On Qualitative Experimental Design for PDE Parameter Identification Inverse Problems

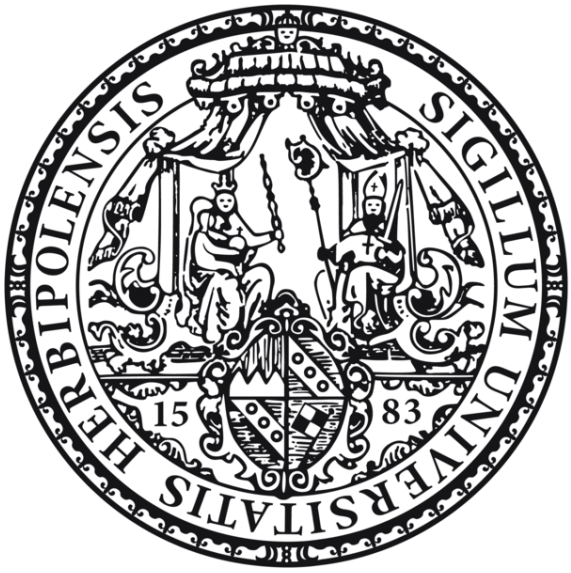
- Doctoral Thesis -

submitted by

Kathrin Hellmuth

under the supervision of

Prof. Dr. Christian Klingenberg



Julius-Maximilians-Universität Würzburg

Faculty of Mathematics and Computer Science

Würzburg, February 2025

Abstract

In inverse problems, the feasibility and quality of a reconstruction for a model parameter from indirect experimental measurements strongly depends on the quality of the available data. It is the goal of experimental design to select suitable experimental setups that generate informative data for this task, which is classically approached by an optimization framework through optimal experimental design techniques. In this work, we lay out two alternative approaches, that are based on sensitivity analysis and rooted in the identifiability analysis framework, allowing them to contribute in bridging the gap between the theoretical input-to-output map based analysis, and the practical finite data setting. As they search for a sufficient instead of an optimal design, these approaches are of qualitative nature.

The first approach derives the finite experimental design through relaxation of a possibly infinite experimental designs on which the theoretical uniqueness proof is based. It is applied to an inverse problem in mathematical biology that seeks to reconstruct the mesoscopic chemotaxis tumbling parameter, that describes the directional change of bacteria in their directed random walk towards attracting chemical stimuli, from macroscopic data on the bacterial density. After developing a uniqueness proof, that is based on an explicit construction of a sequence of experimental designs whose measurements are informative about the parameter by means of the singular decomposition technique, we relax this non-parametric infinite dimensional model parameter to a parametric, locally constant form that in turn allows relaxation to one feasible finite design. Similar strategies as in the uniqueness proof provide sufficiency of this design analytically, which is underpinned by numerical results.

The second approach is more generally applicable, and assumes a predefined parametric form of the model parameter. Design selection can then be regarded as a down-sampling task from the large set of all possible setups in the input-to-output map. The preservation of sensitivity of the design w.r.t. the parameter throughout the down-sampling process is ensured by invoking an importance sampling distribution, that is derived from a matrix sketching algorithm from randomized linear algebra for the sensitivity matrix. We propose a numerical pipeline for implementation and numerical tests demonstrate the potential of this method.

Zusammenfassung

Der Erfolg der Rekonstruktion eines Modellparameters aus indirekten experimentellen Messungen durch Lösen des inversen Problems hängt maßgeblich von der Qualität der zur Verfügung stehenden Daten ab. Ziel der Versuchsplanung ist daher die Auswahl von geeigneten experimentellen Setups, die für diese Aufgabe informative Daten erzeugen. Dies wird im Zuge der optimalen Versuchsplanung klassischerweise als Optimierungsproblem formuliert. In dieser Arbeit stellen wir zwei alternative Ansätze vor, die auf einer Sensitivitätsanalyse basieren und in der Identifizierbarkeitsanalyse beheimatet sind. Diese tragen dazu bei, die Forschungslücke zwischen der theoretischen Analyse des Problems auf Basis der input-to-output Abbildung und der praktischen Anwendung mit begrenzten Daten zu verstehen. Da nach einem hinreichenden anstelle eines optimalen Designs gesuchen wird, sind diese Ansätze qualitativer Natur.

