models

The purpose of this chapter is to first introduce the phenomenon of chemotaxis and the common mathematical forward models used to describe the motion of bacteria on the kinetic and macroscopic level, introduced for example in [34]. The relation between the two arising models will be discussed and some results from the current state of research will be outlined, especially the results from [10], where the drift-diffusion limit for the scaled chemotaxis euqation is carried out. The second part of this section deals with the extension of the results in [10] with a reaction term that encodes the birth and death of bacteria, which is not always included in the commonly used models.

The phenomenon of bacterial movement as a response to a chemical attractant or repellent is called chemotaxis. The mechanism, how the movement occurs, may differ between different species. Some bacteria, like *Myxobacteria*, can pull themselves forward or push themselves back by extending a long hair they have in front of them, called *pilus*. In contrast to that, other bacteria, such as *E. Coli*, move by rotating their *flagella* that act like „propellers“, which they have several of at each cell. By rotating the flagella counter-clockwise, the bacterium can move in a straight line, by rotating them clockwise they start a tumbling motion which can result in a change of direction. This change of direction is not deterministic but involves a random process [34]. The two types of movement are illustrated in Figure 4.

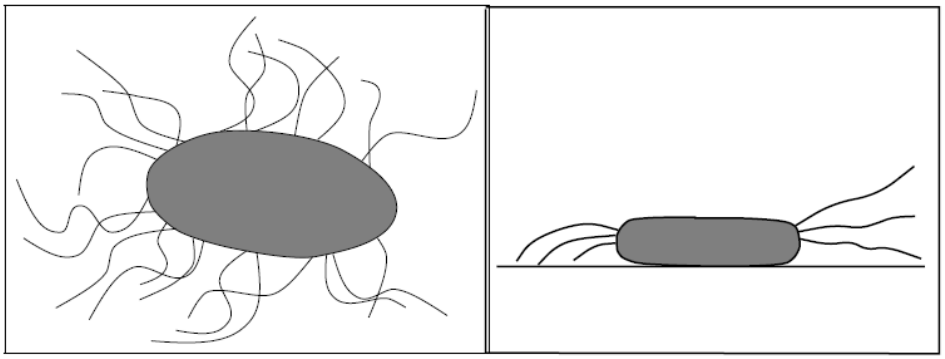


Figure 4: An Illustration of bacterium *E. Coli* and its flagella on the left, and of *Myxobacteria* with its pilus on the right [34, Figure 5.1].

3.1 Kinetic and Macroscopic Description

Mathematically speaking, there a three main models to describe chemotaxis, one is of macroscopic, one of microscopic and one of kinetic nature. In this work, we focus on the macroscopic and kinetic formulation. The most famous model on the macroscopic population level is the Keller-Segel system

$$\begin{aligned} \partial_t \rho(t,x) - \nabla(D(t,x;c)\nabla\rho(t,x)) + \nabla(\rho(t,x)\Gamma(t,x;c)) &= 0 \\ \partial_t c(t,x) - D_c(t,x)\Delta c(t,x) &= \phi(c,\rho) \end{aligned} \tag{3.1}$$

with $D(t,x;c) \in \mathbb{R}^{3\times 3}$ being the diffusion matrix of the cells, $D_c(t,x) \in \mathbb{R}^3$ the diffusion matrix of the chemoattractant and $\Gamma(t,x;c) \in \mathbb{R}^3$ the drift vector, which is introduced in [10, 34]. The first equation traces the evolution of the bacteria density $\rho(t,x)$ at time $t > 0$ and point $x \in \mathbb{R}^3$ when a chemical stimulant is introduced. The second equation describes the diffusion of the chemical stimulus described by the density function $c : (t,x) \mapsto c(t,x)$. Both, the diffusion matrix and the drift vector are determined by the density function c of the chemoattractant. The function $\phi(c,\rho)$ describes the interaction between the chemoattractant and the bacteria, for example if the bacteria produce the chemoattractant. During this work the concentration of the chemoattractant is assumed

to be a known function and is neither consumed or produced by the bacteria, thus reducing the Keller-Segel system to one equation

$$\frac{\partial}{\partial t}\rho(t,x) - \nabla(D(t,x;c)\nabla\rho(t,x)) + \nabla(\rho(t,x)\Gamma(t,x;c)) = 0. \tag{3.2}$$

The Keller-Segel system has been widely studied as outlined in [20,21]. We want to highlight some of the properties because of which the mathematics of the system is interesting and refer to [34, Chapter 5] for a more detailed explanation.

Remark 3.1 (Properties of the Keller-Segel System) In dimension 1 there exist unique and global smooth solutions to the system (3.1) and for small initial mass, it is well-posed globally in time. The solution of the system remains non-negative and the total mass is conserved. For large initial mass blow-up of the coupled system occurs if one does not consider a diffusion coefficient in the equation, meaning that the solution does not remain bounded. For dimension 2 the situation can be made more precise, since there one can calculate a critical mass under which no blow-up occurs. The blow-up scenario can be prevented, for example by using a non-linear diffusion matrix [22] or nonlinear drift vector [26]. A more detailed explanation can also be found in [34, Chapter 5.5.1].

While this model is macroscopic in nature, it overlooks the reaction of the individual bacteria to the chemoattractant and is thus inaccurate in certain regimes. This leads us to the mesoscopic model, the kinetic chemotaxis equation, which we will consider for bacteria that move by rotating their flagella. As mentioned above, the movement of these bacteria consists of two parts: one where they run in a straight line with some velocity v' and one where they tumble and change directions from v' to v as shown in Figure 5.

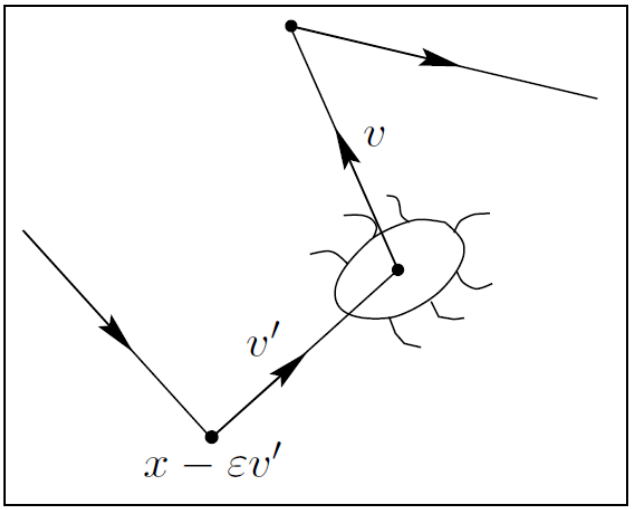


Figure 5: Illustration of run and tumble motion for *E. Coli* [34, Figure 5.6].

The according model, introduced in [4,10], is termed to be a run-and-tumble model of the form

$$\partial_t f(t,x,v) + v \cdot \nabla_x f(t,x,v) = \mathcal{K}(f;c) \tag{3.3}$$

where the function

$$\mathcal{K}(f;c) = \int_V K(t,x,v,v';c)f(t,x,v') - K(t,x,v',v;c)f(t,x,v)\mathrm{d}v'$$

encodes the tumbling motion. The function $f(t,x,v)$ describes the density of bacteria at a point $x \in \mathbb{R}^3$ and time $t > 0$ with a certain velocity $v \in V \subset \mathbb{R}^3$. For V it comes naturally to choose a sphere

in $\mathbb{R}^3$, so we will choose the unit sphere $V = \mathbb{S}^1$. The density function f is solely a mathematical entity, since it is difficult to be measured physically. By $K(t,x,v,v';c)$ the probability of bacteria changing velocity from v' to v is denoted and its shape is determined by the chemoattractants density function. In order to simplify the notation, the abbreviations $f' = f(t,x,v')$ and $K' = K(t,x,v',v;c)$ are introduced. With this, the right hand side of (3.3) looks like

$$\mathcal{K}(f;c) = \int_V \big(Kf' - K'f\big)\mathrm{d}v'.$$

We cite the following theorem about existence from [10, Theorem 1].

Theorem 3.2 (Boundedness and Existence) *Consider the Sobolev-space $W^{1,\infty}(\mathbb{R}^3)$. Assume $f_0 \in L^1_+ \cap L^\infty(\mathbb{R}^3 \times V)$ and that there exists $C > 0$ such that $\forall x \in \mathbb{R}^3$, $v, v' \in V$, $t > 0$ and $c \in W^{1,\infty}(\mathbb{R}^3)$ it holds that*

$$0 \le K(x,v,v',t;c) \le C\big(1 + c(t,x+v) + c(t,x-v')\big). \tag{3.4}$$

Then there exists a global solution $f \in L^\infty\big((0,\infty); L^1_+ \cap L^\infty(\mathbb{R}^3 \times V)\big)$ of (3.3) (with $\varepsilon = 1$).

Remark 3.3 If the tumbling kernel depends on the gradient of c, then no global existence result is available, even if we weaken the assumption in the theorem above to

$$0 \le K(x,v,v',t;c) \le C\Big(1 + \|c(t,\cdot)^\beta_{L^\infty(\mathbb{R}^3)}\|\Big)$$

with $\beta \ge 1$. For $\beta \le 1$ a global existence result can be acquired. [10, Chapter 3]

Since we will assume that the chemoattractants concentration is a given function and is neither consumed or produced by the bacteria, the dependence of D, Γ and K on c will be dropped in the notation.

3.2 Introduction of Birth-Death-Term

Now, the two forward models will be extended by adding a birth-death-term. We assume the term on the macroscopic level to be of the form $b(\rho)\rho$, that there exists a saturation density ρ^* and that the rate at which the bacteria reproduce is bounded. The saturation density correlates with the carrying capacity which is the maximum population size. Overall, the birth-death-term will be described by the Fisher-term [30]

$$\mathcal{R}(\rho) = r(1 - \frac{\rho}{c})\rho = b(\rho)\rho$$

with r being the intrinsic growth rate and c the carrying capacity.

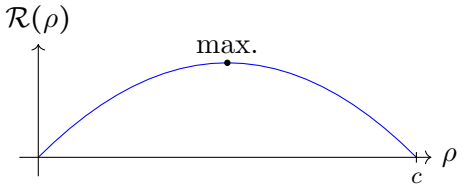


Figure 6: Illustration of the Fisher-term.

On the macroscopic level this results in the Keller-Segel-Fisher-equation

$$\begin{aligned} \partial_t\rho - \nabla\cdot(D\cdot\nabla\rho) + \nabla(\rho\Gamma) &= \mathcal{R}(\rho) \\ \rho(0,x) = \rho_0(x) &= \int_V f_0(x,v)\mathrm{d}v. \end{aligned} \tag{3.5}$$

On the mesoscopic level we expect that the birth-death term is of the form $b(\rho)f$ following [33] for some function $b(\rho)$. Thus, within the kinetic framework the Fisher-term reads

$$\mathcal{R}(f) = r\left(1 - \frac{\int_V f \mathrm{d}v}{c}\right) f = r(1 - \frac{\rho}{c}) f$$

and results in

$$\begin{aligned} \partial_t f + v \cdot \nabla_x f &= \mathcal{K}(f) + \mathcal{R}(f) \\ f_0(x, v) &= f(0, x, v). \end{aligned} \tag{3.6}$$

The relation $\int_V f(t, x, v)\mathrm{d}v = \rho(t, x)$ will be made more clear in the next section.

In the following we will use that we have an upper bound for the norm of $b(\rho)$, either it is bounded by r, 1 or $|1 - \frac{\rho^*}{c}|$, where ρ^* denotes the saturation density. The bound will be denoted by b.

3.3 Drift-Diffusion Limit

In this section, the asymptotic relation between the mesoscopic and macroscopic models is examined. The results of [10] and [17] will be extended to the case where we have an additional birth-death term. The goal is to show that (3.5) and (3.6) are asymptotically equivalent in the long-time large-space regime. By ε we denote a scaling parameter and get for a parabolic scaling the chemotaxis equation

$$\begin{aligned} \varepsilon^2 \partial_t f_\varepsilon(t, x, v) + \varepsilon v \cdot \nabla_x f_\varepsilon(t, x, v) =& \mathcal{K}_\varepsilon(f_\varepsilon) + \varepsilon^2 \mathcal{R}_\varepsilon(f_\varepsilon) \\ :=& \int_V \big(K_\varepsilon(t, x, v, v') f_\varepsilon(t, x, v) - K_\varepsilon(t, x, v', v) f_\varepsilon(t, x, v)\big)\mathrm{d}v' \\ &+ \varepsilon^2 r\left(1 - \frac{\rho_\varepsilon(t, x)}{c}\right) f_\varepsilon(t, x, v) \\ f_\varepsilon(0, x, v) =& f_0(x, v). \end{aligned} \tag{3.7}$$

The idea behind the parabolic scaling is to scale the time-processes with ε^2 and the processes in space with ε. The reproduction term is scaled by ε^2 since it is a process in time. Formally, when $\varepsilon \to 0$ the tumbling kernel dominates the equation and we expect that

$$f_\varepsilon \to f_* \qquad \mathcal{K}_*(f_*) = 0$$

where here $\mathcal{K}_*$ denotes the limiting operator for $\mathcal{K}_\varepsilon$. This means that the limiting solution lies almost in the null space of the operator $\mathcal{K}_*$ [17]. It is shown in [10, Chap. 2] that in absence of a birth-death term the null space of $\mathcal{K}_*$ can be assumed to be one dimensional.

Now we want to discuss this assumption a bit further and for this we assume to only have tumbling kernels of the form

$$K_\varepsilon = K_0 + \varepsilon K_1 + \mathcal{O}(\varepsilon^2)$$

and will omit any terms of a higher order in ε. This is done, since our goal is to compare the tumbling kernel for the kinetic and macroscopic model and higher order terms would not affect the macroscopic equation [10, 17]. Therefore, it would not be possible to reconstruct those from the macroscopic inverse model [17]. Thus, we have that $\mathcal{K}_* = \mathcal{K}_0$ is the limiting operator. We cite the following assumption from [10, Assumption (A0)] and apply it to our case.

(A0) There exists a bounded velocity distribution $F(v) > 0$, independent of t, x and the chemoattractant c, such that the detailed balance $K_0(t, x, v, v')F(v') = K_0(t, x, v', v)F(v)$ holds. The flow produced by this equilibrium distribution vanishes and F is normalized, meaning

$$\int_V vF(v)\mathrm{d}v = 0, \quad \int_V F(v)\mathrm{d}v = 1. \tag{3.8}$$

The turning rate K_0 is bounded, and there exists a constant $\gamma = \gamma(c) > 0$ such that $K_0/F \geq \gamma, \forall (v, v') \in V \times V, x \in \mathbb{R}^3$ and $t > 0$.

Hence, we write the null-space of $\mathcal{K}_0$ as

$$\mathcal{N}(\mathcal{K}_0) = \left\{ \alpha F \mid \alpha \in \mathbb{R}, F > \delta > 0, \int_V F \mathrm{d}v = 1, \int_V vF \mathrm{d}v = 0 \right\} \tag{3.9}$$

and $f_* \in \mathcal{N}(\mathcal{K}_0)$ is of the form $\rho(t,x)F(v)$ for a density function $\rho(t,x)$. The function F is often called *local equilibrium*, for example in [17]. Because of assumption (A0) the function F is a function only of v. Following [10], inserting the formula into the Keller-Segel-Fisher equation and carrying out the asymptotic expansion shows that ρ fulfils

$$\begin{aligned} \partial_t \rho - \nabla \cdot (D \cdot \nabla \rho) + \nabla(\rho \Gamma) &= \mathcal{R}(\rho) \\ \rho(0,x) = \rho_0(x) &= \int_V f_0(x,v) \mathrm{d}v. \end{aligned}$$

In the following, this statement will be justified. In order to do so, (K_0, K_1) is supposed to be an element of

$$\begin{aligned} \mathcal{A} = \big\{ (K_0, K_1) \in C^1([0,\infty) \times \mathbb{R}^3 \times V \times V)^2 \;\big|\; & \|K_0\|_{C^1}, \|K_1\|_{C^1} \leq C, 0 < \alpha \leq K_0 \text{ is symmetric and} \\ & K_1 \text{ antisymmetric in } (v, v') \big\} \end{aligned} \tag{3.10}$$

for some constants $\alpha, C > 0$. Further, we will need the following two results from [10, Lemma 1, Lemma 2].

Lemma 3.4 *Let $\eta : \mathbb{R} \to \mathbb{R}$ and $g : V \to \mathbb{R}$ and denote by*

$$\begin{aligned} \phi_\varepsilon^S &= \frac{K_\varepsilon F' + K_\varepsilon' F}{2}, \\ \phi_\varepsilon^A &= \frac{K_\varepsilon F' - K_\varepsilon' F}{2} \end{aligned}$$

the symmetric and respectively antisymmetric parts of $K_\varepsilon F$. Then it holds that

$$\begin{aligned} \int_V \mathcal{K}_\varepsilon(Fg)\eta(g) \mathrm{d}v = &\frac{1}{2} \int_V \int_V \phi_\varepsilon^S (g - g') \big(\eta(g) - \eta(g')\big) \mathrm{d}v' \, \mathrm{d}v \\ &- \frac{1}{2} \int_V \int_V \phi_\varepsilon^A (g - g') \big(\eta(g) - \eta(g')\big) \mathrm{d}v' \, \mathrm{d}v. \end{aligned}$$

The same holds for $\mathcal{K}_0$ and $\mathcal{K}_1$ with analogous definitions of $\phi_{0,1}^S$ and $\phi_{0,1}^A$.

For the symmetric and antisymmetric parts, we have the expansions

$$\phi_\varepsilon^S = K_0 F' + \mathcal{O}(\varepsilon) \tag{3.11}$$

$$\phi_\varepsilon^A = \varepsilon \frac{K_1 F' - K_1' F}{2} + \mathcal{O}(\varepsilon^2), \tag{3.12}$$

where we used that $K_0 F' = K_0' F$ by assumption (A0), leading to $\phi_0^A = 0$ and $\phi_0^S = K_0 F'$.

Lemma 3.5 *Let (A0) hold. Then the entropy equality*

$$\int_V \mathcal{K}_0(f) \frac{f}{F} \mathrm{d}v = \frac{1}{2} \int_V \int_V \phi_0^S \left(\frac{f}{F} - \frac{f'}{F'}\right)^2 \mathrm{d}v' \mathrm{d}v \geq 0$$

holds. Further, for $g \in L^2(V; \mathrm{d}v/F)$, the equation $\mathcal{K}_0(f) = g$ has a unique solution $f \in L^2(V; \mathrm{d}v/F)$ satisfying $\int_V f \mathrm{d}v = 0$ if and only if $\int_V g \mathrm{d}v = 0$.

Remark 3.6 The entropy equality is a consequence of Lemma 3.4 applied to $g = f/F$ and $\eta = \mathrm{id}$. Further, we note that in (A0) it is assumed that a constant $\gamma > 0$ exists such that $K_0/F \geq \gamma$. Thus, $\phi_0^S \geq \gamma F F'$ and therefore also

$$\int_V \mathcal{K}_0(f)\frac{f}{F}\mathrm{d}v \geq \frac{\gamma}{2}\int_V\int_V FF'\left(\frac{f}{F}-\frac{f}{F}\right)^2\mathrm{d}v'\,\mathrm{d}v = \gamma\int_V \frac{f^2}{F}\mathrm{d}v. \tag{3.13}$$

From this and the Lax-Milgram lemma the statement of Lemma 3.5 follows.

Regarding the birth-death term, the notation

$$\mathcal{R}_\varepsilon(f_\varepsilon) = r\left(1-\frac{\rho_\varepsilon}{c}\right)f_\varepsilon,\quad \mathcal{R}_0(f_0) = r\left(1-\frac{\rho_0}{c}\right)f_0,\quad \mathcal{R}_0(f_1) = r\left(1-\frac{\rho_0}{c}\right)f_1,$$
$$\mathcal{R}_1(f_0) = -\frac{r}{c}\rho_1 f_0 \quad\text{and}\quad \mathcal{R}_1(f_1) = -\frac{r}{c}\rho_1 f_1.$$

is used if it is more convenient to be in the operator notation. Overall, the $\mathcal{R}_0$ operator contains the zeroth order terms of the reaction term and $\mathcal{R}_1$ the first order term which are then multiplied by f. Otherwise, we will write

$$b(\rho_\circ)f_\square = r\left(1-\frac{\rho_\circ}{c}\right)f_\square$$

with $\circ, \square = \varepsilon, 0$ if we want to highlight which ρ is taken into account.

For the proof of the following theorem as well as in other parts of this work, the lemma of Grönwall will be used which is why we recall it here and refer to [38] for a proof.

Lemma 3.7 (Grönwall's inequality) *Suppose $u(t), \alpha(t) : [t_0, T] \to \mathbb{R}$ are continuous functions, $\beta : [t_0, T] \to \mathbb{R}_+$ and the condition*

$$u(t) \leq \alpha(t) + \int_{t_0}^t \beta(s)u(s)\mathrm{d}s \tag{3.14}$$

is satisfied. Then it holds that

$$u(t) \leq \alpha(t) + \int_{t_0}^t \alpha(s)\beta(s)\exp\left(\int_s^t \beta(\sigma)\mathrm{d}\sigma\right)\mathrm{d}s. \tag{3.15}$$

If in addition $\alpha(t)$ is non-decreasing, then

$$u(t) \leq \alpha(t)\exp\left(\int_{t_0}^t \beta(s)\mathrm{d}s\right). \tag{3.16}$$

Taking [17, Theorem 1] as our basis we apply it to the chemotaxis equation including a reaction term. This make the next Theorem 3.8 into a result which extends [17, Theorem 1].

Theorem 3.8 *Suppose that the initial data is smooth in the sense that $f_0 \in C_c^1(\mathbb{R}^3 \times V)$ and that $K_\varepsilon = K_0 + \varepsilon K_1$ with $(K_0, K_1) \in \mathcal{A}$. The solution f_ε of the scaled chemotaxis equation (3.7) satisfies the following:*

(i) The solution f_ε exists for sufficiently small ε and is bounded in $L^\infty([0,T], L^1_+ \cap L^\infty(\mathbb{R}^3 \times V))$ for finite $T < \infty$.

(ii) The solution f_ε converges to $\rho F \in L^\infty([0,T], L^1_+ \cap L^\infty(\mathbb{R}^3 \times V))$ where F lies in the null space $\mathcal{N}(K_0)$ as defined in (3.9) and ρ satisfies the Keller-Segel-Fisher system (3.5) with coefficients

$$D = \int_V v \otimes \kappa(t,x,v)\mathrm{d}v \tag{3.17}$$

$$\Gamma = -\int_V v\Theta(t,x,v)\mathrm{d}v \tag{3.18}$$

where κ and θ fulfil

$$\mathcal{K}_0(\kappa) = vF \quad\text{and}\quad \mathcal{K}_0(\Theta) = \mathcal{K}_1(F).$$

(iii) The boundedness and convergence is uniform in $\mathcal{A}$.

PROOF: (i) First, we show the boundedness of the solution f_ε for ε small enough. The growth-rate of the birth-death term is bounded, hence it is reasonable to assume to have a maximum density, the so-called saturation density f^*. Since we are interested in the density of bacteria, we only consider positive solutions for the chemotaxis equation. It is important to note, that the saturation density is not assumed to be a constant function, but its $L^1(\mathbb{R}^3 \times V)$-norm to be a bound for every f_ε

$$\|f_\varepsilon\|_{L^1(\mathbb{R}^3\times V)} \leq \|f^*\|_{L^1(\mathbb{R}^3\times V)} < \infty.$$

Now, we want to estimate the $L^\infty(\mathbb{R}^3 \times V)$-norm of f_ε. Integrating the scaled chemotaxis equation in time along the characteristic $(t-s, x-\frac{vs}{\varepsilon}, v)$ yields

$$\begin{aligned} f_\varepsilon(t,x,v) &= f_0(x,v) + \int_0^t \mathcal{K}_\varepsilon(f_\varepsilon)(t-s, x-\frac{vs}{\varepsilon}, v)\mathrm{d}s + \int_0^t \Big(b(\rho_\varepsilon) f_\varepsilon(t-s, x-\frac{vs}{\varepsilon}, v)\Big)\mathrm{d}s \\ &\leq f_0(x,v) + \int_0^t \int_V K_\varepsilon(t-s, x-\frac{vs}{\varepsilon}, v, v') f_\varepsilon(t-s, x-\frac{vs}{\varepsilon}, v')\mathrm{d}v'\,\mathrm{d}s \\ &\quad + b\int_0^t f_\varepsilon(t-s, x-\frac{vs}{\varepsilon}, v)\mathrm{d}s \\ &\leq f_0(x,v) + 2C\int_0^t \int_V f_\varepsilon(t-s, x-\frac{vs}{\varepsilon}, v')\mathrm{d}v'\,\mathrm{d}s + b\int_0^t f_\varepsilon(t-s, x-\frac{vs}{\varepsilon}, v)\mathrm{d}s. \quad (3.19) \end{aligned}$$

Note, that the shift $(t-s, x-\frac{vs}{\varepsilon}, v)$ was chosen in order to reduce the transportation part of the equation $\int_0^t \partial_t f_\varepsilon + v\cdot\nabla f_\varepsilon \mathrm{d}s$ to the expression $f_\varepsilon(t,x,v) - f_0(x,v)$. Also we used that

$$\mathcal{K}_\varepsilon(f_\varepsilon) = \int_V K_\varepsilon f_\varepsilon' - K_\varepsilon' f_\varepsilon \mathrm{d}v \leq \int_V K_\varepsilon f_\varepsilon' \mathrm{d}v$$

due to K_ε, f_ε being positive. Additionally, it holds that $0 < K_\varepsilon = K_0 + \varepsilon K_1 \leq (1+\varepsilon)C \leq 2C$ for small enough ε and that $b(\rho)$ is bounded by a constant b. The occurring ε were estimated by 1. Thus we get

$$\|f_\varepsilon(t,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \leq \|f_0\|_{L^\infty(\mathbb{R}^3\times V)} + (2C|V| + b)\int_0^t \|f_\varepsilon\|_{L^\infty(\mathbb{R}^3\times V)}\mathrm{d}s.$$

Calling Lemma 3.7 (Grönwall's inequality) and using that $f_0 \in L^1_+ \cap L^\infty$, one obtains a bound on $\|f_\varepsilon(t,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)}$. For sufficiently small ε the bound is independent of ε and uniform in $\mathcal{A}$. The boundedness of f_ε in $L^\infty([0,T], L^1_+ \cap L^\infty(\mathbb{R}^3 \times V))$ for some bounded $T < \infty$ then follows directly. By standard methods existence of the solution follows [10].

(ii) Now, let f_ε, $f_{\bar\varepsilon}$ be solutions for the chemotaxis equation (3.3) with scaling parameter ε, $\bar\varepsilon$ respectively, denote their difference by $\hat f_{\varepsilon,\bar\varepsilon} = f_\varepsilon - f_{\bar\varepsilon}$ and let $\mathcal{K}_\varepsilon = \mathcal{K}_0 + \varepsilon\mathcal{K}_1$. Subtracting the equations for f_ε and $f_{\bar\varepsilon}$,

$$\varepsilon^2\partial_t f_\varepsilon + \varepsilon v\cdot\nabla_x f_\varepsilon = \mathcal{K}_\varepsilon(f_\varepsilon) + \varepsilon^2\mathcal{R}_\varepsilon(f_\varepsilon)$$

and

$$\bar\varepsilon^2\partial_t f_{\bar\varepsilon} + \bar\varepsilon v\cdot\nabla_x f_{\bar\varepsilon} = \mathcal{K}_{\bar\varepsilon}(f_{\bar\varepsilon}) + \bar\varepsilon^2\mathcal{R}_{\bar\varepsilon}(f_{\bar\varepsilon})$$

yields

$$\varepsilon^2\partial_t f_\varepsilon + \varepsilon v\cdot\nabla_x f_\varepsilon - \bar\varepsilon^2\partial_t f_{\bar\varepsilon} - \bar\varepsilon v\cdot\nabla_x f_{\bar\varepsilon} = \mathcal{K}_\varepsilon(f_\varepsilon) + \varepsilon^2\mathcal{R}_\varepsilon(f_\varepsilon) - \mathcal{K}_{\bar\varepsilon}(f_{\bar\varepsilon}) - \varepsilon^2\mathcal{R}_{\bar\varepsilon}(f_{\bar\varepsilon}).$$

For the expression $\mathcal{K}_\varepsilon(f_\varepsilon) - \mathcal{K}_{\bar\varepsilon}(f_{\bar\varepsilon})$ we have that

$$\begin{aligned}
\mathcal{K}_\varepsilon(f_\varepsilon) - \mathcal{K}_{\bar\varepsilon}(f_{\bar\varepsilon}) &= \mathcal{K}_0(f_\varepsilon) - \mathcal{K}_0(f_{\bar\varepsilon}) + \varepsilon\mathcal{K}_1(f_\varepsilon) - \bar\varepsilon\mathcal{K}_1(f_{\bar\varepsilon}) \\
&= \mathcal{K}_0(f_\varepsilon - f_{\bar\varepsilon}) + \varepsilon\mathcal{K}_1(f_\varepsilon) - \bar\varepsilon\mathcal{K}_1(f_{\bar\varepsilon}) + \varepsilon\mathcal{K}_1(f_{\bar\varepsilon}) - \varepsilon\mathcal{K}_1(f_{\bar\varepsilon}) \\
&= \mathcal{K}_0(\hat f_{\varepsilon,\bar\varepsilon}) + \varepsilon\mathcal{K}_1(\hat f_{\varepsilon,\bar\varepsilon}) + (\varepsilon - \bar\varepsilon)\mathcal{K}_1(f_{\bar\varepsilon}) \\
&= \mathcal{K}_\varepsilon(\hat f_{\varepsilon,\bar\varepsilon}) + (\varepsilon - \bar\varepsilon)\mathcal{K}_1(f_{\bar\varepsilon}).
\end{aligned}$$

For the expression $\varepsilon^2\mathcal{R}_\varepsilon(f_\varepsilon) - \bar\varepsilon^2\mathcal{R}_{\bar\varepsilon}(f_{\bar\varepsilon})$ we have

$$\begin{aligned}
\varepsilon^2\mathcal{R}_\varepsilon(f_\varepsilon) - \bar\varepsilon^2\mathcal{R}_{\bar\varepsilon}(f_{\bar\varepsilon}) &= \varepsilon^2 b(\rho_\varepsilon) f_\varepsilon - \bar\varepsilon^2 b(\rho_{\bar\varepsilon}) f_{\bar\varepsilon} \\
&= \varepsilon^2 b(\rho_\varepsilon) f_\varepsilon - \bar\varepsilon^2 b(\rho_{\bar\varepsilon}) f_{\bar\varepsilon} + \varepsilon^2 b(\rho_\varepsilon) f_{\bar\varepsilon} - \varepsilon^2 b(\rho_\varepsilon) f_{\bar\varepsilon} \\
&= \varepsilon^2 b(\rho_\varepsilon) \hat f_{\varepsilon,\bar\varepsilon} - \bar\varepsilon^2 b(\rho_{\bar\varepsilon}) f_{\bar\varepsilon} + \varepsilon^2 b(\rho_\varepsilon) f_{\bar\varepsilon} \\
&= \varepsilon^2 b(\rho_\varepsilon) \hat f_{\varepsilon,\bar\varepsilon} - \bar\varepsilon^2 b(\rho_{\bar\varepsilon}) f_{\bar\varepsilon} + \varepsilon^2 b(\rho_\varepsilon) f_{\bar\varepsilon} + \varepsilon^2 b(\rho_{\bar\varepsilon}) f_{\bar\varepsilon} - \varepsilon^2 b(\rho_{\bar\varepsilon}) f_{\bar\varepsilon} \\
&= \varepsilon^2 b(\rho_\varepsilon) \hat f_{\varepsilon,\bar\varepsilon} + \varepsilon^2 \underbrace{\big(b(\rho_\varepsilon) - b(\rho_{\bar\varepsilon})\big)}_{=\frac{r}{c}(\rho_\varepsilon - \rho_{\bar\varepsilon})} f_{\bar\varepsilon} + (\varepsilon^2 - \bar\varepsilon^2) b(\rho_{\bar\varepsilon}) f_{\bar\varepsilon}
\end{aligned}$$

and further it holds that

$$\begin{aligned}
\varepsilon^2\partial_t f_\varepsilon - \bar\varepsilon^2\partial_t f_{\bar\varepsilon} &= \varepsilon^2\partial_t f_\varepsilon - \bar\varepsilon^2\partial_t f_{\bar\varepsilon} + \varepsilon^2\partial_t f_{\bar\varepsilon} - \varepsilon^2\partial_t f_{\bar\varepsilon} \\
&= \varepsilon^2\partial_t(f_\varepsilon - f_{\bar\varepsilon}) + (\varepsilon^2 - \bar\varepsilon^2) f_{\bar\varepsilon} \\
&= \varepsilon^2\partial_t \hat f_{\varepsilon,\bar\varepsilon} + (\varepsilon^2 - \bar\varepsilon^2) f_{\bar\varepsilon}
\end{aligned}$$

and with a similar calculation for $\varepsilon v\cdot\nabla_x f_\varepsilon - \bar\varepsilon v\cdot\nabla_x f_{\bar\varepsilon}$ the chemotaxis equation above then is equivalent to

$$\begin{aligned}
\varepsilon^2\partial_t \hat f_{\varepsilon,\bar\varepsilon} + \varepsilon v\cdot\nabla_x \hat f_{\varepsilon,\bar\varepsilon} = \mathcal{K}_\varepsilon(\hat f_{\varepsilon,\bar\varepsilon}) + \varepsilon^2 b(\rho_\varepsilon)\hat f_{\varepsilon,\bar\varepsilon} + \varepsilon^2\frac{r}{c}(\rho_\varepsilon - \rho_{\bar\varepsilon}) f_{\bar\varepsilon} \\
\underbrace{- \Big((\varepsilon^2 - \bar\varepsilon^2)\partial_t f_{\bar\varepsilon} + (\varepsilon - \bar\varepsilon) v\cdot\nabla_x f_{\bar\varepsilon} - (\varepsilon - \bar\varepsilon)\mathcal{K}_1(f_{\bar\varepsilon}) - (\varepsilon^2 - \bar\varepsilon^2) b(\rho_{\bar\varepsilon}) f_{\bar\varepsilon}\Big)}_{\text{Source term S of order } \varepsilon - \bar\varepsilon}.
\end{aligned} \tag{3.20}$$

Therefore, the function $\hat f_{\varepsilon,\bar\varepsilon}$ satisfies a chemotaxis equation with a linear growth term $b(\rho_\varepsilon)\hat f_{\varepsilon,\bar\varepsilon}$, a source term S and an additional reaction term $\frac{r}{c}(\rho_\varepsilon - \rho_{\bar\varepsilon}) f_{\bar\varepsilon}$. The growth part is linear, since it does not depend on $\hat\rho_{\varepsilon,\bar\varepsilon} = \int_V \hat f_{\varepsilon,\bar\varepsilon}\mathrm{d}v$ but on ρ_ε. Using the same arguments as above the L^∞-boundedness of time and spatial derivative in the source term S can be shown, which is why S is of order $\varepsilon - \bar\varepsilon$. Running the same calculations as in (3.19) with the extra source term S yields

$$\begin{aligned}
\hat f_{\varepsilon,\bar\varepsilon}(t,x,v) &\le \overbrace{\hat f_{\varepsilon,\bar\varepsilon}(0,x,v)}^{=0} + \int_0^t \int_V K_\varepsilon(\hat f_{\varepsilon,\bar\varepsilon})\mathrm{d}v\,\mathrm{d}s + \int_0^t \mathcal{R}_\varepsilon(\hat f_{\varepsilon,\bar\varepsilon})\mathrm{d}s \\
&\quad + \frac{r}{c}\int_0^t f_{\bar\varepsilon}\int_V \hat f_{\varepsilon,\bar\varepsilon}\mathrm{d}v\,\mathrm{d}s + \int_0^t S\mathrm{d}s \\
&\le (2C|V| + b)\int_0^t \hat f_{\varepsilon,\bar\varepsilon}\mathrm{d}s + \frac{r}{c}c_f\int_0^t\int_V \hat f_{\varepsilon,\bar\varepsilon}\mathrm{d}v\,\mathrm{d}s + \int_0^t S\mathrm{d}s
\end{aligned}$$

where we omitted the shift $(t-s, x-\frac{vs}{\varepsilon}, v)$ in the notation for a better readability and used that

$$\|f_{\bar\varepsilon}\|_{L^\infty([0,T],L^\infty(\mathbb{R}\times V))} \le c_f$$

as shown in part (i). Also note, that we replaced $\rho_\varepsilon - \rho_{\bar\varepsilon}$ by $\int_V \hat f_{\varepsilon,\bar\varepsilon}\mathrm{d}v$. This leads to

$$\|\hat f_{\varepsilon,\bar\varepsilon}\|_{L^\infty(\mathbb{R}\times V)} \le \left(2C|V| + b + \frac{r}{c}|V|c_f\right)\int_0^t \|\hat f_{\varepsilon,\bar\varepsilon}\|_{L^\infty(\mathbb{R}\times V)}\mathrm{d}s + \underbrace{\int_0^t \|S\|_{L^\infty(\mathbb{R}\times V)}\mathrm{d}s}_{=\mathcal{O}(\varepsilon-\bar\varepsilon)}.$$

Since we have only non-decreasing functions in t, it follows from Grönwall's inequality (3.16) that

$$\|f_\varepsilon - f_{\bar\varepsilon}\|_{L^\infty([0,T];L^\infty(\mathbb{R}^3\times V))} = \mathcal{O}(\varepsilon-\bar\varepsilon).$$

Additionally, the L^1-boundedness of the source-term follows from the L^1-boundedness of f_ε, as shown in (i). Therefore, integrating (3.20) in time, space and velocity along the characteristic $(t-s, x-\frac{vs}{\varepsilon}, v)$ yields

$$\begin{aligned}\|\hat f_{\varepsilon,\bar\varepsilon}\|_{L^1(\mathbb{R}^3\times V)} &\le b\int_0^t\int_{\mathbb{R}^3}\int_V \hat f_{\varepsilon,\bar\varepsilon}\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}s + \int_0^t\int_{\mathbb{R}^3}\int_V S\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}s\\ &\quad + \frac{r}{c}\int_0^t \|\hat f_{\varepsilon,\bar\varepsilon}\|_{L^1(\mathbb{R}^3\times V)}\|f_{\bar\varepsilon}(s,\cdot,\cdot)\|_{L^1(\mathbb{R}^3\times V)}\mathrm{d}s\\ &\le \left(b + \frac{r}{c}\|f^*\|_{L^\infty([0,T];L^1(\mathbb{R}^3\times V))}\right)\int_0^t \|\hat f_{\varepsilon,\bar\varepsilon}\|_{L^1(\mathbb{R}^3\times V)}\mathrm{d}s + \underbrace{\int_0^t \|S\|_{L^1(\mathbb{R}^3\times V)}\mathrm{d}s}_{\mathcal{O}(\varepsilon-\bar\varepsilon)},\end{aligned}$$

where we used that $\hat f_{\varepsilon,\bar\varepsilon}(0,x,v) = 0$. Again, it follows from Grönwall's inequality (3.16) that

$$\|f_\varepsilon - f_{\bar\varepsilon}\|_{L^\infty([0,T];L^1(\mathbb{R}^3\times V))} = \mathcal{O}(\varepsilon-\bar\varepsilon),$$

thus overall we have

$$\|f_\varepsilon - f_{\bar\varepsilon}\|_{L^\infty([0,T];L^1\cap L^\infty(\mathbb{R}^3\times V))} = \mathcal{O}(\varepsilon-\bar\varepsilon).$$

Hence, $(f_\varepsilon)_\varepsilon$ is a Cauchy sequence and thus converges to some $f \in L^\infty([0,T];L^1\cap L^\infty(\mathbb{R}^3\times V))$. It remains to investigate how the limit function looks like. Substituting the expansion f_ε into the scaled chemotaxis equation

$$\varepsilon^2\partial_t f_\varepsilon + \varepsilon v\cdot\nabla_x f_\varepsilon = (\mathcal{K}_0 + \varepsilon\mathcal{K}_1)(f_\varepsilon) + \varepsilon^2\mathcal{R}_\varepsilon(f_\varepsilon)$$

and comparing the different orders of ε yields the following calculations. Consider the operator $\mathcal{K}_0$ and assume its kernel to be spanned by some velocity distribution F with $\int_V F\mathrm{d}v = 1$ and $\int_V vF\mathrm{d}v = 0$, as motivated in (3.9). Then, the comparison of orders in ε results in the following. For $\mathcal{O}(1)$ we have that

$$\mathcal{K}_0(f_0) = 0$$

leading to $f_0(t,x,v) = \rho_0(t,x)F(v)$ since then

$$\mathcal{K}_0(f_0) = \rho(t,x)\mathcal{K}_0(F) = 0$$

is fulfilled due to $F \in \ker(\mathcal{K}_0)$. Further, looking at $\mathcal{O}(\varepsilon)$ yields

$$\begin{aligned}\mathcal{K}_0(f_1) &= v\cdot\nabla_x f_0 - \mathcal{K}_1(f_0)\\ &= vF\nabla_x\rho_0 - \mathcal{K}_1(F)\rho_0,\end{aligned}$$

resulting in

$$f_1(t,x,v) = \kappa(t,x,v)\nabla_x\rho_0(t,x) - \Theta(t,x,v)\rho_0(t,x) + \rho_1 F_1$$

with $F_1 \in \ker(\mathcal{K}_0)$ and κ and Θ being determined by the conditions

$$\mathcal{K}_0(\kappa) = vF \tag{3.21}$$
$$\mathcal{K}_0(\Theta) = \mathcal{K}_1(F). \tag{3.22}$$

The functions κ and Θ exist, following Lemma 3.5 since $vF, \mathcal{K}_1(F) \in L^2(V; \mathrm{d}v/F)$ and

$$\int_V vF\mathrm{d}v = 0$$

holds by assumption and we have

$$\begin{aligned}
\int_V \mathcal{K}_1(F)\mathrm{d}v &= \int_V \Big(\int_V K_1(t,x,v,v')F(v') - K_1(t,x,v',v)F(v)\mathrm{d}v' \Big)\mathrm{d}v \\
&= \int_V \Big(\int_V K_1(t,x,v,v')F(v')\mathrm{d}v' \Big)\mathrm{d}v - \underbrace{\int_V \Big(\int_V K_1(t,x,v',v)F(v)\mathrm{d}v' \Big)\mathrm{d}v}_{v\to v', v'\to v} \\
&= \int_V \Big(\int_V K_1(t,x,v,v')F(v')\mathrm{d}v' \Big)\mathrm{d}v - \int_V \Big(\int_V K_1(t,x,v,v')\mathrm{d}v \Big)F(v')\mathrm{d}v' \\
&= \int_V \Big(\int_V K_1(t,x,v,v')F(v')\mathrm{d}v' \Big)\mathrm{d}v - \int_V \Big(\int_V K_1(t,x,v,v')F(v')\mathrm{d}v' \Big)\mathrm{d}v \\
&= 0.
\end{aligned}$$

Now we integrate the chemotaxis equation in v and divide by ε^2

$$\partial_t \int_V f_\varepsilon \mathrm{d}v + \nabla_x \frac{1}{\varepsilon}\int_V v f_\varepsilon \mathrm{d}v = \frac{1}{\varepsilon^2}\underbrace{\int_V \mathcal{K}_\varepsilon(f_\varepsilon)\mathrm{d}v}_{=0} + b(\rho_\varepsilon)\int_V f_\varepsilon \mathrm{d}v$$

leading to the conservation equation

$$\partial_t \rho_\varepsilon + \nabla_x J_\varepsilon = \mathcal{R}_\varepsilon(\rho_\varepsilon) \tag{3.23}$$

with $\rho_\varepsilon = \int_V f_\varepsilon \mathrm{d}v$ and the flux $J_\varepsilon = \frac{1}{\varepsilon}\int_V v f_\varepsilon \mathrm{d}v$. Carrying out the expansion yields

$$J_\varepsilon = \int_V v f_1 \mathrm{d}v + \mathcal{O}(\varepsilon),$$

where we used that $f_0 = \rho_0 F$ as well as that the assumption $\int_V vF\mathrm{d}v = 0$ holds, resulting in

$$\int_V v f_0 \mathrm{d}v = \rho_0 \int_V vF\mathrm{d}v = 0.$$

Inserting f_1 into (3.23) yields the convection-diffusion equation

$$\partial_t \rho_0 - \nabla(D\nabla\rho_0) + \nabla(\rho_0\Gamma) = \mathcal{R}_0(\rho_0)$$

with

$$\begin{aligned}
D(t,x) &= \int_V v \otimes \kappa(t,x,v)\mathrm{d}v, \\
\Gamma(t,x) &= -\int_V v\Theta(t,x,v)\mathrm{d}v.
\end{aligned}$$

(iii) Since only the boundedness of (K_0, K_1) is used within the proof, the boundedness as well as the convergence are uniform in $\mathcal{A}$. □

Overall, it could be shown that the two forward model are still asymptotic equivalent in the long-time large-space regime when introducing a reaction term. The presented models are well-suited to predict the motion of the bacteria. However, in reality the parameters of the models, namely the diffusion and drift coefficients D, Γ in the macroscopic model or the tumbling kernel K in the kinetic model, are usually unknown except from some well-studied bacteria. Hence, they have to be derived from measuring the densities of the bacteria at certain times. Thus, experiments have to be designed with measurable quantities in order to reconstruct the model parameters. This reconstruction then corresponds to an *inverse problem* and will be discussed in the following section.

4 Chemotaxis: Inverse Problem

In this section, the formulation of the inverse problems in the Bayesian setting is described. Then, the convergence of the two inverse problems occurring from the kinetic and macroscopic forward problem is investigated. Within the literature the study of inverse problems in the Bayesian setting is well studied and prevalent, as for instance in [12, 36] and the references therein. However, relating two inverse problems from different regimes to each other is relatively rare. For example, in [31] the asymptotic limit for the inverse problem coming from kinetic radiative transfer equation and the corresponding macroscopic diffusion equation is established. Additionally, in [17] the convergence of the inverse problems for chemotaxis (without a birth-death term) on kinetic and macroscopic level in the asymptotic limit is shown. The results from [17] will be extended to the case of a chemotaxis equation including a reaction term in this section and we will mainly follow the method presented in [17, Chapter 3-4].

Each of the two forward models gives rise to an inverse problem which will be described within the formulation of Bayesian inference in this section. This means, that for the quantities of interest, e.g. on the kinetic level the tumbling kernel and on the macroscopic level the drift- and diffusion coefficients, are assumed to be random variables. Hence, the goal is to show that the posterior distribution of the kinetic problem converges to the posterior distribution coming from the macroscopic equations. We highlight, that in section 3 the quantities D and Γ where derived in a way, such that they depend implicitly on the tumbling kernel (K_0, K_1) via the functions κ and Θ. In order to be able to compare the two inverse problems, one has to make sure to infer the same quantity in both cases. Either, one could determine the underlying tumbling kernel for the macroscopic equation and compare those. Or one could determine the drift and diffusion coefficients from the macroscopic equation and derive those coefficients from the determined tumbling kernel of the chemotaxis equation [17]. In this work, we will choose the first option, thus comparing the underlying tumbling kernels, due to it being of kinetic nature and therefore more detailed. Since the calculation of (D, Γ) is unique for fixed (K_0, K_1), the convergence then follows as a consequence.

Since the dependence of the Keller-Segel-Fisher equation on (K_0, K_1) is only implicit via the functions κ and Θ, one has to clarify how to derive the posterior distribution $\mu^y_{\mathrm{KS}}(D, \Gamma)$ from a prior distribution on (K_0, K_1). One way would be, to transform the prior to a prior distribution on (D, Γ) and then reconstruct $\mu^y_{\mathrm{KS}}(D, \Gamma)$ in the inverse problem, as illustrated in the lower path in Figure 7. Another approach is, to infer the posterior $\mu^y_{\mathrm{KS}}(K_0, K_1)$ from the prior on (K_0, K_1) in the inverse problem and to transform it to a distribution on (D, Γ) afterwards, as demonstrated in the upper path of Figure 7. In the following sections, the second way is chosen, but one could rightly

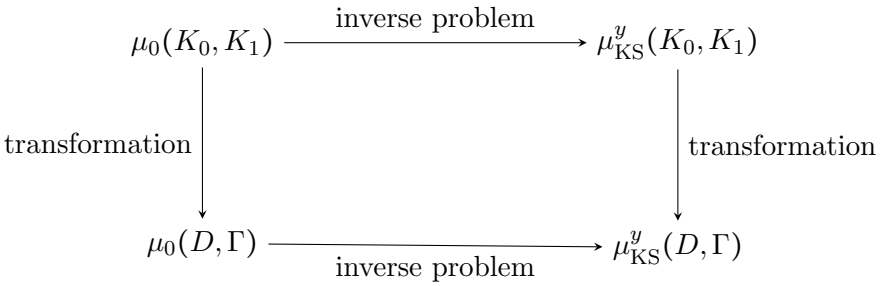


Figure 7: Determination of the posterior distribution $\mu^y_{\mathrm{KS}}(D, \Gamma)$ from a prior distribution on (K_0, K_1) in two different ways. [17, Figure 3]

ask themselves, whether or not the two possibilities yield the same posterior distribution. Looking at the second approach, only the prior distribution is transformed since the likelihood only depends on (D, Γ) in the macroscopic regime. This is also true for the first way, hence they indeed yield the

same results [17, Remark 5].