Der erste Ansatz leitet das endliche experimentelle Design durch eine Relaxation eines konkreten, möglicherweise unendlichen experimentellen Designs aus geeigneten konstruktiven theoretischen Eindeutigkeitsbeweisen her. Wir wenden diesen Ansatz auf ein inverses Problem in der mathematischen Biologie an, das darauf abzielt, den mesoskopischen Chemotaxis-Tumbling-Parameter auf Basis von makroskopischen Daten zur Bakteriendichte zu rekonstruieren, der die Richtungsänderung von Bakterien in ihrem gerichteten Zufallsbewegungsmuster hin zu anziehenden chemischen Stimulus beschreibt. Zunächst wird ein Eindeutigkeitsbeweis basierend auf der input-to-output Abbildung entwickelt, der mithilfe der Singular Decomposition Technik eine Folge von experimentellen Designs explizit konstruiert, deren Messungen den gesuchten Parameter widerspiegeln. Durch Relaxation des nichtparametrischen, unendlichdimensionalen Modellparameters zu einer parametrischen, lokal konstanten Form kann auch die konstruierte Folge von Designs zu einem machbaren, endlichen Design relaxiert werden. Ähnliche Argumente wie im Eindeutigkeitsbeweis belegen die Eignung dieses Designs analytisch, was durch numerische Ergebnisse untermauert wird.

Der zweite Ansatz ist allgemeiner anwendbar, setzt dafür aber eine vordefinierte parametrische Form des Modellparameters voraus. Die Auswahl des Designs kann dann als ein Downsampling-Problem aus einer großen Menge aller möglicher Setups betrachtet werden. Die Erhaltung der Sensitivität des Designs in Bezug auf den Parameter wird im Downsamplingprozess durch die Verwendung einer Importance-Sampling-Verteilung sichergestellt, die aus einem Matrix-Sketching-Algorithmus der randomisierten linearen Algebra für die Sensitivitätsmatrix abgeleitet wird. Wir empfehlen eine numerische Pipeline zur Implementierung, und demonstrieren das Potenzial dieser Methode in numerischen Tests.

Acknowledgements

I would like to thank several people that have shaped my personal and scientific development in the past years.

First and foremost, I would like to express my deep gratitude to my advisor, Christian Klingenberg, for his constant and dedicated support and guidance, and his expertise. I greatly benefited from the international, high-level scientific environment that he provided, be it through international guests in our group seminars or through our numerous scientific travels, which I greatly enjoyed. His training prepared me well for the scientific world.

My research project emerged from this environment through a cooperation with my coauthors Qin Li and Min Tang. I am incredibly thankful for their active support and the opportunity to learn from their immense expertise, which greatly shaped my understanding. I very much enjoyed our regular Zoom discussions, and deeply admire the scientific creativity that each of them demonstrated. Moreover, I am particularly grateful for the opportunity to visit each of them, and work together in person.

In addition, I would like to thank all researchers that have crossed my path for the engaging discussions. I want to particularly mention Herbert Egger, Nora Philippi and Matthias Schlottbom for the opportunity to broaden my scope during our joint project, and Maria Han-Veiga for her invaluable support in setting up the lecture and exercise classes on 'Mathematics of Machine Learning'.

On this stage, I would also like to thank my current and former colleagues in the work group for the welcoming atmosphere and great joint experiences.

My PhD project would not have been possible without the funding I obtained from the WMCCI (Würzburg Mathematics Center of Communication and Interaction), as well as the PhD scholarship from the German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes). I am honoured by their trust, and greatly appreciated the extracurricular activities. Moreover, the additional support through the Marianne-Plehn-Program allowed me to significantly improve my teaching skills.

On a personal level, I would like to thank all my friends for their understanding and encouragement, especially during the past months.

I consider myself very fortunate to have grown up in a family that has always supported me in my decisions, made my studies possible, and always cared for me. A big thank you goes to my parents for all their love, and to my brother his support by equipping me with the screen that this thesis was written on.

Finally, I want to particularly thank Alex for his endless love, understanding and emotional support that helped me persevere even tough times I faced when preparing this thesis.