For the experiment it is assumed to have no boundary effects which correlates to having a large enough plate for the bacteria in the lab environment. The initial condition describes the distribution of cells at the beginning of the experiment. After injecting the chemoattractant into the plate in a controlled environment such that the concentration function is known, the bacteria density at time $t > 0$ and point $x \in \mathbb{R}^3$ is measured. Usually, the measurements are done by taking a picture of the plate and counting the bacteria in a small neighbourhood of x. There are other ways to do them, for example by taking a sample and measure the density by techniques such as flow cytometry [9, 15]. But these kind of techniques are invasive which is why the measurement can only be taken at one time.

4.1 Bayesian Inverse Setup

Depending on the model we use in the forward problem, we collect data of

$$\mathcal{A}^{\varepsilon}_{K_0,K_1} : f_0 \mapsto \int_V f_\varepsilon(t,x,v)\mathrm{d}v$$
$$\mathcal{A}^{0}_{K_0,K_1} = \mathcal{A}_{D,\Gamma} : \rho_0 := \int_V f_0 \mathrm{d}v \mapsto \rho(t,x)$$

respectively. Those operators are also called *Albedo-Operators* and relate the initial condition to the measured data.

When performing multiple experiments with various initial conditions, the same controlled $c(t,x)$ has to be used in order to ensure that the to-be-reconstructed tumbling kernel is the same for all experiments. The indices of the different setups are denoted by $k \in [1,\dots,K]$ and we indicate the measuring time and location by $t_j = t_{j_1}$ and $\chi_j = \chi_{j_2} \in C_c(\mathbb{R}^3)$ where $j = (j_1, j_2) \in [1,\dots,J_1] \otimes [1,\dots,J_2]$. We consider the measurements

$$\mathcal{G}^{\varepsilon,\mathrm{chem}}_{j,k}(K_0,K_1) = \int_{\mathcal{R}^3}\int_V f^{(k)}_\varepsilon(t_j,x,v)\mathrm{d}v\chi_j(x)\mathrm{d}x, \tag{4.1}$$
$$\mathcal{G}^{\mathrm{KS}}_{j,k}(K_0,K_1) = \int_{\mathcal{R}^3}\rho^{(k)}(t_j,x)\chi_j(x)\mathrm{d}x \tag{4.2}$$

where the test functions χ_j can be interpreted as compactly supported blob function at a certain location.

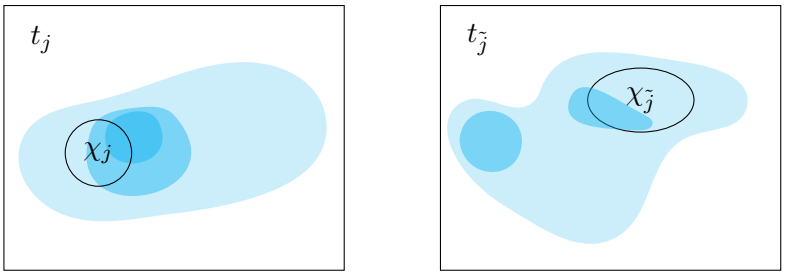


Figure 8: Representation of measurements at two different time steps $t_j, t_{\tilde{j}}$. The density of bacteria is illustrated as blue, where the darkness of the color indicates the number of bacteria. The location of the measurement is illustrated by the support of the test functions $\chi_j, \chi_{\tilde{j}}$. [17, Figure 2]

Further, we collect data of the form

$$y^{\varepsilon,\mathrm{chem}}_{j,k} = \mathcal{G}^{\varepsilon,\mathrm{chem}}_{j,k}(K_0,K_1) + \eta_{jk},$$
$$y^{\mathrm{KS}}_{j,k} = \mathcal{G}^{\mathrm{KS}}_{j,k}(K_0,K_1) + \eta_{jk}$$

where the noise η_{jk} is assumed to be a random variable with a Gaussian distribution $\mathcal{N}(0,\gamma^2)$ with zero mean and known variance $\gamma^2 > 0$.

In the Bayesian setting (K_0, K_1) is assumed to be a random variable and the goal is to reconstruct its distribution $\mu_\circ^y(K_0,K_1)$ via

$$\mu_\circ^y(K_0,K_1) = \frac{1}{Z^\circ}\mu_\circ^{(K_0,K_1)}(y)\mu_0(K_0,K_1) \tag{4.3}$$

with

$$\mu_\circ^{(K_0,K_1)}(y) = \exp\left(-\frac{1}{2\gamma^2}\|\mathcal{G}^\circ(K_0,K_1) - y\|^2\right)$$

where $\circ = \varepsilon, \text{chem}$ or $\circ = \text{KS}$ and Z° being the normalization constant, μ_0 being the prior distribution and $\mu_\circ^{(K_0.K_1)}$ the likelihood of observing data y from a model with tumbling kernel (K_0, K_1).

4.2 Convergence of Posterior Distributions

Having two different forward models results in two different posterior distributions for the inverse problems. The question arises whether or not, the posterior distribution coming from the kinetic equation does also converge towards the posterior distribution from the Keller-Segel-Fisher equation. Since we have the convergence of the forward models one might assume that the convergence of the inverse problem follows directly, however, this is not the case. There are several issues arising during the studies of the convergence, as discussed in [17]. One arises from the control of initial data and the measurement operator. The solution converges in $L^\infty([0,T], L^1_+ \cap L^\infty(\mathbb{R}^3 \times V))$ for each initial condition. Then a uniform convergence is needed when tested on the dual space, since we have a list of initial data and the solutions are tested on a set of measuring operators. Another important aspect is the choice of the metric on the probability function space. An additional issue comes from the translation of convergence for one set of fixed (K_0, K_1) to the convergence on the entire admissible set.

In the following we will assume that the initial data and the measuring operators are bounded

$$\begin{aligned} &\|f_0^{(k)}\|_{L^1}, \|f_0^{(k)}\|_{L^\infty} < C_\rho && \forall k \\ &\max\{\|\chi_j\|_{L^1}, \|\chi_j\|_{L^2}, \|\chi_j\|_{L^\infty}, |\text{supp}\chi_j|_{\mathrm{d}x}\} < C_x && \forall j, \end{aligned} \tag{4.4}$$

as done in [17].

Remark 4.1 When the uniform boundedness and convergence of f_ε was shown in subsection 3.3, we imposed an assumption on the tumbling kernel (K_0, K_1) by choosing the admissible set. The usage of this a priori knowledge will also play a crucial role in the convergence proof for the inverse problems and the prior distribution μ_0 is assumed to be supported on $\mathcal{A}$.

Within this section the *Lax-Milgram-Theorem* will be needed, especially some details from its proof. Therefore, the proof following [13, Chapter 6.2.1., Theorem 1] will be carried out, even though the theorem is well known within linear functional analysis.

Theorem 4.2 (Lax-Milgram Theorem) *Suppose that for the bilinear mapping $B : H \times H \to \mathbb{R}$ with H being a real Hilbert space there exist constants $\alpha, \beta > 0$ such that*

(i) B is bounded

$$|B(u,v)| \le \alpha\|u\|\|v\|$$

(ii) and coercive

$$\beta\|u\|^2 \le B(u,u).$$

Let $f : H \to \mathbb{R}$ be a bounded linear function. Then there exists a unique $u \in H$ such that

$$B(u, v) = \langle f, v \rangle \tag{4.5}$$

holds true for all $v \in H$.

PROOF: The mapping $v \mapsto B(u, v)$ is a bounded linear functional on H for each fixed element $u \in H$ thus it follows from the Riesz Representation Theorem that there exists a unique element $w \in H$ with

$$B(u, v) = \langle w, v \rangle.$$

We will write $Au = w$ leading to

$$B(u, v) = \langle Au, v \rangle \tag{4.6}$$

and we want to show that $A : H \to H$ is a bounded linear operator. Let $\lambda_1, \lambda_2 \in \mathbb{R}$ and $u_1, u_2 \in H$. Linearity of A follows directly from the bilinearity of B and (4.6). Further, it follows from the boundedness of B that

$$\|Au\|^2 = \langle Au, Au \rangle = B(u, Au) \leq \alpha \|u\| \|Au\|$$

holds and so A is bounded. From the coercivity of B it follows that

$$\beta \|u\|^2 \leq B(u, u) = \langle Au, u \rangle \leq \|Au\| \|u\|.$$

Therefore, $\beta \|u\| \leq \|Au\|$ and A is one-to-one and the range $R(A)$ of A is closed in H. Since $R(A)$ is closed in H, it follows that $R(A) = H$. Otherwise, there would be a non-zero element $w \in H$ with $w \in R(A)^\perp$, meaning that

$$\beta \|w\|^2 \leq B(w, w) = \langle Aw, w \rangle = 0$$

which is a contradiction to w being nonzero. From the *Riesz Representation Theorem* it now follows that

$$\langle f, v \rangle = (w, v)$$

for all $v \in H$ showing (4.5). It remains to show that there is at most one element $u \in H$ verifying (4.5). Assume there exists another $\tilde{u} \in H$ with $B(\tilde{u}, v) = \langle f, v \rangle$. Then $B(u - \tilde{u}, v) = 0$ for $v \in H$. With $v = u - \tilde{u}$ we get $\beta \|u - \tilde{u}\|^2 \leq B(u - \tilde{u}, u - \tilde{u}) = 0$, thus $u = \tilde{u}$ follows. □

Remark 4.3 In the following, if we use the Lax-Milgram Theorem we will mostly use the fact that the operator A is bounded from below and above

$$\beta \|u\| \leq \|Au\| \leq \alpha \|u\|.$$

In order to use the theorem of Lax-Milgram, we first have to show, that we get a bilinear mapping B from the operators $\mathcal{K}_0, \mathcal{K}_1$ that fulfils the conditions of Theorem 4.2.

Proposition 4.4 *The mappings $B_{0,1}(u, w) : H \times H \to \mathbb{R}$ for $H = L^2(V; \mathrm{d}v/F)$ defined via*

$$B_0(u, w) = \int_V \mathcal{K}_0(u) w \frac{\mathrm{d}v}{F},$$
$$B_1(u, w) = \int_V \mathcal{K}_1(u) w \frac{\mathrm{d}v}{F}$$

are bilinear, bounded and coercive.

PROOF: We will prove the proposition for B_0, the results for B_1 follow analogously. We will use the notation $u' = u(v')$ and $u = u(v)$ as well as $K_0' = K_0(v', v)$ and $K_0 = K_0(v, v')$. The dependence on (t, x) will be dropped for the sake of better readability. Since the operator $\mathcal{K}_0$ is linear in u, it follows directly that B_0 is a bilinear mapping. It holds that

$$\begin{aligned} |B(u,w)| &\leq \int_V |\mathcal{K}_0(u) w \frac{1}{F}| \mathrm{d}v \\ &\overset{(*)}{\leq} \|\mathcal{K}_0(u)\|_H \|w\|_H \end{aligned}$$

where in $(*)$ we used Hölder's inequality. Remember, that since $(K_0, K_1) \in \mathcal{A}$ we have $K_0 \leq C$. Hence, for $\|\mathcal{K}_0(u)\|_H$ it holds that

$$\begin{aligned} \|\mathcal{K}_0(u)\|_H^2 &= \int_V \Big(\int_V \big(K_0 u' - K_0' u \big) \mathrm{d}v' \Big)^2 \frac{\mathrm{d}v}{F(v)} \\ &\overset{(\Delta)}{\leq} \int_V \int_V \big(K_0 u' - K_0' u \big)^2 \mathrm{d}v' \frac{\mathrm{d}v}{F(v)} \\ &\leq \int_V \Big(\int_V (K_0 u')^2 + (K_0' u)^2 + 2 K_0 K_0' u u' \Big) \mathrm{d}v' \frac{\mathrm{d}v}{F(v)} \\ &\leq C^2 \int_V \left(\frac{\int_V (u')^2 \mathrm{d}v'}{F(v)} + \int_V \frac{u^2}{F(v)} \mathrm{d}v' \right) \mathrm{d}v + \int_V \int_V K_0 K_0' |2uu'| \frac{1}{F(v)} \mathrm{d}v' \, \mathrm{d}v \\ &\overset{(\star)}{\leq} C^2 \int_V \left(\frac{\int_V (u')^2 \mathrm{d}v'}{F(v)} + \int_V \frac{u^2}{F(v)} \mathrm{d}v' \right) \mathrm{d}v + C^2 \int_V \left(\frac{\int_V (u')^2 \mathrm{d}v'}{F(v)} + \int_V \frac{u^2}{F(v)} \mathrm{d}v' \right) \mathrm{d}v \\ &= 2C^2 \int_V \left(\int_V \frac{1}{F(V)} \mathrm{d}v \int_V (u')^2 \mathrm{d}v' + \int_V 1 \mathrm{d}v' \frac{u^2}{F(v)} \right) \mathrm{d}v \\ &\leq 2C^2 \int_V \left(\|F\|_\infty \int_V \frac{1}{F(v)} \mathrm{d}v \int_V \frac{(u')^2}{F(v')} \mathrm{d}v' + |V| \int_V \frac{u^2}{F(v)} \mathrm{d}v \right) \mathrm{d}v \\ &\leq 2C^2 |V| \left(\frac{\|F\|_\infty}{\min F} + 1 \right) \int_V \frac{u^2}{F(v)} \mathrm{d}v \end{aligned}$$

where in (Δ) we used Jensen's inequality and in $(\star)$ we used $K_0 K_0' \leq C^2$ and Cauchy-Schwarz in the sense that $|2uu'| \leq u^2 + (u')^2$. Additionally, we used that

$$\int_V u^2 \mathrm{d}v \leq \|F\|_\infty \int_V \frac{u^2}{F(v)} \mathrm{d}v \quad \text{and} \quad \int_V \frac{1}{F(v)} \mathrm{d}v \leq |V| \frac{1}{\min F}.$$

Overall, this yields

$$|B(u,w)| \leq C \sqrt{2|V| \left(\frac{\|F\|_\infty}{\min F} + 1 \right)} \|u\|_H \|w\|_H.$$

That B_0 is coercive follows from (3.13)

$$\int_V \mathcal{K}_0(f) \frac{f}{F} \mathrm{d}v \geq \frac{\gamma}{2} \int_V \int_V F F' \left(\frac{f}{F} - \frac{f}{F} \right)^2 \mathrm{d}v' \, \mathrm{d}v = \gamma \int_V \frac{f^2}{F} \mathrm{d}v$$

thus leading to

$$B(u,u) = \int_V \mathcal{K}_0(u) u \frac{\mathrm{d}v}{F}$$

$$\geq \gamma \int_V \frac{u^2}{F} \mathrm{d}v$$
$$= \gamma \|u\|_H^2,$$

which concludes the proof. □

For the proof of the lemma below, we will need the following result from [24, Chapter 6, Theorem 6.3].

Theorem 4.5 *Let $\Omega \subset \mathbb{R}^3$ be a bounded and open subset and $T > 0$ and consider the equation*

$$u_t - \sum_{i,j=1}^{3} \partial_i (a^{ij} \partial_j u) + \sum_{j=1}^{3} b^j \partial_j u + cu = f \tag{4.7}$$
$$u(0,x) = g(x) \tag{4.8}$$

where $a^{i,j}(t,x), b^i(t,x), c(t,x)$ are bounded coefficient functions with $a^{i,j} = a^{j,i}$. Additionally, assume that there exists $\psi > 0$ such that

$$\sum_{i,j=1}^{3} a_{ij}(t,x)\zeta_i\zeta_j \geq \psi|\zeta|^2 \quad \text{for all } (t,x) \in (0,T) \times \Omega \text{ and } \zeta \in \mathbb{R}^3.$$

Further, we assume to have a $L^2(0,T;H^{-1})$ source term f and $L^2(\Omega)$ initial condition g with H^{-1} being the L^2-Sobolev space. Then equation (4.7) has a unique weak solution

$$u \in C\big([0,T]; L^2(\Omega)\big) \cap L^2\big(0,T : H_0^1(\Omega)\big)$$

where $H^{-1}(\Omega) = H_0^1(\Omega)'$. Moreover, there exists a constant C depending on Ω, T and the coefficient functions, such that

$$\|u\|_{L^\infty(0,T;L^2)} + \|u\|_{L^2(0,T;H_0^1)} + \|u_t\|_{L^2(0,T;H^{-1})} \leq C\big(\|f\|_{L^2(0,T;H^{-1})} + \|g\|_{L^2}\big).$$

Taking [17, Lemma 2] as our basis we apply it to the chemotaxis equation including a reaction term. This makes the following Lemma 4.6 into a result extending [17, Lemma 2].

Lemma 4.6 *Suppose that the initial condition $f_0(x,v) \in C_c^1(\mathbb{R}^3 \times V)$ and the test functions $\chi_j(x) \in C_c(\mathbb{R}^3)$ fulfil condition (4.4). Then the following properties hold true.*

(i) The measurements $\mathcal{G}_{j,k}^{\varepsilon,\mathrm{chem}}$ and $\mathcal{G}_{j,k}^{\mathrm{KS}}$ are uniformly bounded in $\mathcal{A}$ and ε.

(ii) The measurements $\mathcal{G}_{j,k}^{\varepsilon,\mathrm{chem}}$ and $\mathcal{G}_{j,k}^{\mathrm{KS}}$ are Lipschitz continuous for small enough ε with respect to the tumbling kernel (K_0, K_1) under the norm $\|(K_0,K_1)\|_* = \max(\|K_0\|_\infty, \|K_1\|_\infty)$ on $\mathcal{A}$.

(iii) The posterior distributions are well-posed and absolutely continuous with respect of each other.

Proof: (i) For all tupel (j,k) we have

$$\begin{aligned}
|\mathcal{G}_{j,k}^{\mathrm{KS}}(K_0,K_1)| &= \Big|\int_{\mathcal{R}^3} \rho^{(k)}(t_j,x)\chi_j(x)\mathrm{d}x\Big| \\
&\leq \|\chi_j(x)\|_\infty \|\rho^{(k)}(t_j,\cdot)\|_{L^1(\mathbb{R}^3)} \\
&\leq \|\chi_j(x)\|_\infty \|\rho^*\|_{L^1(\mathbb{R}^3)} \\
&\leq C_x C_\rho
\end{aligned}$$

with ρ^* being the saturation density. Boundedness of $\mathcal{G}_{j,k}^{\varepsilon,\mathrm{chem}}(K_0,K_1)$ follows analogously.

(ii) First we show Lipschitz continuity of $\mathcal{G}^{\varepsilon,\mathrm{chem}}_{j,k}(K_0,K_1)$ with respect to (K_0,K_1) under the norm $\|(K_0,K_1)\|_* := \max(\|K_0\|_\infty,\|K_1\|_\infty)$. For this, let $(K_0,K_1),(\tilde K_0,\tilde K_1)\in\mathcal{A}$.
Then we have

$$\begin{aligned}
|\mathcal{G}^{\varepsilon,\mathrm{chem}}_{j,k}(K_0,K_1)-\mathcal{G}^{\varepsilon,\mathrm{chem}}_{j,k}(\tilde K_0,\tilde K_1)| &= |\int_{\mathbb{R}^3}\int_V \overbrace{\Big(f^{(k)}_\varepsilon-\tilde f^{(k)}_\varepsilon\Big)}^{\bar f^{(k)}_\varepsilon}\mathrm{d}v\chi_j(x)\mathrm{d}x| \\
&\le \|\chi_j(x)\|_\infty\int_{\mathrm{supp}(\chi_j(x))}\int_V\|\bar f^{(k)}_\varepsilon\|\mathrm{d}v\,\mathrm{d}x \\
&\le C_x|V|\|\mathrm{supp}(\chi_j(x))|_{\mathrm{d}x}\|\bar f^{(k)}_\varepsilon\|_{L^\infty(\mathbb{R}^3\times V)} \\
&\le C_x^2|V|\underbrace{\|\bar f^{(k)}_\varepsilon(t,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)}}_{\text{needs to be estimated}}.
\end{aligned}$$

The function $\bar f^{(k)}_\varepsilon$ satisfies the scaled equation

$$\varepsilon^2\partial_t\Big(f^{(k)}_\varepsilon-\tilde f^{(k)}_\varepsilon\Big)+\varepsilon v\cdot\nabla_x\Big(f^{(k)}_\varepsilon-\tilde f^{(k)}_\varepsilon\Big)=\underbrace{\mathcal{K}_\varepsilon(f^{(k)}_\varepsilon)-\tilde{\mathcal{K}}_\varepsilon(\tilde f^{(k)}_\varepsilon)}_{(2)}+\underbrace{\varepsilon^2\mathcal{R}_\varepsilon(f^{(k)}_\varepsilon)-\varepsilon^2\tilde{\mathcal{R}}_\varepsilon(\tilde f^{(k)}_\varepsilon)}_{(1)}.$$

For (1) it holds that

$$\begin{aligned}
b(\rho_\varepsilon)f^{(k)}_\varepsilon-b(\tilde\rho_\varepsilon)\tilde f^{(k)}_\varepsilon &= r(1-\frac{\rho_\varepsilon}{c})f^{(k)}_\varepsilon-r(1-\frac{\tilde\rho_\varepsilon}{c})\tilde f^{(k)}_\varepsilon \\
&= r(f^{(k)}_\varepsilon-\tilde f^{(k)}_\varepsilon)-\frac{r}{c}(\rho_\varepsilon f^{(k)}_\varepsilon-\tilde\rho_\varepsilon\tilde f^{(k)}_\varepsilon) \\
&= r(f^{(k)}_\varepsilon-\tilde f^{(k)}_\varepsilon)-\frac{r}{c}(\rho_\varepsilon f^{(k)}_\varepsilon-\tilde\rho_\varepsilon\tilde f^{(k)}_\varepsilon+\tilde\rho_\varepsilon f^{(k)}_\varepsilon-\tilde\rho_\varepsilon f^{(k)}_\varepsilon) \\
&= r(1-\frac{\tilde\rho_\varepsilon}{c})(f^{(k)}_\varepsilon-\tilde f^{(k)}_\varepsilon)-\frac{r}{c}(\rho_\varepsilon-\tilde\rho_\varepsilon)f^{(k)}_\varepsilon \\
&= b(\tilde\rho_\varepsilon)\bar f^{(k)}_\varepsilon-\frac{r}{c}\bar\rho_\varepsilon f^{(k)}_\varepsilon
\end{aligned}$$

with $\bar\rho_\varepsilon=\rho_\varepsilon-\tilde\rho_\varepsilon$. Rewriting (2) as

$$\begin{aligned}
\mathcal{K}_\varepsilon(f^{(k)}_\varepsilon)-\tilde{\mathcal{K}}_\varepsilon(\tilde f^{(k)}_\varepsilon) &= \mathcal{K}_\varepsilon(f^{(k)}_\varepsilon)-\tilde{\mathcal{K}}_\varepsilon(f^{(k)}_\varepsilon)+\tilde{\mathcal{K}}_\varepsilon(f^{(k)}_\varepsilon)-\tilde{\mathcal{K}}_\varepsilon(\tilde f^{(k)}_\varepsilon) \\
&= \bar{\mathcal{K}}_\varepsilon(f^{(k)}_\varepsilon)+\tilde{\mathcal{K}}_\varepsilon(\bar f^{(k)}_\varepsilon)
\end{aligned}$$

with $\bar{\mathcal{K}}_\varepsilon=\mathcal{K}_\varepsilon-\tilde{\mathcal{K}}_\varepsilon$ yields

$$\begin{aligned}
\varepsilon^2\partial_t\bar f^{(k)}_\varepsilon+\varepsilon v\cdot\nabla_x\bar f^{(k)}_\varepsilon &= \bar{\mathcal{K}}_\varepsilon(f^{(k)}_\varepsilon)+\tilde{\mathcal{K}}_\varepsilon(\bar f^{(k)}_\varepsilon)+\varepsilon^2 b(\tilde\rho_\varepsilon)\bar f^{(k)}_\varepsilon-\varepsilon^2\frac{r}{c}\bar\rho_\varepsilon f^{(k)}_\varepsilon \\
\bar f^{(k)}_\varepsilon(0,x,v) &= 0. \qquad (4.9)
\end{aligned}$$

The goal is to again use Grönwall's inequality for the chemotaxis equation derived for $\bar f^{(k)}_\varepsilon$ in order to get an estimate via $\|\bar{\mathcal{K}}_\varepsilon\|_{L^\infty(\mathbb{R}^3\times V)}\le 2\|\bar K_\varepsilon\|_{L^\infty(\mathbb{R}^3\times V)}\le 4\|(K_0-\tilde K_0,K_1-\tilde K_1)\|_*$.
Integration in s along the characteristic $(t-s,x-\frac{vs}{\varepsilon},v)$ shows

$$\begin{aligned}
\bar f^{(k)}_\varepsilon &= \int_0^t\Big(\tilde{\mathcal{K}}_\varepsilon\Big(\bar f^{(k)}_\varepsilon\Big)(t-s,x-\frac{vs}{\varepsilon},v)+\bar{\mathcal{K}}_\varepsilon\Big(f^{(k)}_\varepsilon\Big)(t-s,x-\frac{vs}{\varepsilon},v)\Big)\mathrm{d}s \\
&\quad+\int_0^t b(\tilde\rho_\varepsilon)\bar f^{(k)}_\varepsilon(t-s,x-\frac{vs}{\varepsilon},v)\mathrm{d}s-\frac{r}{c}\int_0^t\int_V\bar f^{(k)}_\varepsilon(t-s,x-\frac{vs}{\varepsilon},v)\mathrm{d}v f^{(k)}_\varepsilon\mathrm{d}s
\end{aligned}$$

$$
\begin{aligned}
&\leq \int_0^t \int_V \Big(\tilde{K}_\varepsilon \bar{f}'_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) - \tilde{K}'_\varepsilon \bar{f}_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \Big) \mathrm{d}v' \, \mathrm{d}s \\
&\quad + \int_0^t \int_V \Big(\bar{K}_\varepsilon f'_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) - \bar{K}'_\varepsilon f_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \Big) \mathrm{d}v' \, \mathrm{d}s \\
&\quad + \int_0^t b(\tilde{\rho}) \bar{f}_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \mathrm{d}s - \frac{r}{c} \int_0^t f_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \int_V \bar{f}_\varepsilon \mathrm{d}v \, \mathrm{d}s \\
&\leq \int_0^t \int_V \tilde{K}_\varepsilon \bar{f}'_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \mathrm{d}v' \, \mathrm{d}s + \int_0^t \int_V \tilde{K}'_\varepsilon \bar{f}_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \mathrm{d}v' \, \mathrm{d}s \\
&\quad + \int_0^t \int_V \bar{K}_\varepsilon f'_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \mathrm{d}v' \, \mathrm{d}s + \int_0^t \int_V \bar{K}'_\varepsilon f_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \mathrm{d}v' \, \mathrm{d}s \\
&\quad + \int_0^t b(\tilde{\rho}) \bar{f}_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \mathrm{d}s - \frac{r}{c} \int_0^t f_\varepsilon(t-s, x - \frac{vs}{\varepsilon}, v) \int_V \bar{f}_\varepsilon \mathrm{d}v \, \mathrm{d}s
\end{aligned}
$$

leading to

$$
\begin{aligned}
\|\bar{f}_\varepsilon^{(k)}(t,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} &\leq 2C|V| \int_0^t \|\bar{f}_\varepsilon^{(k)}(t-s,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \mathrm{d}s \\
&\quad + 2C|V| \int_0^t \|\bar{f}_\varepsilon^{(k)}(t-s,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \mathrm{d}s \\
&\quad + 2\|(K_0 - \tilde{K}_0, K_1 - \tilde{K}_1)\|_* |V| \int_0^t c_f \mathrm{d}s \\
&\quad + 2\|(K_0 - \tilde{K}_0, K_1 - \tilde{K}_1)\|_* |V| \int_0^t c_f \mathrm{d}s \\
&\quad + b \int_0^t \|\bar{f}_\varepsilon^{(k)}(t-s,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \mathrm{d}s \\
&\quad + \frac{r}{c}|V| c_f \int_0^t \|\bar{f}_\varepsilon^{(k)}(t-s,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \mathrm{d}s \\
&\leq \Big(4C|V| + b + \frac{r}{c}|V| c_f \Big) \int_0^t \|\bar{f}_\varepsilon^{(k)}(t-s,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \mathrm{d}s \\
&\quad + 4\|(K_0 - \tilde{K}_0, K_1 - \tilde{K}_1)\|_* |V| c_f T
\end{aligned}
$$

where we used that $\tilde{K}_\varepsilon \leq 2C$ by assumption since $(K_0, K_1) \in \mathcal{A}$ and that $f_\varepsilon^{(k)} \leq c_f$ is bounded in L^∞ uniformly on $\mathcal{A}$ by Theorem 3.8 (i). Using the Grönwall lemma, (3.16) gives an upper bound

$$
\|\bar{f}_\varepsilon^{(k)}(t,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \leq L(T, c_f, C, b, r, c) \|(\mathcal{K}_0 - \tilde{\mathcal{K}}_0, \mathcal{K}_1 - \tilde{\mathcal{K}}_1)\|_*,
$$

from which the Lipschitz continuity of $\mathcal{G}_{j,k}^{\varepsilon,\mathrm{chem}}(K_0, K_1)$ follows.

Now we look at the Lipschitz continuity of $\mathcal{G}_{j,k}^{\mathrm{KS}}(K_0, K_1)$. It holds that

$$
\begin{aligned}
|\mathcal{G}_{j,k}^{\mathrm{KS}}(K_0, K_1) - \mathcal{G}_{j,k}^{\mathrm{KS}}(\tilde{K}_0, \tilde{K}_1)| &\leq \|\chi_j\|_{L^2} \|\rho^{(k)} - \tilde{\rho}^{(k)}\|_{L^2} \\
&\leq C_x \underbrace{\|\rho^{(k)} - \tilde{\rho}^{(k)}\|_{L^2}}_{(*)}.
\end{aligned}
$$

Taking a closer look at $(*)$ yields the following. Let $\bar{\rho}^{(k)} = \rho^{(k)} - \tilde{\rho}^{(k)}$. Then $\bar{\rho}^{(k)}$ fulfils

$$
\partial_t \bar{\rho}^{(k)} = \nabla\Big(\tilde{D} \nabla \bar{\rho}^{(k)} - \tilde{\Gamma} \bar{\rho}^{(k)} \Big) + r(1 - \frac{\rho + \tilde{\rho}}{c}) \bar{\rho}^{(k)} + \underbrace{\nabla\Big((\overbrace{D - \tilde{D}}^{=\bar{D}}) \nabla \rho^{(k)} - (\overbrace{\Gamma - \tilde{\Gamma}}^{=\bar{\Gamma}}) \rho^{(k)} \Big)}_{\text{Source-term}}
$$

with initial condition $\bar{\rho}^{(k)}(0,x) = 0$. We used that

$$\begin{aligned} b(\rho)\rho - b(\tilde{\rho})\tilde{\rho} &= r(1-\frac{\rho}{c})\rho - r(1-\frac{\tilde{\rho}}{c})\tilde{\rho} \\ &= r(\rho-\tilde{\rho}) - \frac{r}{c}(\rho\rho - \tilde{\rho}\tilde{\rho}) \\ &= r(\rho-\tilde{\rho}) - \frac{r}{c}(\rho\rho - \tilde{\rho}\tilde{\rho} + \rho\tilde{\rho} - \rho\tilde{\rho}) \\ &= r(\rho-\tilde{\rho}) - \frac{r}{c}((\rho-\tilde{\rho})\tilde{\rho} + (\rho-\tilde{\rho})\rho) \\ &= r\bar{\rho} - \frac{r}{c}(\rho-\tilde{\rho})(\rho+\tilde{\rho}) \\ &= r\bar{\rho} - \frac{r}{c}(\rho+\tilde{\rho})\bar{\rho}. \end{aligned}$$

Since the growth term $r(1-\frac{\rho+\tilde{\rho}}{c})$ does not depend directly on $\bar{\rho}$ we can use Theorem 4.5, which leads to

$$\|\bar{\rho}^{(k)}\|_{L^\infty((0,T);L^2(\mathbb{R}^3))} \leq C(T,D,\Gamma)\big(\|\text{Source}-\text{term}\|_{L^2((0,T);H^{-1})} + \|\text{initial cond.}\|_{L^2(\mathbb{R}^3)}\big). \tag{4.10}$$

Theorem 4.5 can be applied here for $\mathbb{R}^3$ instead of a bounded and open subset Ω since we assume the initial condition to have compact support and the birth-death term is bounded, thus the support of ρ stays finite for finite time.

In order to get an estimate with (K_0, K_1) and $(\tilde{K}_0, \tilde{K}_1)$, we have to investigate the source term further.

Following [24, Chapter 4.3], we can estimate

$$\|\nabla\Big(\bar{D}\nabla\rho^{(k)} - \bar{\Gamma}\rho^{(k)}\Big)\|_{L^2((0,T);H^{-1})} \leq \int_0^T \sum_{i=0}^n \int_{\mathbb{R}^3}\Big(\bar{D}\nabla\rho^{(k)} - \bar{\Gamma}\rho^{(k)}\Big)_i^2 \mathrm{d}x\,\mathrm{d}t \tag{4.11}$$

where we use

$$g = g_0 + \sum_{i=1}^n \partial_i g_i, \quad g_0, g_i \in L^2(\mathbb{R}^3)$$

for g being an element of the Sobolev space $H^{-1}(\mathbb{R}^3)$ [24, Theorem 4.7]. Thus, we want to estimate the matrix norms $\|\cdot\|_2$ of $\bar{D}$ and $\bar{\Gamma}$.

$$\begin{aligned} \|\bar{D}\|_2 &= \|\int_V v\otimes(\kappa-\tilde{\kappa})\mathrm{d}v\|_2 \\ &\overset{(\Delta)}{\leq} |V|\int_V \|v\otimes(\kappa-\tilde{\kappa})\|_2 \mathrm{d}v \\ &= |V|\int_V \|v\|_2\|\kappa-\tilde{\kappa}\|_2 \mathrm{d}v \\ &\leq |V|\|v\|_{L^2}\|F\|_\infty\|\kappa-\tilde{\kappa}\|_{L^2(V;\mathrm{d}v/F)} \end{aligned}$$

where in (Δ) we used Jensen's inequality. Further, from the definition of $\kappa, \tilde{\kappa}$

$$\mathcal{K}_0(\kappa) = vF \quad \text{and} \quad \tilde{\mathcal{K}}_0(\tilde{\kappa}) = vF$$

it follows that

$$\begin{aligned} 0 &= \mathcal{K}_0(\kappa) - \tilde{\mathcal{K}}_0(\tilde{\kappa}) \\ &= \mathcal{K}_0(\kappa) - \tilde{\mathcal{K}}_0(\tilde{\kappa}) + \tilde{\mathcal{K}}_0(\kappa) - \tilde{\mathcal{K}}_0(\kappa) \end{aligned}$$

$$= \tilde{\mathcal{K}}_0(\kappa - \tilde{\kappa}) + (\mathcal{K}_0 - \tilde{\mathcal{K}}_0)(\kappa)$$
$$= \tilde{\mathcal{K}}_0(\kappa - \tilde{\kappa}) + \bar{\mathcal{K}}_0(\kappa).$$

Thus, $\kappa - \tilde{\kappa}$ solves

$$\tilde{\mathcal{K}}_0(\kappa - \tilde{\kappa}) = -\bar{\mathcal{K}}_0(\kappa).$$

From Theorem 4.2 (Lax-Milgram) and (3.21) we have that

$$\|\kappa\|_{L^2(V;\mathrm{d}v/F)} \leq \frac{\|F\|_\infty}{\tilde{\beta}} \|vF\|_{L^2(V;\mathrm{d}v/F)} \tag{4.12}$$

and

$$\|\kappa - \tilde{\kappa}\|_{L^2(V,\frac{\mathrm{d}v}{F})} \leq \frac{\|F\|_\infty}{\beta} \|\bar{\mathcal{K}}_0(\kappa)\|_{L^2(V;\mathrm{d}v/F)}. \tag{4.13}$$

Moreover, it holds that

$$\begin{aligned}
\|\bar{\mathcal{K}}_0(\kappa)\|^2_{L^2(V,\mathrm{d}v)/F} &= \int_V \frac{\left(\bar{\mathcal{K}}_0(\kappa)\right)^2}{F}\mathrm{d}v \\
&= \int_V \frac{\left(\int_V \bar{K}_0\kappa' - \bar{K}_0'\kappa \mathrm{d}v'\right)^2}{F}\mathrm{d}v \\
&\overset{(*)}{\leq} 2\int_V \frac{1}{F}\left(\int_V \left(\bar{K}_0\kappa'\right)^2 \mathrm{d}v' + \int_V \left(\bar{K}_0'\right)^2 \mathrm{d}v'\kappa^2\right)\mathrm{d}v \\
&\leq 2\max|\bar{K}_0|^2\left(\int_V \frac{1}{F}\mathrm{d}v \cdot \|F\|_\infty \int_V \frac{(\kappa')^2}{F'}\mathrm{d}v' + |V|\int_V \frac{\kappa^2}{F}\mathrm{d}v\right) \\
&\leq 2\max|\bar{K}_0|^2\left(\frac{\|F\|_\infty}{\min F}|V|\|\kappa\|^2_{L^2(V;\mathrm{d}v/F)} + |V|\|\kappa\|^2_{L^2(V;\mathrm{d}v/F)}\right) \\
&\leq 2\|(K_0, K_1) - (\tilde{K}_0, \tilde{K}_1)\|_* \underbrace{\left(\frac{\|F\|_\infty}{\min F} + 1\right)|V|\frac{\|F\|_\infty}{\tilde{\beta}}\|vF\|_{L^2}}_{:=\tilde{\nu}/2}.
\end{aligned}$$

for the norm $\|(K_0, K_1)\|_* = \max(\|K_0\|_\infty, \|K_1\|_\infty)$. Note, that F is bounded away from zero and that

$$\int_V (\kappa')^2 \mathrm{d}v \leq \|F\|_\infty \int_V \frac{(\kappa')^2}{F(v')}\mathrm{d}v' \quad \text{and} \quad \int_V \frac{1}{F(v)}\mathrm{d}v \leq |V|\frac{1}{\min F}.$$

holds. In $(*)$ we again used Cauchy-Schwarz inequality, as in the proof of Proposition 4.4 and we also used (4.12) in order to estimate $\|\kappa\|^2_{L^2(V;\mathrm{d}v/F)}$. It follows, that

$$\begin{aligned}
\|\bar{D}\|_2 &\leq |V|\|v\|_{L^2}\|F\|_\infty\left(\tilde{\nu}\frac{1}{\beta}\|(K_0, K_1) - (\tilde{K}_0, \tilde{K}_1)\|_*\right) \\
&:= \nu\|(K_0, K_1) - (\tilde{K}_0, \tilde{K}_1)\|_*.
\end{aligned}$$

Now, the norm of the drift vector is estimated as

$$\begin{aligned}
\|\bar{\Gamma}\|_2 &= \|\int_V v(\Theta - \tilde{\Theta})\mathrm{d}v\|_2 \\
&\overset{(\Delta)}{\leq} |V|\int_V \|v\|_2\|\Theta - \tilde{\Theta}\|_2 \mathrm{d}v \\
&\leq |V|\|v\|_{L^2}\|F\|_\infty\|\Theta - \tilde{\Theta}\|_{L^2(V;\mathrm{d}v/F)}
\end{aligned}$$

where in (Δ) we used Jensen's inequality since the norm $\|\cdot\|_2$ is convex.
Using Lax-Milgram yields

$$\begin{aligned}
\|\Theta\|^2_{L^2(V;\mathrm{d}v/F)} &\leq \frac{\|F\|_\infty}{\delta}\|\mathcal{K}_1(F)\|^2_{L^2(V;\mathrm{d}v/F)} \\
&\leq \frac{\|F\|_\infty}{\delta}\int_V \frac{\left(\int_V K_1F' - K_1'F\mathrm{d}v'\right)^2}{F}\mathrm{d}v \\
&\overset{(*)}{\leq} 2\frac{\|F\|_\infty}{\delta}\int_V \frac{\left(\int_V K_1F'\mathrm{d}v'\right)^2 + \left(\int_V K_1'F\mathrm{d}v'\right)^2}{F}\mathrm{d}v \\
&\leq 2\frac{\|F\|_\infty}{\delta}C^2\int_V \frac{\left(\int_V F(v')\mathrm{d}v'\right)^2 + |V|^2F^2\right)}{F}\mathrm{d}v \\
&\leq 2\frac{\|F\|_\infty}{\delta}C^2\left(\frac{|V|}{\min F} + |V|^2\right)
\end{aligned}$$

with $\|F\|_{L^2(V;\mathrm{d}v/F)} = \int_V \frac{F^2}{F}\mathrm{d}v = \int_V F\mathrm{d}v = 1$ and F being bounded away from zero by assumption as well as $\int_V \frac{1}{F}\mathrm{d}v \leq \frac{|V|}{\min F}$. In $(*)$ we again used Cauchy-Schwarz inequality, as in the proof of Proposition 4.4. Thus, we have

$$\|\Theta\|_{L^2(V;\mathrm{d}v/F)} \leq \frac{\|F\|_\infty}{\tilde{\delta}}C\sqrt{\frac{|V|}{\min F} + |V|^2} := \tilde{\eta}.$$

Additionally, $\Theta - \tilde{\Theta}$ solves

$$\tilde{\mathcal{K}}_0(\Theta - \tilde{\Theta}) + (\mathcal{K}_0 - \tilde{\mathcal{K}}_0)(\Theta) = \mathcal{K}_0(\Theta) - \tilde{\mathcal{K}}_0(\tilde{\Theta}) = \mathcal{K}_1(F) - \tilde{\mathcal{K}}_1(F) = \bar{\mathcal{K}}_1(F),$$

leading to

$$\tilde{\mathcal{K}}_0(\Theta - \tilde{\Theta}) = \bar{\mathcal{K}}_1(F) - \bar{\mathcal{K}}_0(\Theta).$$

Also with Lax-Milgram, it then follows

$$\begin{aligned}
\|\Theta - \tilde{\Theta}\|_{L^2(V;\mathrm{d}v/F)} &\leq \frac{\|F\|_\infty}{\delta}\left(\|\bar{\mathcal{K}}_0(\Theta)\|_{L^2(V;\mathrm{d}v/F)} + \|\bar{\mathcal{K}}_1(F)\|_{L^2(V;\mathrm{d}v/F)}\right) \\
&\leq \frac{\|F\|_\infty}{\delta}\sqrt{\frac{\|F\|_\infty}{\min F} + 1}\sqrt{|V|}\left(\|\Theta\|_{L^2(V;\mathrm{d}v/F)}\max|\bar{K}_0| + \|F\|_{L^2(V;\mathrm{d}v/F)}\max|\bar{K}_1|\right) \\
&\leq \underbrace{\frac{\|F\|_\infty}{\delta}\sqrt{\frac{\|F\|_\infty}{\min F} + 1}\sqrt{|V|}(\tilde{\eta} + 1)}_{:=\hat{\eta}}\|(K_0, K_1) - (\tilde{K}_0, \tilde{K}_1)\|_*,
\end{aligned}$$

where the coefficient $\sqrt{\frac{\|F\|_\infty}{\min F} + 1}\sqrt{|V|}$ can be derived in a similar fashion as has been done previously for $\bar{\mathcal{K}}_0(\kappa)$. Therefore,

$$\begin{aligned}
\|\bar{\Gamma}\|_2 &\leq |V|\|v\|_{L^2}\|F\|_\infty\left(\hat{\eta}\|(K_0, K_1) - (\tilde{K}_0, \tilde{K}_1)\|_*\right) \\
&:= \eta\|(K_0, K_1) - (\tilde{K}_0, \tilde{K}_1)\|_*
\end{aligned}$$

All together with $\xi := \max(\eta, \nu)$ this results in

$$\begin{aligned}
\|\nabla\left(\bar{D}\nabla\rho^{(k)} - \bar{\Gamma}\rho^{(k)}\right)\|_{L^2((0,T);H^{-1})} &= \int_0^T \|\nabla\left(\bar{D}\nabla\rho^{(k)} - \bar{\Gamma}\rho^{(k)}\right)\|_{H^{-1}}\mathrm{d}t \\
&\overset{(*)}{\leq} \int_0^T \sum_{i=1}^n \int_{\mathbb{R}^3}\left(\bar{D}\nabla\rho^{(k)} - \bar{\Gamma}\rho^{(k)}\right)_i^2\mathrm{d}x\,\mathrm{d}t
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^T \int_{\mathbb{R}^3} \|\bar{D}\nabla\rho^{(k)} - \bar{\Gamma}\rho^{(k)}\|_2^2 \mathrm{d}x\,\mathrm{d}t \\
&\leq \int_0^T \int_{\mathbb{R}^3} 2\Big(\|\bar{D}\|_2^2\|\nabla\rho^{(k)}\|_2^2 + \|\bar{\Gamma}\|_2^2\|\rho^{(k)}\|_2^2\Big)\mathrm{d}x\,\mathrm{d}t \\
&\leq \int_0^T \int_{\mathbb{R}^3} \Big(\|\nabla\rho^{(k)}\|_2^2 + \|\rho^{(k)}\|_2^2\Big)\big(\xi\|(\bar{K}_0,\bar{K}_1)\|_*^2\big)\mathrm{d}x\,\mathrm{d}t \\
&= \|\rho^{(k)}\|^2_{L^2((0,T);H_0^1)}\big(\xi\|(\bar{K}_0,\bar{K}_1)\|_*^2\big) \\
&\leq \xi\tilde{C}^2\|(\rho^*)^{(k)}\|_{L^2}\|(\bar{K}_0,\bar{K}_1)\|_*^2
\end{aligned}
$$

where

$$
\|(\bar{K}_0,\bar{K}_1)\|_*^2 = \|(K_0,K_1) - (\tilde{K}_0,\tilde{K}_1)\|_*^2
$$

and in $(*)$ we used (4.11). Overall, we have that

$$
\|\rho^{(k)} - \tilde{\rho}^{(k)}\|_{L^2(\mathbb{R}^3)} \leq \xi\tilde{C}^2\|(\rho^*)^{(k)}\|_{L^2}\|(K_0,K_1) - (\tilde{K}_0,\tilde{K}_1)\|_*^2
$$

Integrating in time proves the Lipschitz continuity of $\mathcal{G}_{j,k}^{\mathrm{KS}}$.

(iii) From the boundedness of the measurements shown in (i) it follows that the likelihood

$$
\mu_\circ^{(K_0,K_1)}(y) = \exp\Big(-\frac{1}{2\gamma^2}\|\mathcal{G}^\circ(K_0,K_1) - y\|^2\Big)
$$

and normalization constants Z° are bounded away from zero and are bounded uniformly in $\mathcal{A}$ and ε. From the Lipschitz-continuity of the measurements shown in (ii) the measurability of the likelihoods follows. Thus, the posterior distributions are well-defined and absolutely continuous with respect to each other. The well-posedness of the distributions follows from the continuous dependency of the likelihoods on the data y. □

Now, the convergence of the two measurements can be shown.

Corollary 4.7 *Suppose that the initial condition $f_0(x,v) \in C_c^1(\mathbb{R}^3 \times V)$ and the test functions $\chi_j(x) \in C_c(\mathbb{R}^3)$ fulfil condition (4.4). Then the kinetic measurement $\mathcal{G}_{j,k}^{\varepsilon,\mathrm{chem}}$ converges to $\mathcal{G}_{j,k}^{\mathrm{KS}}$ uniformly in $\mathcal{A}$ for $\varepsilon \to 0$.*

Proof: The convergence of the measurements is a direct consequence of the convergence of f_ε to ρF in $L^\infty\big([0,T], L_+^1 \cap L^\infty(\mathbb{R}^3 \times V)\big)$ as shown in Theorem 3.8. Then it holds that

$$
\begin{aligned}
|\mathcal{G}_{j,k}^{\varepsilon,\mathrm{chem}} - \mathcal{G}_{j,k}^{\mathrm{KS}}| &= |\int_{\mathbb{R}^3}\int_V f_\varepsilon^{(k)}(t_j,x,v)\mathrm{d}v\chi_j(x)\mathrm{d}x - \int_{\mathbb{R}^3}\rho^{(k)}(t_j,x)\chi_j(x)\mathrm{d}x| \\
&\leq \int_{\mathbb{R}^3}\int_V |f_\varepsilon^{(k)}(t_j,x,v) - \rho^{(k)}(t_j,x)F(v)|\mathrm{d}v|\chi_j(x)|\mathrm{d}x \\
&\leq \|f_\varepsilon^{(k)}(t_j,\cdot,\cdot) - \rho^{(k)}(t_j,\cdot)F(\cdot)\|_{L^\infty(\mathbb{R}^3\times V)}|V|\|\chi_j(x)\|_{L^1(\mathbb{R}^3)} \\
&\to 0,
\end{aligned}
$$

where we used that

$$
\int_{\mathbb{R}^3}\rho^{(k)}(t_j,x)\chi_j(x)\mathrm{d}x = \int_{\mathbb{R}^3}\rho^{(k)}(t_j,x)\chi_j(x)\mathrm{d}x\int_V F(v)\mathrm{d}v = \int_{\mathbb{R}^3}\int_V \rho^{(k)}(t_j,x)F(v)\mathrm{d}v\chi_j(x)\mathrm{d}x
$$

because of the assumption on F to fulfil $\int_V F\mathrm{d}v = 1$. Since the convergence of f_ε to ρF is uniformly, the calculations above are uniform in $\mathcal{A}$. The uniform convergence over (j,k) follows from the boundedness of the initial data and the test functions. □

Now, all preconditions are proven that are needed to show the convergence of the posterior distributions. As already mentioned, one issue may arise with the metric one chooses to compare the distributions with each other. Here, we will choose two metrics to calculate the difference between the posterior distributions, as done in [17]. Both metrics require the to-be-compared probability measurements μ_1 and μ_2 to be either absolutely continuous with respect to each other or with respect to a third measure μ_0, in order to evaluate the distance between them. The first is called the *Kullback-Leibler divergence*

$$d_{\mathrm{KL}}(\mu_1, \mu_2) := \int_{\mathcal{A}} \left(\log \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}(u) \right) \mathrm{d}\mu_2(\mathrm{u}). \tag{4.14}$$

While this is in fact no metric, since it lacks the symmetry and triangle-equality properties, it is still used widely due to its close connection to several information concepts such as the Fisher information metric or the Shannon entropy [27]. The second metric - that is indeed a metric - is called *Hellinger metric*

$$d^2_{\mathrm{Hell}}(\mu_1, \mu_2) := \frac{1}{2} \int_{\mathcal{A}} \left(\sqrt{\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}(u)} - \sqrt{\frac{\mathrm{d}\mu_2}{\mathrm{d}\mu_0}(u)} \right)^2 \mathrm{d}\mu_0(\mathrm{u}), \tag{4.15}$$

where μ_1 and μ_2 are absolutely continuous with respect to μ_0. Convergence in this metric also implies that the expectations of any polynomially bounded function converges with respect to either of their posterior distributions, as explained in [36]. For instance, the mean, covariance and higher moments of the distribution converge.