Kathrin Hellmuth

Contents

Contents v

1. Introduction 1

2. Inverse Problems 5
- 2.1. General Formulation . . . 5
- 2.2. Cost Minimization . . . 8
  - 2.2.1. Numerical Optimization by Gradient Descend . . . 9
- 2.3. The Bayesian Approach . . . 10
  - 2.3.1. Bayesian Posterior Sampling . . . 12
  - 2.3.2. Connection between Bayesian and Cost Minimization . . . 15
- 2.4. Identifiability analysis . . . 16
  - 2.4.1. Structural identifiability . . . 18
  - 2.4.2. Cost Function and Sensitivity Based Identifiability . . . 18
  - 2.4.3. Practical identifiability . . . 22
  - 2.4.4. Dealing with Non-Identifiability. . . . 22
  - 2.4.5. Literature Survey . . . 23
- 2.5. Optimal Experimental Design . . . 28
  - 2.5.1. The Basic Idea . . . 28
  - 2.5.2. Optimality Criteria. . . . 29
  - 2.5.3. Sequential Optimal Experimental Design. . . . 31

I. Relaxation of Theory Approach 33

3. The Inverse Problem for Chemotaxis 35
- 3.1. Modelling Chemotaxis . . . 35
  - 3.1.1. Biological Background. . . . 36
  - 3.1.2. Kinetic Chemotaxis Model . . . 38
  - 3.1.3. Chemotaxis Models across the Scales . . . 40
  - 3.1.4. Refined Models . . . 43
- 3.2. Inverse Problem for Chemotaxis . . . 44
  - 3.2.1. Experimental Setting . . . 45
  - 3.2.2. Research question. . . . 47
  - 3.2.3. Literature on Inverse Problems related to Chemotaxis. . . . 48

3.2.4. Novelty of this Work and Outline of Part I. 50

4. Structural identifiability: Experimental Setting **51**
4.1. Setting and Main Results. 51
4.2. Proof of Theorem 4.1 in Dimension $d = 2, 3$ 56
4.2.1. The Ballistic Part f_0 63
4.2.2. The One Tumble Part f_1 64
4.2.3. The Multiple Tumble Part $f_{\geq 2}$ 68
4.3. Proof of Theorem 4.1 in Dimension $d = 1$ 69
4.3.1. The Ballistic Part f_0 74
4.3.2. The Tumble Part $f_{\geq 1}$ 74

5. Sensitivity Based & Cost Function Identifiability: Experimental Design **77**
5.1. General Setting 79
5.2. Qualitative Experimental Design 83
5.2.1. Non Identifiability for Close Measurement Specifications 83
5.2.2. An Identifiable Design in Dimension $d = 1$ 84
5.2.3. Extension to $d = 2$ 90
5.3. Proofs of this Chapter 93
5.3.1. Proof of Lemma 5.5 93
5.3.2. Proof of Theorem 5.6 95
5.3.3. Proof of Proposition 5.14 96

6. Numerical experiments **101**
6.1. Numerical Setting 102
6.1.1. Optimization by Gradient Descent 102
6.1.2. Calculation of the Gradient 102
6.1.3. Integration of Advanced Methods. 103
6.1.4. Computational Setting. 104
6.2. 1D Numerical examples. 105
6.2.1. Well-Posedness of Design (1D). 105
6.2.2. Ill-conditioning for Close Experimental Setups 109
6.2.3. A Locally but Non-Globally Convex Design 113
6.2.4. Comparison of the Convexity of considered Designs. 115
6.3. Extension to 2D 116

7. Conclusion of Part I **119**

II. Sampling Approach **123**

8. Experimental Design through Sampling **125**
8.1. Prerequisites: Randomized Quasimatrix Multiplication 128
8.2. General Program 130
8.2.1. Starting Point 130

8.2.2. Experimental Design as Sampling . . . 132
8.2.3. Importance Sampling and Matrix Sketching . . . 132
8.2.4. Choice of Sampling Algorithms . . . 134
8.2.5. Greedy Sampling . . . 135
8.3. Application to the Schrödinger Potential Reconstruction . . . 136
8.3.1. Importance Sampling Distributions. . . . 138
8.3.2. Effect of Sampling . . . 140
8.3.3. Extension to Controllable Source . . . 144
8.4. Conclusion . . . 147
8.5. Derivation of the formula for $\nabla_p u_p(\hat{x})$ . . . 149

Discussion **153**

9. Discussion **153**

Appendices **159**

A. Existence Theory **159**
A.1. Existence Results for the Kinetic Chemotaxis Equation via Semigroup Theory 159
A.1.1. Forward Problem Existence . . . 160
A.1.2. Adjoint Problem Existence . . . 165
A.2. Existence Results for the Stationary Schrödinger Equation via Elliptic Theory 167
A.2.1. L^2 Data . . . 167
A.2.2. Dirac Data . . . 168