The following theorem can be proven analogously to [17, Theorem 2] and [31, Theorem 3.2].

Theorem 4.8 *Consider the measurements of the macroscopic bacteria density given by* (4.1) *in case of the kinetic model and* (4.2) *in case of the Keller-Segel-Fisher model respectively. Suppose that the initial condition* $f_0(x, v) \in C^1_c(\mathbb{R}^3 \times V)$ *and the test functions* $\chi_j(x) \in C_c(\mathbb{R}^3)$ *fulfil condition* (4.4). *For a given prior* μ_0 *on* $\mathcal{A}$ *and an additive Gaussian noise in the data the two posterior distributions coming from the kinetic chemotaxis equation and the macroscopic Keller-Segel-Fisher equation are asymptotically equivalent in the Kullback-Leibler divergence*

$$d_{\mathrm{KL}}(\mu^y_{\varepsilon,\mathrm{chem}}, \mu^y_{\mathrm{KS}}) \xrightarrow{\varepsilon \to 0} 0. \tag{4.16}$$

Proof: First, the order of the Kullback-Leibler divergence can be calculated via Bayes' rule by

$$\begin{aligned}
\log \frac{\mathrm{d}\mu^y_{\varepsilon,\mathrm{chem}}}{\mathrm{d}\mu^y_{\mathrm{KS}}}(K_0, K_1) &= \log \left(\frac{\mu_0(K_0, K_1) \mu^{(K_0,K_1)}_{\varepsilon,\mathrm{chem}}(y)}{Z^{\varepsilon,\mathrm{chem}}} \frac{Z^{\mathrm{KS}}}{\mu_0(K_0, K_1) \mu^{(K_0,K_1)}_{\mathrm{KS}}(y)} \right) \\
&= \log \frac{Z^{KS}}{Z^{\varepsilon,\mathrm{chem}}} + \log \frac{\mu^{(K_0,K_1)}_{\varepsilon,\mathrm{chem}}(y)}{\mu^{(K_0,K_1)}_{\mathrm{KS}}(y)} \\
&= \mathcal{O}\big(|Z^{\varepsilon,\mathrm{chem}} - Z^{KS}|\big) + \mathcal{O}\Big(|\mu^{(K_0,K_1)}_{\varepsilon,\mathrm{chem}}(y) - \mu^{(K_0,K_1)}_{\mathrm{KS}}(y)|\Big) \\
&= \mathcal{O}\Big(|\mu^{(K_0,K_1)}_{\varepsilon,\mathrm{chem}}(y) - \mu^{(K_0,K_1)}_{\mathrm{KS}}(y)|\Big).
\end{aligned}$$

Therefore, the likelihoods have to be estimated via

$$\begin{aligned}
|\mu^{(K_0,K_1)}_{\varepsilon,\mathrm{chem}}(y) - \mu^{(K_0,K_1)}_{\mathrm{KS}}(y)| &= \left| \exp\left(-\frac{1}{\gamma^2} \|y - \mathcal{G}^{\varepsilon,\mathrm{chem}}(K_0, K_1)\|^2 \right) - \exp\left(-\frac{1}{\gamma^2} \|y - \mathcal{G}^{\mathrm{KS}}(K_0, K_1)\|^2 \right) \right| \\
&\leq c \left| -\frac{1}{\gamma^2} \|y - \mathcal{G}^{\varepsilon,\mathrm{chem}}(K_0, K_1)\|^2 - \frac{1}{\gamma^2} \|y - \mathcal{G}^{\mathrm{KS}}(K_0, K_1)\|^2 \right|,
\end{aligned}$$

where $c < \infty$ is the Lipschitz constant of the function $x \mapsto \exp\left(-\frac{|x|}{2\gamma^2}\right)$. Further it holds that

$$\begin{aligned}
&\left|\|y - \mathcal{G}^{\varepsilon,\mathrm{chem}}(K_0, K_1)\|^2 - \|y - \mathcal{G}^{\mathrm{KS}}(K_0, K_1)\|^2\right| \\
&= \left|\mathrm{tr}\Big[\big(2y - \mathcal{G}^{\varepsilon,\mathrm{chem}}(K_0, K_1) - \mathcal{G}^{\mathrm{KS}}(K_0, K_1)\big)^{\mathrm{T}}\big(\mathcal{G}^{\varepsilon,\mathrm{chem}}(K_0, K_1) - \mathcal{G}^{\mathrm{KS}}(K_0, K_1)\big)\Big]\right| \\
&\leq \|2y - \mathcal{G}^{\varepsilon,\mathrm{chem}}(K_0, K_1) - \mathcal{G}^{\mathrm{KS}}(K_0, K_1)\| \cdot \|\mathcal{G}^{\varepsilon,\mathrm{chem}}(K_0, K_1) - \mathcal{G}^{\mathrm{KS}}(K_0, K_1)\|
\end{aligned}$$

where the first factor is bounded uniformly in $\mathcal{A}$ and ε by Lemma 4.6. The second factor converges to 0 uniformly on $\mathcal{A}$ due to the convergence of the measurements as shown in Corollary 4.7. Overall, it follows that

$$d_{\mathrm{KL}}(\mu^y_{\varepsilon,\mathrm{chem}}, \mu^y_{\mathrm{KS}}) \xrightarrow{\varepsilon \to 0} 0. \qquad \square$$

As shown in [39, Lemma 2.4], the Hellinger metric is bounded by the Kullback-Leibler divergence

$$d^2_{\mathrm{Hell}}(\mu_1, \mu_2) \leq d_{\mathrm{KL}}(\mu_1, \mu_2)$$

which, together with the theorem above, shows the convergence in the Hellinger metric of the posterior distributions.

Corollary 4.9 *In the framework of Theorem 4.8, the posterior distributions are asymptotically equivalent in the Hellinger metric*

$$d_{\mathrm{Hell}}(\mu^y_{\varepsilon,\mathrm{chem}}, \mu^y_{\mathrm{KS}}) \xrightarrow{\varepsilon \to 0} 0.$$

4.3 Conclusion

In this chapter, the relation between the posterior distributions coming from the kinetic and macroscopic forward models was investigated by extending the results from [17] for models including a reaction term, namely the birth-death term. In order to show the uniform convergence of the distributions, assumptions were made on the information we put in the model, by imposing conditions on the tumbling kernel. With this, the well-posedness of the posterior distributions could be shown and that they converge in the asymptotic limit under the Kullback-Leibler divergence and the Hellinger metric. This does now confirm the intuition, that the convergence of the forward problems as shown in section 3 should translate into convergence of the inverse problems in the Bayesian setting.

5 Ill conditioning of Inverse Keller-Segel

Now that the convergence of the two inverse problems is shown we want to take a look at the stability of the inverse chemotaxis equation as $\varepsilon \to 0$ in the following section. This will be done for the case, where no reaction term occurs in the models. In the asymptotic limit the chemotaxis equation can be approximated by the Keller-Segel-equation as discussed in section 3 and [10] and the inverse problems converge as well, as shown in section 4 and [17]. Since our goal is to reconstruct the tumbling kernel - a kinetic quantity - it is intuitive that we lose some information when entering the macroscopic regime. In this chapter, the goal is to quantify this loss of information. In order to do so, we try to extend the results from [28], where the ill-conditioning of stationary inverse radiative transfer in the diffusive limit is being investigated, to the non-stationary case. On a global level, the difference between the prior distribution and the posterior distribution under the Kullback-Leibler divergence will be estimated. Since the posterior distribution takes the prior information as well as the data into account, the difference tells us how much the posterior distribution is influenced by the data. On a local level, we will take a closer look at the maximum a posteriori (MAP) point and more precise at the flatness of the distribution at this point. If the distribution is rather flat around the MAP point it means that different values for the quantity of interest are equally likely to actually be the value one is looking for. Therefore, it does not matter whether or not we are exactly at the MAP point and therefore the flatness characterizes the level of uncertainty of the maximizer. For both approaches the dependence on ε will be analysed, exploring the loss of information between the kinetic and macroscopic regime, following [28]. Within this work, the two presented approaches do not suffice to quantify the loss of information for the non-stationary chemotaxis equation in the diffusive limit.

In [28] those considerations were made regarding the stationary radiative transfer equation, which has a similar structure as our chemotaxis equation without a time-dependency and thus being a boundary-value problem instead of an initial-value problem. For this equation, which can be applied to the field of optical imaging for instance, the goal is to reconstruct certain properties of an examined tissue by sending near infrared light into the tissue and measuring the outgoing photons [31]. As one decreases the photon energy, and thus enters the diffusive limit, the picture becomes less crisp and the reconstruction more unstable. There exist various studies on the subject of loss of information in the diffusive limit for radiative transfer, for example one can find numerical observations in [1, 19] and a rigorous proof in [2]. The main problem with those results is, that they assume that we have access to the full Albedo Operator

$$\mathcal{A}_{K_0,K_1} : f_0 \mapsto \int_V f(t,x,v)\mathrm{d}v \tag{5.1}$$

which in reality is not the case. Knowing the full operator $\mathcal{A}$ would correspond to send in every set of initial data and take measurements over the whole space. But in the lab one only has a finite number of incoming data as well as a finite number of measurements, which is why the mentioned results have been extended to the discrete case in [28] and we follow those elaborations in this section.

For a finite series of initial values $\{f_0^{(k)}\}_k$ and blob functions $\{\chi_j\}_j$ the parameter-to-measurement map is defined as

$$\mathcal{G}_{j,k} : K(t,x,v,v') \mapsto \left\{ \int_{\mathbb{R}^3} \int_V f^{(k)}(t_j,x,v)\mathrm{d}v\, \chi_j(x)\, \mathrm{d}x \right\}_{j,k}, \tag{5.2}$$

where we will drop (j,k) in the notation of $\mathcal{G}$ if the statement is true for all (j,k). Additionally, we define the parameter-to-solution map as

$$\mathcal{S}_{j,k} : K(t,x) \mapsto f^{(k)}(t_j,x,v) \tag{5.3}$$

leading to the relation

$$\mathcal{G}(K) = \int_{\mathbb{R}^3} \int_V \mathcal{S}(K) \mathrm{d}v \chi_j(x) \mathrm{d}x. \tag{5.4}$$

In order to investigate the global loss of information, we want to estimate the Kullback-Leibler divergence by the derivative of the parameter-to-solution map. Following [28] we have the next proposition about the Kullback-Leibler divergence of the prior and posterior distribution.

Proposition 5.1 *For $K \in \mathcal{B}$ the Kullback-Leibler divergence between μ_{post} and μ_0 can be estimated via*

$$d_{KL}(\mu_0, \mu_{\text{post}}) \leq C \int_{\mathcal{A}} \int_{\mathcal{A}} |(\mathcal{G}(K') - \mathcal{G}(K))(\mathcal{G}(K) + \mathcal{G}(K') - 2y)| \, \mathrm{d}\mu_0(K') \, \mathrm{d}\mu_0(K) \tag{5.5}$$

for some positive constant C that does not depend on the admissible set $\mathcal{A}$.

Remark 5.2 As shown in the proof of [28, Theorem 2], we can rewrite Proposition 5.1 such that we get an estimate of the Kullback-Leibler divergence by the derivative of $\mathcal{G}'$, thus

$$\begin{aligned} d_{KL}(\mu_0, \mu_{\text{post}}) &\leq C \int_{\mathcal{A}} \int_{\mathcal{A}} |(\mathcal{G}(K') - \mathcal{G}(K))(\mathcal{G}(K) + \mathcal{G}(K') - 2y)| \, \mathrm{d}\mu_0(K') \, \mathrm{d}\mu_0(K) \\ &= \int_{\mathcal{A}} \int_{\mathcal{A}} \int_0^t |\mathcal{G}'\big(K + s(K' - K)\big)\big(K' - K\big)\big(\mathcal{G}(K) + \mathcal{G}(K') - 2y\big)| \mathrm{d}s \, \mathrm{d}\mu_0(K') \, \mathrm{d}\mu_0(K). \end{aligned}$$

Hence, to estimate the Kullback-Leibler divergence, the derivative of $\mathcal{G}(K)$ respectively the derivative of $\mathcal{S}(K)$ needs to be investigated. We call a variation $\tilde{K}$ of K admissible if $K + s\tilde{K} \in \mathcal{A}$ for all sufficiently small parameter s. For such a variation we have that

$$\begin{aligned} \mathcal{G}'_{jk}(K)\tilde{K} &= \lim_{s \to 0} \frac{1}{s} \int_V \int_{\mathbb{R}^3} \Big(f^{(k)}_{K + s\tilde{K}}(t_j, x, v) - f^{(k)}_K(t_j, x, v) \Big) \chi_j(x) \mathrm{d}x \, \mathrm{d}v \\ &= \int_V \int_{\mathbb{R}}^3 \mathcal{S}'_{jk}(K)\tilde{K} \mathrm{d}v \chi_j(x) \mathrm{d}x. \end{aligned} \tag{5.6}$$

Therefore, if G varies slowly over the admissible set $\mathcal{A}$, then the information gain from the data is small because μ_{post} and μ_0 only differ a little from each other. Thus the goal is to show that the derivative of $\mathcal{G}$ is of order ε, since we assume to have a loss of information.

Up until now, it was explained how to identify the global loss of information, but we are also interested in the local behaviour around the MAP point. If, for example, the distribution is flat around the MAP it means that the probability remains unchanged in a certain area around the MAP point. This results in the reconstruction being insensitive to the data illustrating the instability. Following [28, Chap. 5], an intuition for the characterisation of the flatness of a distribution will be derived. Suppose we have a linear problem, meaning

$$\mathcal{G}(K) = GK$$

and let the prior distribution μ_0 be Gaussian centered at K_0 with covariance C_0. Then the posterior distribution is uniquely determined by

$$K_{\text{post}} = C^{-1}_{\text{post}}\big(G^{\mathrm{T}} y + C_0^{-1} K_0\big), \quad C_{\text{post}} = \big(G^T G + C_0^{-1}\big)^{-1}$$

and its flatness is characterised by the covariance matrix

$$\int_{\mathcal{A}} \|K_{\text{post}} - K\|^2 \mathrm{d}\mu_{\text{post}} = \mathrm{tr}(C_{\text{post}}),$$

as stated in [28]. Thus, the less informative the forward map is, the smaller is G and the bigger $\mathrm{tr}(C_{\text{post}})$ gets, indicating the higher mean square error which geometrically corresponds to a flatter

Gaussian. In our case, the forward map is nonlinear in K, which is why the argument above only serves as a guidance. If we denote the posterior distribution as

$$\mu_{\text{post}} = \frac{1}{Z}\exp(-\|\mathcal{G}(K) - y\|^2/2)\mu_0 \propto \exp(-A),$$

its convexity is determined by the Hessian of A. From [28, Prop. 7] we have the following result.

Proposition 5.3 *Let K_* be admissible and $\hat{K}$ be an admissible variation. Then we can express the Hessian of the posterior distribution A in terms of the forward map $\mathcal{G}$ via*

$$A''(K_*)[\hat{K},\hat{K}] = \frac{1}{2}\sum_i \big(\mathcal{G}_i'(K_*)[\hat{K}]^T\mathcal{G}_i'(K_*)[\hat{K}] + \mathcal{G}_i(K)^T\mathcal{G}_i''(K_*)[\hat{K},\hat{K}] - y_i^T\mathcal{G}_i''(K_*)[\hat{K},\hat{K}]\big). \quad (5.7)$$

Since this holds true for all admissible K_* it is also valid for the MAP point. Therefore, in order to show the flatness of the distribution we need to show that $\mathcal{G}''$ is small.

In order to have a shorter notation we introduce the abbreviations

$$f = f_K(t,x,v), \qquad f_s = f_{K+s\tilde{K}}(t,x,v), \qquad w = \mathcal{S}'(K)\tilde{K} \quad (5.8)$$

and if the calculations hold true for all j and k the subscripts will be omitted.

5.1 Velocity Independent Case

First, we will simplify our equation and only look at tumbling kernels, that do not depend on the velocity, meaning $K(t,x,v,v') = K(t,x)$. Therefore, the tumbling operator $\mathcal{K}$ is reduced to

$$K(t,x)\mathcal{L}(g) := K(t,x)\Big(\int_V g\,\mathrm{d}v - g\Big)$$

for $g \in L^2(V;\mathrm{d}v)$. Also, the admissible set $\mathcal{A}$ can be replaced by

$$\mathcal{B} = \big\{K \in C^1([0,\infty)\times\mathbb{R}^3) \;\big|\; 0 < \alpha \le K(t,x) \text{ and } \|K\|_{L^\infty(\mathbb{R}^3)} \le C_{\mathcal{B}}\big\}. \quad (5.9)$$

To investigate the dependence of the derivative of the parameter-to-solution map on ε, the drift diffusion limit for the case where $K(t,x,v,v') = K(t,x)$ will be shortly derived, following [28, Theorem 1] and the elaborations in subsection 3.3.

Proposition 5.4 *Consider the scaled chemotaxis equation*

$$\begin{aligned}\varepsilon^2\partial_t f(t,x,v) + \varepsilon v\cdot\nabla_x f(t,x,v) &= K(t,x)\Big(\int_V f(t,x,v')\mathrm{d}v' - f(t,x,v)\Big)\\ f(0,x,v) &= f_0(x,v). \end{aligned} \quad (5.10)$$

Then $f(t,x,v) \to \rho(t,x)$ as $\varepsilon \to 0$ where ρ solves

$$\begin{aligned}\partial_t\rho(t,x) - C_d\nabla_x\frac{1}{K}\nabla_x\rho(t,x) &= 0\\ \rho(0,x) &= \rho_0(x).\end{aligned}$$

Proof: Consider the expansion $f = f_0 + \varepsilon f_1 + \varepsilon^2 f_2 + \varepsilon^3 f_r$ and compare orders in (5.10). For $\mathcal{O}(1)$ we have that

$$\begin{aligned} &K\mathcal{L}f_0 = 0\\ \Leftrightarrow &\int_V f_0(t,x,v')\mathrm{d}v' = f_0(t,x,v)\end{aligned}$$

meaning that $f_0(t,x,v) = f_0(t,x) := \rho_0(t,x)$. Note, that we used $\int_V 1 \mathrm{d}v = 1$ since we consider V to be the unit sphere in $\mathbb{R}^3$. For $\mathcal{O}(\varepsilon)$ it holds that

$$v \cdot \nabla_x f_0(t,x) = K\mathcal{L} f_1 \tag{5.11}$$

which holds true for $f_1(t,x,v) = -\frac{1}{K} v \cdot \nabla_x \rho_0(t,x)$, since $\mathcal{L}(v) = -v$ and the other terms do not depend on v. Looking at $\mathcal{O}(\varepsilon^2)$ yields the relation

$$K\mathcal{L}(f_2) = \partial_t f_0 + v \cdot \nabla_x f_1 \tag{5.12}$$

$$\Leftrightarrow f_2 = \mathcal{L}^{-1}\left(\frac{1}{K}\partial_t f_0 + \frac{1}{K} v \cdot \nabla_x f_1\right). \tag{5.13}$$

In order to f_2 being well-defined, $\frac{1}{K}(\partial_t f_0 + v \cdot \nabla_x f_1)$ needs to lie within the range of $\mathcal{L}$. With the condition

$$-\int_V \frac{v}{K} \nabla_x f_1 \mathrm{d}v = \frac{1}{K}\partial_t f_0 \tag{5.14}$$

$$\Leftrightarrow \int_V \frac{v}{K} \nabla_x \left(\frac{v}{K} \nabla_x \rho_0 \mathrm{d}v\right) = \frac{1}{K}\partial_t \rho_0, \tag{5.15}$$

f_2 is well defined since we have the inverse image

$$\mathcal{L}\left(-\frac{1}{K}\partial_t f_0 - \frac{1}{K} v \cdot \nabla_x f_1\right) = \overbrace{\mathcal{L}\left(-\frac{1}{K}\partial_t f_0\right)}^{=0} + \mathcal{L}\left(-\frac{v}{K}\nabla_x f_1\right)) \tag{5.16}$$

$$= \int_V -\frac{v}{K}\nabla_x f_1 \mathrm{d}v - \frac{1}{K}(-v \cdot \nabla_x f_1) \tag{5.17}$$

$$\overset{(5.14)}{=} \frac{1}{K}\partial_t f_0 + \frac{1}{K} v \cdot \nabla_x f_1. \tag{5.18}$$

The condition (5.15) equipped with suited initial data determines the function ρ_0.

Now it remains to show, of which order f_r is. Rearranging (5.10) and dividing by ε^3 yields

$$\begin{aligned}
\varepsilon^2 \partial_t f_r(t,x,v) + \varepsilon v \cdot \nabla_x f_r(t,x,v) - K(t,x)\mathcal{L}(f_r) &= -\left(\frac{1}{\varepsilon}\partial_t f_0(t,x) + \frac{1}{\varepsilon^2} v \cdot \nabla_x f_0(t,x)\right) \\
&\quad - \left(\partial_t f_1(t,x,v) + \frac{1}{\varepsilon} v \cdot \nabla_x f_1(t,x,v) - \frac{1}{\varepsilon^2} K(t,x)\mathcal{L}(f_1)\right) \\
&\quad - \left(\varepsilon \partial_t f_2(t,x,v) + v \cdot \nabla_x f_2(t,x,v) - \frac{1}{\varepsilon} K(t,x)\mathcal{L}(f_2)\right) \\
&= -\varepsilon \partial_t f_2(t,x,v) - v \cdot \nabla_x f_2(t,x,v) - \partial_t f_1(t,x,v)
\end{aligned}$$

where we used the relations (5.11) and (5.12). We impose $f_r(0,x,v) = -\varepsilon^2 f_1(0,x,v) - \varepsilon f_2(0,x,v)$ as initial condition. Thus, f_r fulfils a chemotaxis equation with $\mathcal{O}(\varepsilon)$ initial condition and $\mathcal{O}(1)$ source term, hence from applying Lemma 3.7 (Grönwall's inequality) it follows that $\|f_r\|_{L^\infty(\mathbb{R}^3 \times V)} \leq \mathcal{O}(1)$. Looking at $\mathcal{O}(\varepsilon^2)$ gives

$$\partial_t f_0(t,x) + v \cdot \nabla_x f_1(t,x,v) = K\mathcal{L}(f_2).$$

Integrating the equation in v and inserting the formula for $f_1(t,x,v)$ leads to

$$\begin{aligned}
\partial_t \rho_0(t,x) - \nabla_x \frac{1}{K} \int_V v^2 \mathrm{d}v \cdot \nabla_x \rho_0(t,x)) = 0 \\
\Rightarrow \partial_t \rho_0(t,x) - C_d \nabla_x \frac{1}{K} \nabla_x \rho_0(t,x) = 0
\end{aligned}$$

where $C_d = \int_V v^2 \mathrm{d}v$. □

5.1.1 Global Loss of Information

The proposition below characterizes the derivative of the parameter-to-solution map.

Proposition 5.5 *Denote*

$$\mathcal{T}f = \varepsilon^2 \partial_t f + \varepsilon v \cdot \nabla_x f, \qquad C_K f = -K(t,x)\mathcal{L}f \tag{5.19}$$

with $\mathcal{L}f = \int_V f(t,x,v)\mathrm{d}v - f$. *For* $K \in \mathcal{B}$ *and an admissible variation* $\tilde{K}$, *the derivative of the parameter-to-solution map* w *is the unique solution of*

$$\begin{aligned} \mathcal{T}w + C_K w &= -C_{\tilde{K}} f, \\ w(0,x,v) &= 0, \end{aligned} \tag{5.20}$$

with $f \in \mathcal{F}$ *being the solution of the chemotaxis equation with parameter* K. *Also, for* $\varepsilon = 1$ *it holds that*

$$\|w\|_{L^\infty(0,T;L^\infty(\mathbb{R}^3\times V))} \le C(C_{\mathcal{B}}, T)\|f\|_{L^\infty(0,T;L^\infty(\mathbb{R}^3\times V))} \tag{5.21}$$

where $C_{\mathcal{B}}$ *denotes the constant appearing in the definition of the admissible set* $\mathcal{B}$ *as defined in* (5.9).