B. Transformations on the Sphere **171**
B.1. Stereographic projection . . . 171
B.2. On the transformation $\mathcal{T}_a^w$ . . . 174

C. Numerical Schemes for PDE Discretization **177**
C.1. Finite Difference Schemes for Transport Equations . . . 177
C.1.1. Review on Properties of Numerical Schemes. . . . 177
C.1.2. The Lax Wendroff Scheme . . . 179
C.1.3. Finite Difference Schemes for the Forward and Adjoint Chemotaxis Equations . . . 180
C.2. Finite Element Scheme for the Stationary Schrödinger Equation. . . . 182

Bibliography **187**

1 Introduction

Mathematical models have proven valuable tools to describe phenomena in various areas of life, be it in physics, biology, engineering, medicine, economics or even in social studies [Bol72, KS70, LL92, RS08, MT14]. These models often include adjustable parameters, which give them the flexibility to describe the phenomenon in different scenarios and often reflect physical characteristics of the phenomenon, for instant, a speed of motion or a material specific optical density. Parameter identification inverse problems study the reconstruction of parameters, which are not directly observable, from experimental data, as obtained from real world observation of the phenomenon. If successful, this provides valuable insight into the physical characteristics, describing for instance the composition of tissue in medical imaging [RS08], and allows simulation and prediction of the phenomenon's future behaviour in the considered scenario by means of the fitted model [BC09]. Obtaining a good parameter reconstruction, however, is in many cases anything but simple, and depends predominantly on the quality of the available experimental data with respect to (w.r.t.) the considered reconstruction task. Identifiability analysis provides a holistic framework to categorize this quality according to three sources of information degradation [GDI85], in order to allow a structured search for remedies:

(i) In the worst case, the experiment is intrinsically uninformative of the parameter, no matter how exactly data is collected.

(ii) The experiment might be informative in general, but suffer from a poor choice in data collection, as described by the experimental design, for instance measurements may be taken at uninformative observation locations.

(iii) Noise might pollute the data and distort the reconstruction up to a degree where it cannot be trusted anymore - this is related to stability issues as encountered in the famous Calderon problem [Cal80].

The framework is wide spread in ordinary differential equation (ODE) inverse problems [GDI85, BÅ70, LG94], but application to partial differential equation (PDE) parameter identification is rather limited.

In this thesis, we shall focus on experimental design (ii) for PDE parameter identification, that is (i.e.) the question how to select a finite number of data that is informative about the parameter. We lay out two approaches that contrast the classical optimal experimental design methodology, and relax the search to a sufficient instead of an optimal design, putting them in the realm of qualitative experimental design. The identifiability analysis framework above suggests to consider the designing process as a transition from an informative infinite set of data as given by the input-to-output map, that summarizes the data of all possible experimental setups as considered in (i), to a practically more feasible finite data set that is still capable of representing the parameter according to (ii). This work thus contributes to bridging the gap between theory of inverse problems in the infinite-data setting and the practical finite data setting, which is poorly understood, and developed methodologies in parts diverge [BJ09, Ren10].

Approach 1: Relaxation of Theory. The first approach describes a relaxation of a suitable theoretical proof of identifiability under source (i), and allows a simultaneous, compatible dimension reduction of the data and the parameter. This parameter discretization becomes necessary, especially for non-parametric infinite dimensional parameter inverse problems, when going over to a finite amount of data that can only characterize a finite dimensional parameter. To sustain identifiability, a-priori information on the parameter is injected that leverages the loss of information in the data. This approach is exemplified for a specific parameter identification problem for a kinetic PDE model from mathematical biology, where inverse problems are still at their fancy, but become more and more relevant as more refined measurement techniques allow finer modelling of biological phenomena. The model describes the directed motion of bacteria in response to an external stimulus such as a chemical substance, termed chemotaxis. Our goal is to determine the microscopic parameter that uniquely describes their motion through the likelihood of changing the running direction. We assume the experimental data on the bacterial density to be of sub-optimal macroscopic type, i.e. non-local in the movement direction, which is more tractable for practitioners, but poses challenges to the reconstruction [Bal09, BLM08, Lan08, BJ09, BJ10, BJJ10]. This precise inverse problem has not been studied in literature before, and in particular the experimental design aspect under (ii) is novel to inverse problems related to chemotaxis phenomena.