Proof: Let f be the solution of $\mathcal{T}f + C_K f = 0$ and f_s of $\mathcal{T}f_s + C_{K+s\hat{K}} f_s = 0$, where s describes the variation of the parameter K. Then it holds that

$$\mathcal{T}f_s + C_{K+s\hat{K}} f_s = \mathcal{T}f_s + C_K f_s + sC_{\hat{K}} f_s,$$

since C_K is a linear operator with respect to the parameter K. Subtracting the equation for f by the one for f_s yields

$$\begin{aligned} \mathcal{T}(f - f_s) + C_K(f - f_s) &= -sC_{\hat{K}} f_s \\ \Leftrightarrow \mathcal{T}w_s + C_K w_s &= -C_{\hat{K}} f_s \end{aligned}$$

with $w_s = \frac{1}{s}(f - f_s)$. Carrying out the limit $s \to 0$ shows the result, where uniqueness follows by standard methods.

In order to show the boundedness we use that

$$|\mathcal{L}(w)| \le 2\|w\|_{L^\infty(\mathbb{R}^3\times V)}, \quad |\mathcal{L}(f)| \le 2\|f\|_{L^\infty(\mathbb{R}^3\times V)}$$

and integrate the equation in time and integrate along the characteristic $(t-s, x - \frac{vs}{\varepsilon}, v)$ leading to

$$\begin{aligned} w(t,x,v) &= \overbrace{w(0,x,v)}^{=0} + \int_0^t K(t-s, x-\frac{vs}{\varepsilon})\mathcal{L}(w)(t-s, x-\frac{vs}{\varepsilon}, v)\mathrm{d}s \\ &\quad + \int_0^t K(t-s, x-\frac{vs}{\varepsilon})\mathcal{L}(f)(t-s, x-\frac{vs}{\varepsilon}, v)\mathrm{d}s \\ \Leftrightarrow |w(t,x,v)| &= \Big| \int_0^t K(t-s, x-\frac{vs}{\varepsilon})\mathcal{L}(w)(t-s, x-\frac{vs}{\varepsilon}, v)\mathrm{d}s \\ &\quad + \int_0^t K(t-s, x-\frac{vs}{\varepsilon})\mathcal{L}(f)(t-s, x-\frac{vs}{\varepsilon}, v)\mathrm{d}s \Big| \\ &\le \int_0^t |K(t-s, x-\frac{vs}{\varepsilon})| \cdot |\mathcal{L}(w)(t-s, x-\frac{vs}{\varepsilon}, v)|\mathrm{d}s \\ &\quad + \int_0^t |K(t-s, x-\frac{vs}{\varepsilon})| \cdot |\mathcal{L}(f)(t-s, x-\frac{vs}{\varepsilon}, v)|\mathrm{d}s \\ &\le C_{\mathcal{B}} \int_0^t |\mathcal{L}(w)(t-s, x-\frac{vs}{\varepsilon}, v)|\mathrm{d}s + C_{\mathcal{B}} \int_0^t |\mathcal{L}(f)(t-s, x-\frac{vs}{\varepsilon}, v)|\mathrm{d}s \end{aligned}$$

$$\Leftrightarrow \|w(t,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \le 2C_{\mathcal{B}} \int_0^t \|w(t-s,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \mathrm{d}s + 2C_{\mathcal{B}} \int_0^t \|f(t-s,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \mathrm{d}s.$$

From Grönwall's inequality it then follows that

$$\|w(t,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \le 2C_{\mathcal{B}} \int_0^t \|f(t-s,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^3\times V)} \mathrm{d}s \cdot \exp(2C_{\mathcal{B}}T)$$

with $t \le T$. This concludes the proof. □

Now we are ready to investigate the dependence of the derivative $\mathcal{G}'$ on ε. The desired statement would be, that for every $K \in \mathcal{B}$ and admissible variation K' it holds that

$$\mathcal{G}'(K)K' = \mathcal{O}(\varepsilon). \tag{5.22}$$

With this, we could then use Proposition 5.1 for the chemotaxis equation with a tumbling kernel that does not depend on the velocity v and show that the Kullback-Leibler divergence vanishes as $\varepsilon \to 0$,

$$d_{\mathrm{KL}}(\mu_0, \mu_{post}) = \mathcal{O}(\varepsilon).$$

First, we expand

$$w = w_0 + \varepsilon w_1 + \varepsilon^2 w_r, \qquad f = f_0 + \varepsilon f_1 + \varepsilon^2 f_2 + \varepsilon^3 f_r$$

and plug into (5.20). Note, that $\varepsilon^3 f_r = \mathcal{O}(\varepsilon^3)$. Comparing orders yields the following results. For $\mathcal{O}(1)$ we have that

$$-K\mathcal{L}w_0 = K'\mathcal{L}f_0 = 0,$$

meaning that $w_0(t,x,v) = w_0(t,x)$ since $f_0(t,x) = \rho_0(t,x)$. Further, at $\mathcal{O}(\varepsilon)$ we can see that

$$\begin{aligned} v\cdot\nabla_x w_0 - K\mathcal{L}w_1 &= K'\mathcal{L}f_1 \\ \Rightarrow K\mathcal{L}w_1 &= v\cdot\nabla_x w_0 - K'\mathcal{L}f_1 \\ &= v\cdot\nabla_x w_0 - \frac{K'}{K}\underbrace{\mathcal{L}v}_{=-v}\nabla_x\rho_0 \\ \Rightarrow w_1(t,x,v) &= -\frac{1}{K}v\cdot\nabla_x w_0(t,x) - \frac{K'}{K^2}v\cdot\nabla_x\rho_0(t,x) \end{aligned}$$

where we used that $\mathcal{L}(v) = \int_V v\mathrm{d}v - v = -v$, thus

$$\mathcal{L}(w_1) = \frac{1}{K}v\cdot\nabla_x w_0(t,x) + \frac{K'}{K^2}v\cdot\nabla_x\rho_0(t,x)$$

yields the correct result. For $\mathcal{O}(\varepsilon^2)$ we get

$$\begin{aligned} \partial_t w_0 + v\cdot\nabla_x w_1 - K\mathcal{L}w_r &= K'\mathcal{L}(f_2) \qquad (5.23)\\ \Leftrightarrow \partial_t w_0 - v\cdot\nabla_x\left(\frac{1}{K}v\cdot\nabla_x w_0\right) - v\cdot\nabla_x\left(\frac{K'}{K^2}v\cdot\nabla_x\rho_0\right) &= K\mathcal{L}w_r + K'\mathcal{L}(f_2). \end{aligned}$$

When averaging over v the right hand side becomes zero and we have

$$\partial_t w_0 - \int_V v\cdot\nabla_x v\frac{1}{K}\nabla_x w_0 \mathrm{d}v = \int_V v\cdot\nabla_x\frac{K'}{K^2}v\cdot\nabla_x\rho_0\mathrm{d}v \tag{5.24}$$

with zero initial condition $w_0(0,x) = 0$. This estimate shows the boundedness of $\partial_t w_0$ and $\nabla_x w_0$.

It is left to show of which order w_r is. For w_r it holds that

$$\varepsilon^2\partial_t w_r + \varepsilon v\cdot\nabla_x w_r - K\mathcal{L}w_r = -\partial_t w_0 - \frac{1}{\varepsilon}v\cdot\nabla_x w_0 - \varepsilon\partial_t w_1 - v\cdot\nabla_x w_1 + \frac{1}{\varepsilon}K\mathcal{L}w_1 + K'\mathcal{L}\big(\frac{1}{\varepsilon}f_1 + f_2 + \varepsilon f_r\big)$$

with initial condition $w_r(0,x,v) = -\frac{1}{\varepsilon}w_1(0,x,v)$. Note that w_0 and f_0 lie within the kernel of $\mathcal{K}_0$ which is why those terms are neglected. Using the boundedness of $\partial_t w_0$, $\nabla_x w_0$, $\mathcal{L}w_1$, $\mathcal{L}f_1$, $\mathcal{L}f_2$ and that $\mathcal{L}f_r \le \mathcal{O}(1)$, and applying the lemma of Grönwall to $|w_r|$ after integrating along the characteristic yields $\|w_r\|_{L^\infty(\mathbb{R}^3\times V)} \le \mathcal{O}\big(\frac{1}{\varepsilon}\big)$, since we have a $\mathcal{O}\big(\frac{1}{\varepsilon}\big)$ source term.

Overall, this leads to

$$\begin{aligned}\mathcal{G}'(K)K' &= \int_{\mathbb{R}^3}\int_V w(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x\\ &= \int_{\mathbb{R}^3}\int_V \big(w_0(t,x) + \varepsilon w_1(t,x,v) + \varepsilon^2 w_r(t,x,v)\big)\mathrm{d}v\chi(x)\mathrm{d}x\\ &= \int_{\mathbb{R}^3}\int_V w_0(t,x)\mathrm{d}v\chi(x)\mathrm{d}x + \mathcal{O}(\varepsilon),\end{aligned}$$

thus, we have to take a closer look at the Integral over $w_0(t,x)$.

Since we know the initial condition of $w_0(t,x)$ to be zero, we consider two approaches to estimate the integral above. First, one could try to show that

$$\partial_t\int_{\mathbb{R}^3}\int_V w_0(t,x)\mathrm{d}v\chi(x)\mathrm{d}x = 0,$$

from which it would then follow that

$$\int_{\mathbb{R}^3}\int_V w_0(t,x)\mathrm{d}v\chi(x)\mathrm{d}x = 0$$

because of the zero initial condition. From integrating (5.23) in velocity and space

$$\int_{\mathbb{R}^3}\int_V \partial_t w_0(t,x)\mathrm{d}v\chi(x)\mathrm{d}x = -\int_{\mathbb{R}^3}\int_V v\cdot\nabla_x w_1(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x - \int_{\mathbb{R}^3}\int_V K\mathcal{L}w_r\mathrm{d}v\chi(x)\mathrm{d}x + \int_{\mathbb{R}^3}\int_V K\mathcal{L}(f_2+f_r)\mathrm{d}v\chi(x)\mathrm{d}x$$

the desired result would follow, if it is true that

$$\int_{\mathbb{R}^3}\int_V v\cdot\nabla_x w_1(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x = 0,$$

since the other terms vanish when averaging over the velocity. We assume to have compact initial data for w_1 and that our measures are equal to one on their support and 0 elsewhere, thus from the divergence theorem we get that

$$\int_{\mathbb{R}^3}\nabla_x w_1(t,x,v)\chi(x)\mathrm{d}x = \int_{\mathrm{supp}(\chi(x))}\nabla_x w_1(t,x,v)\mathrm{d}v = \int_{\partial\,\mathrm{supp}(\chi(x))} w_1(t,x,v)n\mathrm{d}v, \tag{5.25}$$

with $\partial\,\mathrm{supp}(\chi(x))$ being the boundary of the compact support of the measurements and n its normal vector. This means, that the integral over $v\cdot\nabla_x w_1$ would vanish if, for instance the first order term of the derivative of the parameter-to-solution map equals zero on the boundary of our measurements. Or in other words, the first order term of the solution itself would have to stay constant on the boundary with respect to a change in the parameter K. Otherwise, one could try to choose another measurement than a compactly supported blob-function, for instance the whole $\mathbb{R}^3$ since the divergence theorem would then be applicable. But in the lab, this corresponds to taking a picture of all the bacteria on

the plate at different time steps, meaning we would not gain any information other than the total number of bacteria and have an ill-conditioned problem by construction of the experiment. Thus, this approach leads to an assumption on the behaviour of w_1 or on the measurements, which can not or shall not be imposed in that way within an experiment.

A second approach could be, to derive a linear differential equation for $\int_{\mathbb{R}^3}\int_V w_0(t,x)\mathrm{d}v\chi(x)\mathrm{d}x$, from which the desired result could be established due to the zero initial condition. But in (5.24) we have a differential equation for the term of interest

$$\partial_t w_0 - \int_V v\cdot\nabla_x v\frac{1}{K}\nabla_x w_0\mathrm{d}v = \int_V v\cdot\nabla_x\frac{K'}{K^2}v\cdot\nabla_x\rho_0\mathrm{d}v, \tag{5.26}$$

which is not linear since we still have a source term depending on ρ_0.

Overall, we could not draw information about $\int_{\mathbb{R}^3}\int_V w_0(t,x)\mathrm{d}v\chi(x)\mathrm{d}x$ from the two considered approaches. A more detailed discussion of this follows in subsection 5.3, where the differences to [28] are explained.

5.1.2 Local Behaviour around the MAP point

Now, the behaviour around the MAP-point for the velocity independent case is investigated in this section. The second derivative of the parameter to solution map is characterized by the proposition below.

Proposition 5.6 *For any admissible K_* and admissible variations K_1, K_2, the second derivative of the parameter-to-solution map $H = S''(K_*)[K_1,K_2]$ is the unique solution of*

$$\begin{aligned}\mathcal{T}H + C_{K_*}H &= -C_{K_1}w^2 - C_{K_2}w^1\\ H(0,x,v) &= 0\end{aligned} \tag{5.27}$$

where $f\in\mathcal{F}$ is the solution to the chemotaxis equation with parameter K_* and $w^{1,2}$ are the solutions to (5.20) with parameters $K_* + sK_{1,2}$. Moreover, it holds that $\|H\|_{L^\infty(\mathbb{R}^3\times V)}$ is bounded.

Proof: First, we note that

$$H_s = \frac{1}{s}\Big(f^{(1,2)} - f^{(1)} - f^{(2)} + f\Big) \tag{5.28}$$

where $f^{(i,j)}$ all satisfy the chemotaxis equation with different tumbling kernel and same initial condition,

$$\begin{aligned}\mathcal{T}f^{(1,2)} + C_{K_*+sK_1+sK_2}f^{(1,2)} &= 0,\\ \mathcal{T}f^{(1)} + C_{K_*+sK_1}f^{(1)} &= 0,\\ \mathcal{T}f^{(2)} + C_{K_*+sK_2}f^{(2)} &= 0,\\ \mathcal{T}f + C_{K_*}f &= 0.\end{aligned}$$

Combining these equations

$$\mathcal{T}\Big(f^{(1,2)} - f^{(1)} - f^{(2)} + f\Big) + C_{K_*}\Big(f^{(1,2)} - f^{(1)} - f^{(2)} + f\Big) = -sC_{K_1}f^{(1,2)} + sC_{K_1}f^{(1)} - sC_{K_2}f^{(1,2)} + sC_{K_2}f^{(2)}$$

and dividing by s^2 yields

$$\mathcal{T}H_s + C_{K_*}H_s = -C_{K_1}\frac{1}{s}\Big(f^{(1,2)} - f^{(1)}\Big) - C_{K_2}\frac{1}{s}\Big(f^{(1,2)} - f^{(2)}\Big).$$

For the expression $\frac{1}{s}\big(f^{(1,2)} - f^{(1)}\big)$ it holds that it is equal to w_s^2. This can be shown by replacing $K + sK_2$ in the proof of Proposition 5.5 with $(K_* + sK_1) + sK_2$. With the same argumentation $\frac{1}{s}\big(f^{(1,2)} - f^{(2)}\big)$ is equal to w_s^1. From carrying out the limit $s \to 0$ it follows that

$$\mathcal{T}H + C_{K_*}H = -C_{K_1}w^2 - C_{K_2}w^1.$$

The boundedness follows from integrating the equation along its characteristic and applying Grönwall's inequality to the absolute value of H. □

Now, we want to investigate the dependence of the second derivative $\mathcal{G}''$ on ε. The desired statement would be, that for $K_* \in \mathcal{A}$ and admissible variation $\hat{K}$, the diagonal elements $A''(K_*)[\hat{K}, \hat{K}]$ fulfil

$$A''(K_*)[\hat{K}, \hat{K}] = \mathcal{O}(\varepsilon).$$

Or in terms of $\mathcal{G}''$, that for every admissible $K_* \in \mathcal{A}$ and admissible variation $\hat{K}$ it holds that

$$\mathcal{G}''(K_*)[\hat{K}, \hat{K}] = \mathcal{O}(\varepsilon) \tag{5.29}$$

for every j, k.

Consider the expansions

$$\begin{aligned} H &= H_0 + \varepsilon H_1 + \varepsilon^2 H_r \\ w^1 &= w_0^1 + \varepsilon w_1^1 + \varepsilon^2 w_r^1 \\ w^2 &= w_0^2 + \varepsilon w_1^2 + \varepsilon^2 w_r^2, \end{aligned}$$

plug them into (5.27) and compare orders. For $\mathcal{O}(1)$ we get that

$$-K_*\mathcal{L}H_0 = K_1\mathcal{L}w_0^2 + K_2\mathcal{L}w_0^1 = 0$$

since $w_0^{1,2}(t, x, v) = w_0^{1,2}(t, x)$. Therefore, H_0 does not depend on the velocity either. At $\mathcal{O}(\varepsilon)$ it holds that

$$\begin{aligned} v \cdot \nabla_x H_0 - K_*\mathcal{L}H_1 &= K_1\mathcal{L}w_1^2 + K_2\mathcal{L}w_1^1 \\ \Leftrightarrow \mathcal{L}H_1 &= \frac{1}{K_*} v \cdot \nabla_x H_0 + \frac{K_1}{K_*} w_1^2 + \frac{K_2}{K_*} w_1^1 \\ \Rightarrow H_1 &= -\frac{1}{K_*} v \cdot \nabla_x H_0 - \frac{K_1}{K_*} w_1^2 - \frac{K_2}{K_*} w_1^1. \end{aligned}$$

Here, we used the assumption that K is bounded away from zero and that $\mathcal{L}(w_1^\circ) = -w_1^\circ$ for $\circ = 1, 2$. Looking at $\mathcal{O}(\varepsilon^2)$ yields

$$\partial_t H_0 + v \cdot \nabla_x H_1 - \underbrace{K_*\mathcal{L}H_r = K_1\mathcal{L}w_r^2 + K_2\mathcal{L}w_r^1}_{=0 \text{ when averaging over } v}. \tag{5.30}$$

Integrating over v and inserting the formula for H_1 results in

$$\partial_t \int_V H_0 \mathrm{d}v - \int_V v \cdot \nabla_x \Big(\frac{1}{K_*} v \cdot \nabla_x H_0 \Big) \mathrm{d}v = \int_V v \cdot \nabla_x \Big(\frac{K_1}{K_*} w_1^2 - \frac{K_2}{K_*} w_1^1 \Big) \mathrm{d}v \tag{5.31}$$

which shows the boundedness of $\partial_t H_0$ and $\nabla_x H_0$.

Now, it remains to show of what order H_r is. For H_r we have

$$\varepsilon^2\partial_t H_r + \varepsilon v\cdot\nabla_x H_r - K_*\mathcal{L}H_r = -\partial_t H_0 - \frac{1}{\varepsilon}v\cdot\nabla_x H_0 + \underbrace{\frac{1}{\varepsilon^2}K_*\mathcal{L}H_0}_{=0} - \varepsilon\partial_t H_1 - v\cdot\nabla_x H_1 + \frac{1}{\varepsilon}K_*\mathcal{L}H_1$$
$$+ \underbrace{\frac{1}{\varepsilon^2}K_1\mathcal{L}w_0^2}_{=0} + \frac{1}{\varepsilon}K_1\mathcal{L}w_1^2 + K_1\mathcal{L}w_r^2 + \underbrace{\frac{1}{\varepsilon^2}K_2\mathcal{L}w_0^1}_{=0} + \frac{1}{\varepsilon}K_2\mathcal{L}w_1^1 + K_2\mathcal{L}w_r^1$$

with initial condition $H_r(0,x,v) = -\frac{1}{\varepsilon}H_1(0,x,v)$. From the boundedness of $\partial_t H_0$ and $\nabla_x H_0$ and the lemma of Grönwall it follows that $\|H_r\|_{L^\infty(\mathbb{R}^3\times V)} \le \mathcal{O}\big(\frac{1}{\varepsilon}\big)$.

Overall, this leads to

$$\begin{aligned}\mathcal{G}''(K_*)[\hat{K},\hat{K}] &= \int_{\mathbb{R}^3}\int_V H(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x\\ &= \int_{\mathbb{R}^3}\int_V \big(H_0(x,v) + \varepsilon H_1(t,x,v) + \varepsilon^2 H_r(t,x,v)\big)\mathrm{d}v\chi(x)\mathrm{d}x\\ &= \int_{\mathbb{R}^3}\int_V H_0(x,v)\mathrm{d}v\chi(x)\mathrm{d}x + \mathcal{O}(\varepsilon)\end{aligned}$$

Thus, we need to take a closer look at the integral over $H_0(t,x)$. Again, we can follow the two different approaches of either using the divergence theorem or finding a linear differential equation for the integral over H_0. Note that $H_0(0,x) = 0$ and from integrating (5.30) in space and velocity

$$\begin{aligned}\int_{\mathbb{R}^3}\int_V \partial_t H_0\mathrm{d}v\chi(x)\mathrm{d}x + \int_{\mathbb{R}^3}\int_V v\cdot\nabla_x H_1\mathrm{d}v\chi(x)\mathrm{d}x &= \int_{\mathbb{R}^3}\int_V \big(K_*\mathcal{L}H_r + K_1\mathcal{L}w_r^2 + K_2\mathcal{L}w_r^1\big)\mathrm{d}v\chi(x)\mathrm{d}x\\ \Leftrightarrow \partial_t\int_{\mathbb{R}^3}\int_V H_0\mathrm{d}v\chi(x)\mathrm{d}x &= -\int_{\mathbb{R}^3}\int_V v\cdot\nabla_x H_1\mathrm{d}v\chi(x)\mathrm{d}x,\end{aligned}$$

where $H_0 = H_0(t,x)$ and $H_1 = H_1(t,x,v)$. But again, as it was the case for the global loss of information, this leads to an assumption on the behaviour of f at the boundary of the measurements which cannot be imposed in that way in an experiment. Also (5.31) gives a non-linear differential equation for H_0, which is why we cannot conclude that the integral vanishes from this. Again, the two approaches do not lead to an estimate for the integral of interest.

5.2 Velocity Dependent Case

Now, the situation for a velocity dependent tumbling kernel will be investigated. Even though, the approaches did not work out for the velocity independent case, we want to consider the more complex problem in order to see, whether or not difficulties arise earlier on in the calculations, due to the velocity dependency. The main difference to the previous case lies within the drift-diffusion limit which is discussed in subsection 3.3 in case of including a birth-death-term and in [10, Chap. 2] in case of no reaction term. In the following, no birth-death-term will be considered.

Remark 5.7 From [10, Chap. 2] we know that for $f_\varepsilon(t,x,v) = f_0(t,x,v) + \varepsilon f_1(t,x,v)$ and $\mathcal{K}_\varepsilon(t,x,v) = \mathcal{K}_0 + \varepsilon\mathcal{K}_1$ it holds that

$$f_0(t,x,v) = \rho_0(t,x)F(v), \tag{5.32}$$
$$f_1(t,x,v) = \kappa(t,x,v)\nabla_x\rho_0(t,x) - \Theta(t,x,v)\rho_0(t,x) + \rho_1(t,x)F(v) \tag{5.33}$$

with $F \in \ker(\mathcal{K}_0)$. It remains to show what f_2 looks like and of which order f_r is, if we look at the expansion $f_\varepsilon = f_1 + \varepsilon f_1 + \varepsilon^2 f_2 + \varepsilon^3 f_r$. Consider the scaled chemotaxis equation

$$\varepsilon^2\partial_t f_\varepsilon + \varepsilon v\cdot\nabla_x f_\varepsilon = (\mathcal{K}_0 + \varepsilon\mathcal{K}_1)(f_\varepsilon)$$

and the equations (5.32) and (5.33). Then looking at $\mathcal{O}(\varepsilon^2)$ gives rise to the relation

$$\mathcal{K}_0(f_2) = \partial_t f_0 + v \cdot \nabla_x f_1 - \mathcal{K}_1 f_1 \tag{5.34}$$

$$\Leftrightarrow f_2 = \mathcal{K}_0^{-1}(\partial_t f_0 + v \cdot \nabla_x f_1 - \mathcal{K}_1 f_1). \tag{5.35}$$

In order for f_2 to be well-defined, $\partial_t f_0 + v \cdot \nabla_x f_1 - \mathcal{K}_1 f_1$ needs to lie within the range of the operator $\mathcal{K}_0$. From Lemma 3.5 it follows, that the equation

$$\mathcal{K}_0(g) = -\mathcal{K}_1(f_1)$$

has a solution $g \in L^2(V; \mathrm{d}v/F)$ since $\mathcal{K}_1(f_1) \in L^2(V; \mathrm{d}v/F)$ and $\int_V \mathcal{K}_1(f_1)\mathrm{d}v = 0$. Also, it follows from Lemma 3.5 that the equation

$$\mathcal{K}_0(h) = \partial_t f_0 + v \cdot \nabla_x f_1$$

has a solution $h \in L^2(V; \mathrm{d}v/F)$ if and only if $\int_V \partial_t f_0 + v \cdot \nabla_x f_1 \mathrm{d}v = 0$ and $\partial_t f_0 + v \cdot \nabla_x f_1 \in L^2(V; \mathrm{d}v/F)$. In order to fulfil the first condition, it must hold true that

$$\int_V v \cdot \nabla_x f_1 \mathrm{d}v = \int_V \partial_t f_0 \mathrm{d}v.$$

Looking at the left hand side yields

$$\begin{aligned}
\int_V v \cdot \nabla_x f_1 \mathrm{d}v &= \int_V (v \cdot \nabla_x(\kappa(t,x,v)\nabla_x \rho_0(t,x) - \Theta(t,x,v)\rho_0(t,x) + \rho_1(t,x)F(v)))\mathrm{d}v \\
&= \int_V v \cdot \nabla_x(\kappa(t,x,v)\nabla_x \rho_0(t,x))\mathrm{d}v - \int_V v \cdot \nabla_x(\Theta(t,x,v)\rho_0(t,x))\mathrm{d}v \\
&\quad + \underbrace{\int_V v \cdot \nabla_x \rho_1(t,x)F(v)\mathrm{d}v}_{=0}
\end{aligned}$$

where we used that $\int_V v \cdot \nabla_x \rho_1(t,x)F(v)\mathrm{d}v = \nabla_x \rho_1(t,x) \int_V vF(v)\mathrm{d}v = 0$ due to the properties we imposed on F. Hence, in order to f_2 being well-defined ρ_0 has to fulfil

$$\int_V v \cdot \nabla_x(\kappa(t,x,v)\nabla_x \rho_0(t,x))\mathrm{d}v - \int_V v \cdot \nabla_x(\Theta(t,x,v)\rho_0(t,x))\mathrm{d}v = \partial_t \rho_0(t,x) \int_V F\mathrm{d}v = \partial_t \rho_0(t,x).$$

This condition equipped with suitable initial condition determines ρ_0. Now it remains to show of what order f_r is. Rearranging the chemotaxis equation and dividing by ε^3 yields

$$\begin{aligned}
\varepsilon^2 \partial_t f_r + v \cdot \nabla_x f_r - (\mathcal{K}_0 + \varepsilon\mathcal{K}_1)(f_r) &= -\left(\frac{1}{\varepsilon}\partial_t f_0 + \frac{1}{\varepsilon^2} v \cdot \nabla_x f_0 - \frac{1}{\varepsilon^3}(\varepsilon\mathcal{K}_1)(f_0)\right) \\
&\quad - \left(\partial_t f_1 + \frac{1}{\varepsilon} v \cdot \nabla_x f_1 - \frac{1}{\varepsilon^2}(\mathcal{K}_0 + \varepsilon\mathcal{K}_1)(f_1)\right) \\
&\quad - \left(\varepsilon\partial_t f_2 + v \cdot \nabla_x f_2 - \frac{1}{\varepsilon}(\mathcal{K}_0 + \varepsilon\mathcal{K}_1)(f_2)\right) \\
&= -\partial_t f_1 - \varepsilon\partial_t f_2 - v \cdot \nabla_x f_2 + \mathcal{K}_1(f_2)
\end{aligned}$$

where we used that $\mathcal{K}_0(f_0) = 0$, $\mathcal{K}_0(f_1) = v \cdot \nabla_x f_0 - \mathcal{K}_1(f_0)$ and (5.34). With initial condition $f_r(0,x,v) = -\varepsilon^2 f_1(0,x,v) - \varepsilon f_2(0,x,v)$ and the source term being of order $\mathcal{O}(1)$ we get from applying the lemma of Grönwall that $\|f_r\|_{L^\infty(\mathbb{R}^3 \times V)} \leq \mathcal{O}(1)$. Further, we assume that the kernels of all K_0 with $(K_0, K_1) \in \mathcal{A}$ are spanned by the same velocity distribution F.

5.2.1 Global Loss of Information

Now, we are ready to look at the global loss of information. The proposition below follows analogously to Proposition 5.5.

Proposition 5.8 *Denote*

$$\mathcal{T}f = \varepsilon^2 \partial_t f + \varepsilon v \cdot \nabla_x f, \qquad C_K f = -\int_V K(t,x,v,v')f(t,x,v') - K(t,x,v',v)f(t,x,v)\mathrm{d}v. \tag{5.36}$$

For $K \in \mathcal{A}$ and an admissible variation $\tilde{K}$, the derivative of the parameter-to-solution map w is the unique solution of

$$\begin{aligned} \mathcal{T}w + C_K w &= -C_{\tilde{K}} f, \\ w(0,x,v) &= 0, \end{aligned} \tag{5.37}$$

with $f \in \mathcal{F}$ being the solution of the chemotaxis equation with parameter K. Also, for $\varepsilon = 1$ it holds that

$$\|w\|_{L^\infty(0,T;L^\infty(\mathbb{R}^3\times V))} \le C(C_{\mathcal{A}}, T)\|f\|_{L^\infty(0,T;L^\infty(\mathbb{R}^3\times V))} \tag{5.38}$$

where $C_{\mathcal{A}}$ denotes the constant C appearing in the admissible set $\mathcal{A}$ as defined in (3.10)

Then, the desired result would be that for every $K \in \mathcal{A}$ and admissible variation K' it holds that

$$\mathcal{G}'(K)K' \le \mathcal{O}(\varepsilon). \tag{5.39}$$

Expand

$$w = w_0 + \varepsilon w_1 + \varepsilon^2 w_r, \quad f = f_0 + \varepsilon f1 + \varepsilon^2 f_2 + \varepsilon^3 f_r, \quad K = K_0 + \varepsilon K_1$$

and plug into (5.37). Comparing orders yields the following results. For $\mathcal{O}(1)$ we have that

$$-\mathcal{K}_0(w_0) = \mathcal{K}'_0(f_0) = 0$$

meaning, that $w_0(t,x,v) = \rho_{w_0}(t,x)F(v)$. Further, at $\mathcal{O}(\varepsilon)$ one can see that

$$\begin{aligned} v\cdot\nabla_x w_0 - \mathcal{K}_1(w_0) - \mathcal{K}_0(w_1) &= \mathcal{K}'_0(f_1) + \mathcal{K}'_1(f_0) \\ \Leftrightarrow \mathcal{K}_0(w_1) &= vF\nabla_x\rho_{w_0} - \rho_{w_0}\mathcal{K}_1(F) - \nabla_x\rho_0\mathcal{K}'_0(\kappa) - \rho_0\mathcal{K}'_0(\Theta) - \rho_0\mathcal{K}'_1(F), \end{aligned}$$

where we used that

$$\mathcal{K}'_0(f_1) = \nabla_x\rho_0\mathcal{K}'_0(\kappa) + \rho_0\mathcal{K}'_0(\Theta) + \rho_1\mathcal{K}'_0(F) = \nabla_x\rho_0\mathcal{K}'_0(\kappa) + \rho_0\mathcal{K}'_0(\Theta).$$

From this, it follows that

$$w_1 = \kappa_{w_1}\nabla_x\rho_{w_0} - \Theta_{w_1}\rho_{w_0} + \kappa'_{w_1}\nabla_x\rho_0 - \Theta'_{w_1}\rho_0 + \rho_{w_1}F \tag{5.40}$$

with

$$\mathcal{K}_0(\kappa_{w_1}) = vF, \tag{5.41}$$
$$\mathcal{K}_0(\Theta_{w_1}) = \mathcal{K}_1(F), \tag{5.42}$$
$$\mathcal{K}_0(\kappa'_{w_1}) = \mathcal{K}'_0(\kappa), \tag{5.43}$$
$$\mathcal{K}_0(\Theta'_{w_1}) = \mathcal{K}'_0(\Theta) + \mathcal{K}'_1(F). \tag{5.44}$$

The equations (5.41) - (5.44) all have a bounded solution according to Lemma 3.5, since the right hand sides RHS each fulfil

$$\mathrm{RHS} \in L^2(V; \mathrm{d}v/F), \qquad \int_V \mathrm{RHS}\ \mathrm{d}v = 0.$$

That the integrals over the right hand sides is equal to zero holds true because of the property that the tumbling operator vanishes when averaging over v and that we assume $\int_V vF\mathrm{d}v = 0$. Additionally, from Lemma 3.5 it holds that $\int_V w_1(t,x,v)\mathrm{d}v = \rho_{w_1}$. Looking at $\mathcal{O}(\varepsilon^2)$ yields

$$\partial_t w_0 + v\cdot\nabla_x w_1 - \mathcal{K}_0(w_r) - \mathcal{K}_1(w_1) = \mathcal{K}_1'(f_1). \tag{5.45}$$

Averaging over v then gives

$$\int_V \partial_t w_0 \mathrm{d}v + \int_V v\cdot\nabla_x w_1 \mathrm{d}v = 0$$

and plugging in (5.40) yields

$$\begin{aligned}\partial_t\rho_{w_0} + \int_V v\cdot\nabla_x(\kappa_{w_1}\nabla_x\rho_{w_0})\mathrm{d}v - \int_V v\cdot\nabla_x(\Theta_{w_1}\rho_{w_0})\mathrm{d}v = &-\int_V v\cdot\nabla_x(\kappa_{w_1}'\nabla_x\rho_0)\mathrm{d}v\\ &-\int_V v\cdot\nabla_x(\Theta_{w_1}'\rho_0)\mathrm{d}v\\ &-\underbrace{\int_V vF\mathrm{d}v\cdot\nabla_x\rho_{w_1}}_{=0}.\end{aligned} \tag{5.46}$$

Since $\rho_0, \nabla_x\rho_0$ are bounded, $\partial_t\rho_{w_0}, \nabla_x\rho_{w_0}$ and thus $\partial_t w_0, \nabla_x w_0$ are bounded as well.

For w_r we get

$$\begin{aligned}\varepsilon^2\partial_t w_r + \varepsilon v\cdot\nabla_x w_r - \mathcal{K}_\varepsilon(w_r) = &-\Big(\partial_t w_0 + \frac{1}{\varepsilon}v\cdot\nabla_x w_0 - \overbrace{\frac{1}{\varepsilon^2}\mathcal{K}_0(w_0)}^{=0} - \frac{1}{\varepsilon}\mathcal{K}_1(w_0)\Big)\\ &-\Big(\varepsilon\partial_t w_1 + v\cdot\nabla_x w_1 - \frac{1}{\varepsilon}\mathcal{K}_0(w_1) - \mathcal{K}_1(w_1)\Big)\\ &+\underbrace{\frac{1}{\varepsilon^2}\mathcal{K}_0'(f_0)}_{=0} + \frac{1}{\varepsilon}\mathcal{K}_0'(f_1) + \mathcal{K}_0'(f_r) + \frac{1}{\varepsilon}\mathcal{K}_1'(f_0) + \mathcal{K}_1'(f_1) + \varepsilon\mathcal{K}_1'(f_r)\end{aligned}$$

with initial condition $w_r(0,x,v) = -\frac{1}{\varepsilon}w_1(0,x,v)$. From the previous results and the lemma of Grönwall it follows that $\|w_r\|_{L^\infty(\mathbb{R}^3\times V)} \le \mathcal{O}\big(\frac{1}{\varepsilon}\big)$. Overall, this leads to

$$\begin{aligned}\mathcal{G}'(K)K' &= \int_{\mathbb{R}^3}\int_V w(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x\\ &= \int_{\mathbb{R}^3}\int_V \big(w_0(t,x,v) + \varepsilon w_1(t,x,v) + \varepsilon^2 w_r(t,x,v)\big)\mathrm{d}v\chi(x)\mathrm{d}x\\ &= \int_{\mathbb{R}^3}\int_V w_0(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x + \mathcal{O}(\varepsilon),\end{aligned}$$

Thus, we need to take a closer look at the integral over $w_0(t,x,v)$. Given that $w_0(0,x,v) = 0$ and integrating (5.45) in velocity and space

$$\begin{aligned}\int_{\mathbb{R}^3}\int_V \partial_t w_0(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x = &-\int_{\mathbb{R}^3}\int_V v\cdot\nabla_x w_1(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x + \int_{\mathbb{R}^3}\int_V \mathcal{K}_0(w_r)\mathrm{d}v\chi(x)\mathrm{d}x\\ &+\int_{\mathbb{R}^3}\int_V \mathcal{K}_1(w_1)\mathrm{d}v\chi(x)\mathrm{d}x - \int_{\mathbb{R}^3}\int_V \mathcal{K}_1'(f_1)\mathrm{d}v\chi(x)\mathrm{d}x\end{aligned}$$

$$\Leftrightarrow \partial_t\int_{\mathbb{R}^3}\int_V w_0(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x = 0.$$

yields

$$\int_{\mathbb{R}^3}\int_V w_0(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x = \int_{\mathbb{R}^3}\int_V v\cdot\nabla_x w_1(t,x,v)\mathrm{d}v\chi(x)\mathrm{d}x.$$

We note that we assume to have compact initial data for w_1 and that our measures are equal to one on their support and 0 elsewhere, thus from the divergence theorem we get that

$$\int_{\mathbb{R}^3} \nabla_x w_1(t,x,v)\chi(x)\mathrm{d}x = \int_{\mathrm{supp}(\chi(x))} \nabla_x w_1(t,x,v)\mathrm{d}v = \int_{\partial\,\mathrm{supp}(\chi(x))} w_1(t,x,v)n\mathrm{d}v, \tag{5.47}$$

with $\partial\,\mathrm{supp}(\chi(x))$ being the boundary of the compact support of the measurements and n its normal vector. Hence, we end up with the same preconditions as it was the case for the velocity-independent case. Additionally, (5.46) gives rise to a nonlinear differential equation for the quantity of interest. Again, neither the use of the divergence theorem or of finding a linear differential equation for the integral of interest work out.

5.2.2 Local Behaviour around the MAP point

Now, the loss of information on a local level will be discussed. The proof of the following proposition is analogously to the proof of Proposition 5.6.

Proposition 5.9 *Denote*

$$\mathcal{T}f = \varepsilon^2\partial_t f + \varepsilon v\cdot\nabla_x f, \qquad C_K f = -\int_V K(t,x,v,v')f(t,x,v') - K(t,x,v',v)f(t,x,v)\mathrm{d}v \tag{5.48}$$

For any admissible K_* and admissible variations K_1, K_2, the second derivative of the parameter-to-solution map $H = S''(K_*)[\hat{K},\hat{K}]$ is the unique solution of

$$\begin{aligned} \mathcal{T}H + C_{K_*}H &= -C_{K_1}w^2 - C_{K_2}w^1 \\ H(0,x,v) &= 0 \end{aligned} \tag{5.49}$$

where $f\in\mathcal{F}$ is the solution to the chemotaxis equation with parameter K_* and $w^{1,2}$ are the solutions to (5.20) with parameters $K_* + sK_{1,2}$. Moreover, it holds that $\|H\|_{L^\infty(\mathbb{R}^3\times V)}$ is bounded.

The desired result would be that for $K_*\in\mathcal{A}$ and admissible variation $\hat{K}$, the diagonal elements $A''(K_*)[\hat{K},\hat{K}]$ fulfill

$$A''(K_*)[\hat{K},\hat{K}] = \mathcal{O}(\varepsilon).$$

Or in terms of $\mathcal{G}''$, that for every admissible $K^*\in\mathcal{A}$ and admissible variation $\hat{K}$ it holds

$$\mathcal{G}''(K_*)[\hat{K},\hat{K}] = \mathcal{O}(\varepsilon) \tag{5.50}$$

for every j,k.

Consider the expansions

$$\begin{aligned} H &= H_0 + \varepsilon H_1 + \varepsilon^2 H_r \\ w^1 &= w_0^1 + \varepsilon w_1^1 + \varepsilon^2 w_r^1 \\ w^2 &= w_0^2 + \varepsilon w_1^2 + \varepsilon^2 w_r^2 \\ K^\circ &= K_0^\circ + \varepsilon K_1^\circ, \end{aligned}$$

with $\circ = *, 1, 2$ and plug them into (5.49) and compare orders. For $\mathcal{O}(1)$ we get that

$$-\mathcal{K}_0^*(H_0) = \mathcal{K}_0^2(w_0^1) + \mathcal{K}_0^1(w_0^2) = 0$$

leading to $H_0(t,x,v) = \rho_{H_0}(t,x)F(v)$. Further, at $\mathcal{O}(\varepsilon)$ one can see that

$$v\cdot\nabla_x H_0 - \mathcal{K}_1^*(H_0) - \mathcal{K}_0^*(H_1) = \mathcal{K}_0^2(w_1^1) + \mathcal{K}_1^2(w_0^1) + \mathcal{K}_0^1(w_1^2) + \mathcal{K}_1^1(w_0^2)$$

$$\Leftrightarrow \mathcal{K}_0^*(H_1) = vF\nabla_x\rho_{H_0} - \rho_{H_0}\mathcal{K}_1^*(F) - \mathcal{K}_0^2(w_1^1) - \mathcal{K}_1^2(w_0^1) - \mathcal{K}_0^1(w_1^2) - \mathcal{K}_1^1(w_0^2). \tag{5.51}$$

We take a closer look at $\mathcal{K}_0^2(w_1^1)$, $\mathcal{K}_1^2(w_0^1)$, $\mathcal{K}_0^1(w_1^2)$ and $\mathcal{K}_1^1(w_0^2)$ using (5.40)

$$w_1^\circ = \kappa_{w_1^\circ}\nabla_x\rho_{w_0^\circ} - \Theta_{w_1^\circ}\rho_{w_0^\circ} + \kappa'_{w_1^\circ}\nabla_x\rho_{f_0} - \Theta'_{w_1^\circ}\rho_{f_0} + \rho_{w_1^\circ}F$$

for $\circ = 1, 2$. First we note that for all combinations of $\circ = 1, 2$ and $\triangle = 1, 2$

$$\mathcal{K}_0^\circ(\rho_{w_1^\triangle}F) = 0.$$

With

$$\mathcal{K}_0^{2,1}(w_1^{1,2}) = \nabla_x\rho_{w_0^{1,2}}\mathcal{K}_0^{2,1}(\kappa_{w_1^{1,2}}) - \rho_{w_0^{1,2}}\mathcal{K}_0^{2,1}(\Theta_{w_1^{1,2}}) - \nabla_x\rho_{f_0}\mathcal{K}_0^{2,1}(\kappa'_{w_1^{1,2}}) - \rho_{f_0}\mathcal{K}_0^{2,1}(\Theta'_{w_1^{1,2}})$$

and

$$\mathcal{K}_1^{2,1}(w_0^{1,2}) = \rho_{w_0^{1,2}}\mathcal{K}_1^{2,1}(F).$$

(5.51) can be rewritten as

$$\begin{aligned}\mathcal{K}_0^*(H_1) = {} & vF\nabla_x\rho_{H_0} - \rho_{H_0}\mathcal{K}_1^*(F) - \nabla_x\rho_{w_0^1}\mathcal{K}_0^2(\kappa_{w_1^1}) - \rho_{w_1^1}\mathcal{K}_0^2(\Theta_{w_1^1}) \\ & - \nabla_x\rho_{w_0^2}\mathcal{K}_0^1(\kappa_{w_1^2}) - \rho_{w_1^2}\mathcal{K}_0^1(\Theta_{w_1^2}) - \nabla_x\rho_{f_0}\Big(\mathcal{K}_0^2(\kappa'_{w_1^1}) + \mathcal{K}_0^1(\kappa'_{w_1^2})\Big) \\ & - \rho_{f_0}\Big(\mathcal{K}_0^2(\Theta'_{w_1^1}) + \mathcal{K}_0^1(\Theta'_{w_1^2})\Big).\end{aligned}$$

This leads to

$$H_1 = \kappa_{H_1}\nabla_x\rho_{H_0} - \rho_{H_0}\Theta_{H_1} - \kappa^1_{H_1}\nabla_x\rho_{w_0^1} - \Theta^1_{H_1}\rho_{w_0^1} \tag{5.52}$$

$$- \kappa^2_{H_1}\nabla_x\rho_{w_0^2} - \Theta^2_{H_1}\rho_{w_0^2} - \kappa^3_{H_1}\nabla_x\rho_{f_0} - \Theta^3_{H_1}\rho_{f_0} + \rho_{H_1}F \tag{5.53}$$

with

$$\mathcal{K}_0^*(\kappa_{H_1}) = vF \tag{5.54}$$

$$\mathcal{K}_0^*(\Theta_{H_1}) = \mathcal{K}_1^*(F) \tag{5.55}$$

$$\mathcal{K}_0^*(\kappa^1_{H_1}) = \mathcal{K}_0^2(\kappa_{w_1^1}) \tag{5.56}$$

$$\mathcal{K}_0^*(\Theta^1_{H_1}) = \mathcal{K}_0^2(\Theta_{w_1^1}) \tag{5.57}$$

$$\mathcal{K}_0^*(\kappa^2_{H_1}) = \mathcal{K}_0^1(\kappa_{w_1^2}) \tag{5.58}$$

$$\mathcal{K}_0^*(\Theta^2_{H_1}) = \mathcal{K}_0^1(\Theta_{w_1^2}) \tag{5.59}$$

$$\mathcal{K}_0^*(\kappa^3_{H_1}) = \mathcal{K}_0^2(\kappa'_{w_1^1}) + \mathcal{K}_0^1(\kappa'_{w_1^2}) \tag{5.60}$$

$$\mathcal{K}_0^*(\Theta^3_{H_1}) = \mathcal{K}_0^2(\Theta'_{w_1^1}) + \mathcal{K}_0^1(\Theta'_{w_1^2}). \tag{5.61}$$

The equations (5.54) - (5.61) all have a bounded solution according to Lemma 3.5, since the right hand sides RHS each fulfil

$$\text{RHS} \in L^2(V; \mathrm{d}v/F), \qquad \int_V \text{RHS}\, \mathrm{d}v = 0.$$

That the integrals of the right hand sides are equal to zero, follows from the property of the tumbling operator to vanish when averaging over V and that we assume $\int_V vF\mathrm{d}v = 0$. Additionally, from Lemma 3.5 it holds that $\int_V H_1(t, x, v)\mathrm{d}v = \rho_{H_1}$.

Looking at $\mathcal{O}(\varepsilon^2)$ yields

$$\partial_t H_0 + v \cdot \nabla_x H_1 - \mathcal{K}_0^*(H_r) - \mathcal{K}_1^*(H_1) = \mathcal{K}_1^2(w_1^1) + \mathcal{K}_1^1(w_1^2) + \mathcal{K}_0^2(w_r^1) + \mathcal{K}_0^1(w_r^2). \tag{5.62}$$

Averaging over v leads to

$$\int_V \partial_t H_0 \mathrm{d}v + \int_V v \cdot \nabla_x H_1 v \mathrm{d}v = 0 \tag{5.63}$$

and plugging in (5.53) gives the boundedness of $\partial_t \rho_{H_0}, \nabla_x \rho_{H_0}$ and thus $\partial_t H_0$ and $\nabla_x H_0$ are bounded as well.

It is left to show, of what order $H_r(t,x,v)$ is. Rearranging (5.49) after plugging in the expansion and dividing by ε^2 gives

$$\begin{aligned}
\varepsilon^2 \partial_t H_r + \varepsilon v \cdot \nabla_x H_r - \mathcal{K}_\varepsilon^*(H_r) = &-\partial_t H_0 - \frac{1}{\varepsilon} v \cdot \nabla_x H_0 + \underbrace{\frac{1}{\varepsilon^2} K_0^* \mathcal{L} H_0}_{=0} + \frac{1}{\varepsilon} \mathcal{K}_1^*(H_0) \\
&- \varepsilon \partial_t H_1 - v \cdot \nabla_x H_1 + \frac{1}{\varepsilon} \mathcal{K}_0^*(H_1) + \mathcal{K}_1^*(H_1) \\
&+ \underbrace{\frac{1}{\varepsilon^2} \mathcal{K}_0^1(w_0^2)}_{=0} + \frac{1}{\varepsilon} \mathcal{K}_1^1(w_0^2) + \frac{1}{\varepsilon} K_0^1(w_1^2) + \mathcal{K}_1^1(w_1^2) + \mathcal{K}_0^1(w_r^2) + \varepsilon \mathcal{K}_1^1(w_r^2) \\
&+ \underbrace{\frac{1}{\varepsilon^2} \mathcal{K}_0^2(w_0^1)}_{=0} + \frac{1}{\varepsilon} \mathcal{K}_1^2(w_0^1) + \frac{1}{\varepsilon} K_0^2(w_1^1) + \mathcal{K}_1^2(w_1^1) + \mathcal{K}_0^2(w_r^1) + \varepsilon \mathcal{K}_1^2(w_r^1)
\end{aligned}$$

with initial condition $H_r(0,x,v) = -\frac{1}{\varepsilon} H_1(0,x,v)$. From the previous results and the lemma of Grönwall it follows that $\|H_r\|_{L^\infty(\mathbb{R}^3 \times V)} \leq \mathcal{O}\big(\frac{1}{\varepsilon}\big)$.

Overall, this leads to

$$\begin{aligned}
\mathcal{G}''(K_*)[\hat{K}, \hat{K}] &= \int_{\mathbb{R}^3} \int_V H(t,x,v) \mathrm{d}v \chi(x) \mathrm{d}x \\
&= \int_{\mathbb{R}^3} \int_V \big(H_0(x,v) + \varepsilon H_1(t,x,v) + \varepsilon^2 H_r(t,x,v) \big) \mathrm{d}v \chi(x) \mathrm{d}x \\
&= \int_{\mathbb{R}^3} \int_V H_0(t,x,v) \mathrm{d}v \chi(x) \mathrm{d}x + \mathcal{O}(\varepsilon)
\end{aligned}$$

Thus, we take a look at the integral over $H_0(t,x,v)$. Given that $H_0(0,x,v) = 0$ and integrating (5.62) in time and space

$$\begin{aligned}
\int_{\mathbb{R}^3} \int_V \partial_t H_0 \mathrm{d}v \chi \mathrm{d}x + \int_{\mathbb{R}^3} \int_V v \cdot \nabla_x H_1 \mathrm{d}v \chi \mathrm{d}x &= \int_{\mathbb{R}^3} \int_V \mathcal{K}_1^2(w_1^1) + \mathcal{K}_1^1(w_1^2) + \mathcal{K}_0^2(w_r^1) + \mathcal{K}_0^1(w_r^2) \mathrm{d}v \chi \mathrm{d}x \\
\Leftrightarrow \partial_t \int_{\mathbb{R}^3} \int_V H_0 \mathrm{d}v \chi \mathrm{d}x &= - \int_{\mathbb{R}^3} \int_V v \cdot \nabla_x H_1 \mathrm{d}v \chi \mathrm{d}x
\end{aligned}$$

with $H_0 = H_0(t,x,v)$, $H_1 = H_1(t,x,v)$ and $\chi = \chi(x)$, as it was the case for the velocity independent problem. Also, (5.63) again gives rise to a non-linear differential equation for the integral of interest. Therefore, we have the same situation as in the case of a tumbling kernel that does not depend on the velocity and cannot quantify the loss of information.

5.3 Conclusion and Comparison with the Ill-Conditioning of Inverse Radiative Transfer

Within this chapter, the stability deterioration in the diffusive limit of the chemotaxis equation was discussed. It has been attempted to extend the results from [28] about the instability deterioration in the Bayesian framework of stationary radiative transfer to the non-stationary case with a parabolic

scaling. In scope of this work, this could not be shown, but the previous sections can be seen as a first attempt of quantifying the loss of information in the diffusive limit and thus a beginning of further research.

We still want to discuss, why the approaches for global and local loss of information where chosen and present some guesses on why it did not work out for the inverse chemotaxis equation. As mentioned before, the chapter mainly followed the elaborations in [28], where the methods were applied to the inverse problem of stationary radiative transfer (RTE). First, we want to shortly summarize the results that can be found in the literature about the multiscale analysis of RTE in order to compare it to our situation.

The model, that describes the propagation of photon particles, namely RTE, is of the form

$$v \cdot \nabla_x f(x,v) = \int_V k(x,v,v') f(x,v') \mathrm{d}v' - \alpha(x,v) f(x,v) \tag{5.64}$$

as introduced for instance in [5,31] with $x \in \Omega \subset \mathbb{R}^d$ and $v \in V = \mathcal{S}^{d-1}$. Here, one sees the similarity to the kinetic equation for chemotaxis, where α corresponds to $\int_V K(t,x,v',v)\mathrm{d}v'$. The forward problem would be to determine the particle distribution f by the optical properties of a tissue described via k, α. We call $k(x,v,v')$ the *scattering coefficient* and $\alpha(x,v)$ the *total absorption coefficient*. The inverse problem for RTE is widely discussed in literature, for example is the unique reconstruction of the medium in dimension 3 shown in [11], in case of a fully known Albedo-Operator. Further reviews can be found in [6,7].

On the macroscopic level, the dynamic of the photons can be described by the diffusion equation (DE)

$$-\nabla_x (a(x) \nabla_x \rho(x)) = 0,$$

with $a(x)$ being the diffusion coefficient [31]. Again, the equation resembles the macroscopic equation in the case of chemotaxis, namely the Keller-Segel model. In [31], the diffusion limit for the case were

$$\alpha(x,v) = k(x,v,v') = \varepsilon^{-1}\sigma(x),$$

where ε is the *Knudsen number* and corresponds to our scaling scaling parameter. For RTE, the Knudsen number stands for the ratio of the mean free path, which is the average distance a particle travels without scattering, and the domain [31]. Thus, we are in the case of velocity-independent coefficients. Further, the boundary is defined as

$$\Gamma_\pm = \{(x,v) \mid x \in \partial\Omega, \pm v \cdot n_x > 0\}$$

or for a fixed point $y \in \partial\Omega$

$$\Gamma_\pm(y) = \{(x,v) \mid x = y, \pm v \cdot n_y > 0\}$$

where n_x is the normal vector pointing out of the domain at point x. The boundary conditions are imposed on Γ_-, the measurements are taken on Γ_+. The following theorem shows the convergence of the two forward problems and the proof can be found in [31, Theorem 2.2].

Theorem 5.10 *Suppose that $f(x,v)$ satisfies*

$$\begin{aligned} v \cdot \nabla_x f(x,v) &= \int_V k(x,v,v') f(x,v') \mathrm{d}v - \alpha(x,v) f(x,v) \\ f|_{\Gamma_-} &= \phi(x,v) \end{aligned} \tag{5.65}$$

with smooth boundary condition and that $\rho(x)$ solves

$$\begin{aligned} -\nabla_x (a(x) \nabla_x \rho(x)) &= 0 \\ \rho|_{\partial\Omega} &= \xi(x). \end{aligned} \tag{5.66}$$

If the boundary data ϕ and ξ are compatible with each other, then $f(x,v) \to \rho(x)$ in the diffusive limit $\varepsilon \to 0$.

With the established convergence of the forward problems, the convergence of the inverse problems in the Bayesian setting could also be shown in [31]. Here, they consider the Albedo operator

$$(\mathcal{H}^{\mathrm{RTE}}(\sigma)\phi)(x) = -\frac{1}{C\varepsilon}\int_{\Gamma_+(x)\cup\Gamma_-(x)} v\cdot n f(x,v)\mathrm{d}v$$

and the map that takes the Dirichlet data to the Neumann outflow, the DtN-map

$$(\mathcal{H}^{DE}(\sigma)\phi)(x) = \frac{1}{\sigma}\partial_n\rho(x)$$

for a constant C depending on the dimension d, the solution f of (5.65) and ρ satisfying (5.66). The forward maps are defined via

$$\mathcal{G}^{RTE}(\sigma)_{j,k} = l_j(\mathcal{H}^{\mathrm{RTE}}(\sigma)\phi_k)$$
$$\mathcal{G}^{DE}(\sigma)_{j,k} = l_j(\mathcal{H}^{\mathrm{DE}}(\sigma)\phi_k)$$

with l_j being a linear functional of the Albedo-operator and the DtN-map respectively. Under the assumption, that the prior distribution of σ is supported on the admissible set

$$\mathcal{A} = \left\{\sigma \in C^3(\Omega) \mid \max\{\|\sigma\|_{L^\infty(\Omega)}, \|\sigma^{-1}\|_{L^\infty(\Omega)}, \|\nabla(\sigma^{-1})\|_{L^\infty(\Omega)}\} < C\right\}$$

the convergence of the forward maps is shown in [31, Proposition 3.1]

Proposition 5.11 *Assume that $\phi_k(x,v) = \xi(x) - \varepsilon\frac{1}{\sigma(x)}v(x)\nabla_x\xi_k(x)$. The the two forward maps $\mathcal{G}^{RTE}$ and $\mathcal{G}^{DE}$ fulfil*

$$\sup_{\sigma\in\mathcal{A}}\|\mathcal{G}^{RTE}(\sigma) - \mathcal{G}^{DE}(\sigma)\|_\infty \le \frac{C_{\mathcal{A}}}{C_d}\varepsilon$$

where $C_{\mathcal{A}}$ depends on C_1. Further, there exists a constant $C = C(C_1,\Omega)$ such that

$$\max\left\{\sup_{\sigma\in\mathcal{A}}\|\mathcal{G}^{RTE}(\sigma)\|_\infty, \sup_{\sigma\in\mathcal{A}}\|\mathcal{G}^{DE}(\sigma)\|_\infty\right\} \le C.$$

Then in [31, Theorem 3.2] the authors conclude from the convergence of the forward maps, the convergence under the Kullback-Leibler divergence and thus also under the Hellinger metric.

Theorem 5.12 *Suppose that the prior distribution of σ is supported on the admissible set and the two forward maps $\mathcal{G}^{RTE}$ and $\mathcal{G}^{DE}$ converge. Then it follows for the posterior distributions of RTE and DE, μ^y_{RTE} and μ^y_{DE}, that*

$$d_{KL}(\mu^y_{\mathrm{RTE}}, \mu^y_{\mathrm{DE}}) \le \mathcal{O}(\varepsilon).$$

Remark 5.13 (Order of Convergence) Note, that in [31, Theorem 3.2] even an order of the convergence in the Kullback-Leibler divergence could be accomplished. They also investigate the order of convergence for a linearized radiative transfer equation, which can then be calculated to be of $\mathcal{O}(\varepsilon^2)$ [31, Chapter 4]. For the inverse problem of chemotaxis, no order of convergence could be established, yet.

From this short overview of the multiscale analysis of stationary radiative transfer, one gets an idea about the similarities - and differences - to the inverse problem for chemotaxis. The kinetic equations for both problems underlie the same structure and can both be approximated by a diffusion equation in the velocity independent case. Also the convergence of the forward problems was proved in a similar fashion than in our case, by plugging an expansion for f into the kinetic equation [31]. Additionally,

the convergence of the inverse problems could be shown in the same metrics, namely the Kullback-Leibler divergence and the Hellinger metric. Because of all those parallels, we attempted to extend the results of [28], where the ill-conditioning of inverse radiative transfer in the diffusive limit was shown, to our case.

The biggest difference of the two problems lies within the kinds of measurements one considers. Whereas we had an initial value problem and measured the density for different time steps at different locations, the stationary RTE is a boundary value problem and the measurements are only taken on the boundary as well. Hence, different „arithmetic tricks“ could be used. Throughout the proofs of this thesis we often took advantage of zero-initial conditions when looking at the differences of two solutions of the chemotaxis equation by using Grönwall's inequality, for instance when proving that f_ε is a Cauchy-sequence in section 3 or when investigating the derivative of the parameter-to-solution map in section 5. For proving the ill-conditioning in the diffusive limit of RTE in [28], they use the fact that their derivative of the parameter-to-solution map has zero boundary condition, thus this also follows for w_0 in their case. Since w_0 does not depend on the velocity for stationary RTE, the zero boundary condition on Γ_- can then be expanded onto the whole boundary of the domain, leading to the measurements on Γ_+ being zero as well. Then, from looking at $\mathcal{G}'(\sigma_*)\sigma = \int_{\Gamma_+} w(x,v)v \cdot n\mathrm{d}v$, they can use that the integral over w_0 vanishes. Therefore, we tried to get a similar result, by taking advantage of the zero initial condition on w_0 in our case of chemotaxis. But our measurements are not taken in a way, that we get this outcome as natural as it comes for the stationary boundary-value problem.

6 Summary and Outlook

In the following, the results of the previous chapters are briefly summarised and an outlook on further research is given. Within this work an overview of Bayesian Inference was given in section 2. This was done by first introducing the conventional approach of solving an inverse problem by the method of least squares and linking the solution arising from this approach to the Maximum a Posteriori (MAP) point of the posterior distributions coming from Bayesian inversion. More precise, in the finite-dimensional setting with Gaussian prior and noise, the MAP estimator corresponds to the solution of a weighted least square problem. The influence of the prior on the solution was discussed and put the choice of prior information into perspective. Afterwards, the concept of well-posed solutions was established for distributions in the Bayesian setting. Under certain assumptions on the prior and the potential, the posterior distribution resulting from Bayes theorem is well defined and stable with respect to perturbations in the data. The chapter was complemented by the discussion of selected numerical algorithms, such as Markov Chain Monte Carlo algorithms or variational and filtering methods, which are used in order to draw information from Bayes formula.

Before Bayesian inversion can be applied to the inverse problem of chemotaxis, the two forward models on the kinetic and macroscopic level were introduced in section 3. In this chapter, the phenomenon of chemotaxis was introduced as well as the mathematical models which describe the forward problems. In the case of bacteria that move by rotating their flagella, the model on the kinetic level is a run and tumble model consisting of a term that describes the transport part and a tumbling kernel, which describes the random process of the change of direction. Here, a probability distribution dependent on time, space and velocity is described by the chemotaxis equation. On the macroscopic level a drift-diffusion-equation, namely the Keller-Segel model, was introduced and its properties described. The two models are asymptotically equivalent in the long-time-large-space regime and the goal of this chapter was to extend this result for the case when a reaction term is introduced to the system. We used the Fisher Term in order to represent the effect of birth and death of bacteria, which depends on the distribution function. The asymptotic equivalence of the extended system could be shown, if one assumes the occurring densities to be bounded and that the kernel of the zeroth order term of the tumbling kernel is one dimensional and spanned by a velocity distribution.

Building up on the results for the forward problem, the inverse problem in the Bayesian setting was introduced in section 4. After establishing the asymptotic limit for the forward problem, the convergence of the corresponding inverse problems with a reaction term shall be investigated. First, the question which quantities have to be inferred in order to be able to compare the inverse problems due to the different scales of the model was conquered. Then the measurements were introduced and the process of collecting data was discussed. Afterwards, it was shown that the posterior distributions resulting from Bayes' rule are well-defined and the convergence under the Kullback-Leibler divergence and the Hellinger-metric could be proven. The results were concluded while putting assumptions on the structure of the tumbling kernel, such as boundedness and (anti-)symmetry as prior information.

The work is concluded by the analysis of the loss of information during the scaling process for chemotaxis without a reaction term in section 5. On the global level, the loss of information was attempted to be quantified by the difference between the prior and posterior distributions under the Kullback-Leibler divergence. Since the posterior takes the prior as well as the data into account, a small difference of the two distributions would mean that the data has little impact on the posterior distribution. On the local level, the flatness of the posterior distribution around the MAP point was investigated. A rather flat distributions implies, that the MAP point cannot be determined uniquely making the reconstruction of the tumbling kernel insensitive to the data. Both, the global and the local loss of information could not be quantified within this work. But the approach was compared with similar results for the stationary radiative transfer equation and can be seen as a first attempt to investigate the loss of information for chemotaxis in the diffusive limit.

We want to conclude the thesis, by giving an outlook on further research an current developments within this field.

The Analysis of the stability deterioration did not yield a quantification of the loss of information in the diffusive limit, thus further research is needed for this topic. For instance, one could try to use a different metric that may be better suited for the types of measurement we take. Another approach would be to show, that the desired statement $\mathcal{G}' = \mathcal{O}(\varepsilon)$ is not true, meaning that there would be a gain of information in the diffusive limit. In order to do so, one would have to prove that $\int_{\mathbb{R}^3} \int_V w_0 \mathrm{dv}\chi \mathrm{dx} \geq \mathcal{O}(1)$. But since we expect to have some sort of information loss when entering the macroscopic regime, this would be a rather unintuitive outcome.

The results of the stability deterioration in the case of radiative transfer in [28] where $\mathcal{G}' = \mathcal{O}(\varepsilon)$ could be shown, suggest that the Bayesian approach may not be suited for all kinds of inverse problems. With Bayesian inference, the unique determination of the quantity of interest arises rather naturally from Bayes' theorem and can be a good foundation for some numerical algorithms, such as Markov chain Monte Carlo algorithms, but it may not yield the most reasonable solution. Thus, the question arises if one should use another approach to inverse problems. In [18], Hellmuth et. Al. approach the inverse problem for chemotaxis with the method of *singular decomposition*, a technique designed to investigate inverse problems coming from kinetic theory. They could show that given a special design of initial data, the population density - a macroscopic quantity - suffices to reconstruct the tumbling kernel, in the case of no occurring birth-death-term and with a known chemical stimulus. Hence, further research of the inverse problem coming from chemotaxis could also include to extend the results of [18] to more complex cases.

Overall, the multiscale analysis of the inverse chemotaxis problem still yields various open research topics. Whether it be the question of ill-conditioning in the Bayesian setting when entering the diffusive limit or of applying other methods to the inverse problem, such as singular decomposition, in order to reconstruct the tumbling kernel from the macroscopic bacteria density. Understanding the relation between the inverse problems coming from the kinetic and macroscopic framework respectively, not only gives more insights about chemotaxis, but the established theory can serve as a guidance for other kinetic equations as well.

References

[1] Chen, K.; Li, Q.;Wang, L. *Stability of stationary inverse transport equation in diffusion scaling.* https://doi.org/10.1088/1361-6420/aa990c, 2018.

[2] Lai, R.Y.; Li, Q.; Uhlmann, G. *Inverse Problems for the Stationary Transport Equation in the Diffusion Scaling.* https://doi.org/10.1137/18M1207582, 2019.

[3] Stigler, S.M. *The history of statistics : the measurement of uncertainty before 1900.* 2. print. - Cambridge, Mass. [u.a.]: Belknap Press of Harvard Univ. Press, XVI, 410 S., 1986.

[4] Alt, W. . *Biased random walk models for chemotaxis and related diffusion approximations.* https://doi.org/10.1007/BF00275919, 1980. 147-177.

[5] Arridge, S.R. *Optical tomography in medical imaging.* https://doi.org/10.1088/0266-5611/15/2/022, 1999.

[6] Bal, G., Jollivet, A. *Time-dependent angularly averaged inverse transport.* Inverse Problems, vol. 25, 075010, 2009.

[7] Bal, G., Langmore, I., Monard, F. *Inverse transport with isotropic sources and angularly averaged measurements.* http://doi.org/10.3934/ipi.2008.2.23, 2008.

[8] Bayes, T. *An Essay towards solving a Problem in the Doctrine of Chances.* https://doi.org/10.1098/rstl.1763.0053, 1763.

[9] Beal, J., Farny, N., Haddock-Angelli, T., Vinoo Selvarajah, V., Baldwin, G.S., Buckley-Taylor, R., Gershater, M., Kiga, D., Marken, J., Sanchania, V., et al. . *Robust estimation of bacterial cell count from optical density.* https://doi.org/10.1038/s42003-020-01127-5, 2020.

[10] Chalub, F., Markowich, P., Perthame, B., Schmeiser, C. *Kinetic Models for Chemotaxis and their Drift-Diffusion Limits.* https://doi.org/0.1007/s00605-004-0234-7, 2004.

[11] Choulli, M., Stefanov, P. *Inverse scattering and inverse boundary value problems for the linear Boltzmann equation.* https://doi.org/10.1080/03605309608821207, 1996.

[12] Dashti, M., Stuart, A. *The Bayesian Approach to Inverse Problems.* https://doi.org/10.48550/arXiv.1302.6989, 2015.

[13] Evans, L. *Partial Differential Equations.* American Mathematical Society, Graduate Studies in Mathematics Vol. 19, 2010. Second Edition.

[14] Fornacon-Wood, I., Mistry, H., Johnson-Hart, C., Faivre-Finn, C., O'Connor, J.P.B., Price, G.J. *Understanding the Differences Between Bayesian and Frequentist Statistics.* https://doi.org/10.1016/j.ijrobp.2021.12.011, 2021.

[15] Hammes, F., Egli, T. *Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications.* https://doi.org/10.1007/s00216-010-3646-3, 2010.

[16] Haug, A.J. . *Bayesian Estimation and Tracking : A Practical Guide.* John Wiley & Sons, Incorporated, 2012.

[17] Hellmuth, K., Klingenberg, C., Li, Q., Tang, M. Multiscale convergence of the inverse problem for chemotaxis in the bayesian setting. https://doi.org/10.3390/computation9110119, 2021.

[18] Hellmuth, K., Klingenberg, C., Li, Q., Tang, M. Kinetic chemotaxis tumbling kernel determined from macroscopic quantities. https://doi.org/10.48550/arXiv.2206.01629, June 2022.

[19] Hielscher, A.; Alcouffe, R.; Barbour, R. *Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues.* https://doi.org/10.1088/0031-9155/43/5/017, 1998.

[20] Horstmann, D. . *From 1970 until present: the Keller-Segel model in chemotaxis and its consequences I.* Jahresberichte der DMV 105, 2003. pp.103-165.

[21] Horstmann, D. . *From 1970 until present: the Keller-Segel model in chemotaxis and its consequences II.* Jahresberichte der DMV 106, 2003. pp.51-69.

[22] Horstmann, D., Winkler, M. . *Boundedness vs. blow-up in a chemotaxis system.* https://doi.org/10.1016/j.jde.2004.10.022, 2005. 52-107.

[23] Huber, R. Variational regularization for systems of inverse problems: Tikhonov regularization with multiple forward operators. https://doi.org/10.1007/978-3-658-25390-5, 2019.

[24] Hunter, J. *Notes on Partial Differential Equations.* Department of Mathematics, University of California at Davis, 2014.

[25] Keng, B. https://bjlkeng.github.io/posts/variational-bayes-and-the-mean-field-approximation/. [Online; accessed 12th October 2022].

[26] Kowalczyk, R. . *Preventing blow-up in a chemotaxis model.* J. Math. Anal. Appl. 305, 2005. no. 2, 566-588.

[27] Kullback, S., Leibler, R.A. . *On Information and Sufficiency.* https://doi.org/10.1214/aoms/1177729694, 1951.

[28] Li, Q., Newton, K., Wang, L. *Bayesian Instability of Optical Imaging: Ill Conditioning of Inverse Linear and Nonlinear Radiative Transfer Equation in the Fluid Regime.* https://doi.org/10.3390/computation10020015, 2022.

[29] Lohmann, G. *Temperatures from energy balance models: the effective heat capacity matters.* https://doi.org/10.5194/esd-11-1195-2020, 2020.

[30] Nadin, G, Perthame, B., Ryzhik, L. . *Traveling waves for the Keller–Segel system with Fisher birth term.* Interfaces and Free Boundaries, July 2007.

[31] Newton, K., Li, Q., Stuart, A.M. *Diffusive Optical Tomography in the Bayesian Framework.* https://doi.org/10.1137/19M1247346, 2020.

[32] Ng, A. Y. *Feature selection, L1 vs. L2 regularization, and rotational invariance.* https://doi.org/10.1145/1015330.1015435, 2004.

[33] Othmer, H.G., Dunabr, S.R., Alt, W. . *Models of dispersal in biological systems.* https://doi.org/10.1007/BF00277392, 1988. 263-298.

[34] Perthame, B. *Transport Equations in Biology.* Birkäuser Verlag, Basel - Boston - Berlin, 2007.

[35] Phjolainen, S., et. Al. *Mathematical Modelling.* https://doi.org/10.1007/978-3-319-27836-0, 2016.

[36] Stuart, A. M. *Inverse Problems: A Bayesian Perspective.* https://doi.org/10.1017/S0962492910000061, 2010.

[37] Sullivan, T.J. *Introduction to Uncertainty Quantification.* https://doi.org/10.1007/978-3-319-23395-6, 2015.

[38] Teschl, G. *Ordinary Differential Equations and Dynamical Systems.* Graduate Studies in Mathematics, Vol. 140, 2012.

[39] Tsybakov, A.B. . *Introduction to Nonparametric Estimation.* https://doi.org/10.1007/b13794, 2009. 1st edition.

[40] U.S. Department of Energy. *Scientific Grand Challenges for National Security: The Role of Computing at the Extreme Scale.* U.S. Department of Energy, 2009.

[41] van Ravenzwaaij, D., Cassey, P., Brown, S.D. . *A simple introduction to Markov Chain Monte–Carlo sampling.* https://doi.org/10.3758/s13423-016-1015-8, 11 March 2016.

[42] Wang, T. *Climatic & Ecological Modelling for Adaptive Forest Applications.* E-Learning Course of the University of British Columbia, 2021. Topic 2.1 - Introduction to Climate Models.

[43] Wiecki, T. https://twiecki.io/blog/2015/11/10/mcmc-sampling/. [Online; accessed 12th October 2022].

[44] Williams, P.H. *Model Building in Mathematical Programming.* John Wiley & Sons, Incorporated, 2013. 5th Edition.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit in allen Teilen selbstständig angefertigt und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt habe. Sämtliche wörtlichen oder sinngemäßen Übernahmen und Zitate sind kenntlich gemacht und nachgewiesen.

Würzburg, den

_____

Kaja Alina Jurak