In a first project we develop the proof of general suitability of this experiment to generate informative data, in the sense of (i). Assuming, theoretically, access to perfect data according to characteristics (ii) and (iii), i.e. access to the noise free infinite-to-infinite dimensional input-to-output map, we invoke an existing technique for kinetic parameter identification, termed singular decomposition [LS20, CS96a, Bal09], to prove analytically that this input-to-output map uniquely determines a continuous bacterial motion parameter. The proof explicitly constructs a sequence of experimental setups, that trigger microscopic information on the turning parameter by inducing increasingly singular initial data, and exploits the freedom of prescribing spatially and directionally concentrated initial data for the bacteria and availability of small time, local interior domain measurement data in this setting. This is the first result to study the combination of all these factors and to obtain unique identifiability of a kinetic turning parameter from directionally averaged

data without any further simplifying assumptions, and has been published in [HKLT24].

In a second project, published under [HKLT25], we then conduct the relaxation, first in the parameter by imposing a locally constant form of the tumbling kernel, which then allows relaxation of the singularity and small measurement time requirement in the previously considered experimental designs, so to obtain a finite number of practically feasible experimental setups. A similar methodology as before allows us to analytically prove suitability of the proposed combination of the parameter discretization and experimental design for reconstruction. This is in contrasted by a study on the decay of information in the data if experimental setups are not sufficiently diverse, which demonstrates the need for well chosen designs. These theoretical results are illustrated in numerical experiments that demonstrate the decay of sensitivity of the data and with it the corresponding parameter reconstructions under decreasing diversity of the experimental setups, and a good sensitivity and reconstruction behaviour under the proposed design.

Approach 2: Sampling. The second approach assumes a prescribed finite-dimensional parametric form of the model parameter and only seeks to reduce the data dimension. Its goal and setting thus coincide with those in optimal experimental design, however, the shift in perspective to search for sufficient instead of optimal designs allows us to relax the experimental design process to a sampling task that can benefit from existing strategies developed for randomized numerical linear algebra [Mah16, MT20]. This perspective and methodology is novel to experimental design. We then propose a numerical pipeline to execute the sampling based on Bayesian posterior samplers and a greedy mechanism and illustrate the potential of this strategy to improve sensitivity of the data w.r.t. the parameter on numerical examples for the sensor placement problem for the Schrödinger potential reconstruction problem. This work is summarized in the thus far unpublished preprint [HKL24].

The thesis is structured as follows. Chapter 2 provides an introduction to inverse problems, with a particular focus on introducing terminology relevant to parameter identification tasks, before the identifiability analysis framework is laid out, followed by a brief overview on optimal experimental design. Part I is then devoted to approach 1 and starts with an introduction of the kinetic chemotaxis model and the corresponding inverse problem under investigation in Chapter 3. The theoretical proof of suitability of the considered experimental setting according to (i) is developed in Chapter 4 and Chapter 5 then lays out the relaxation approach to discretize the motion parameter and construct a finite experimental design, which is then analytically studied for its suitability according to (ii). The analytical framework also demonstrates a decay of information on the parameter in the data under decreasing data diversity. These results are illustrated in a numerical context in Chapter 6. Chapter 8 in Part II then develops the sampling strategy for experimental design and justifies it in numerical experiments. Applicability of both approaches is discussed in Chapter 9 that concludes the thesis with a brief outlook on possible future directions of research.

2

Inverse Problems

In a large variety of contexts, quantities of interest cannot be observed directly. Examples can be found in medical imaging [RS08], where the interior of the body shall be examined without invasive surgery, or similarly in non-destructive testing in engineering [Bil20], in subsurface structure analysis in geophysics [ILS14], where drilling is expensive and thus avoided if possible, in image deblurring [BBDM21], where only a blurry version of the original image is available, or in air pollutant emission tracing in environmental physics [ZK20], where the locations of pollutant sources are searched according to measurements of a fixed monitoring network, just to mention some. Another important application is the fitting of mathematical models through identification of non-observable model parameters, for instance in the famous Calderon problem [Cal80].

In all these cases, quantities of interest have to be inferred from experimental data collected from related observable quantities. This means taking the inverse perspective.

2.1. General Formulation

In the following, a short introduction to inverse problem is given in order to fix notation, following [Bal12]. A plethora of further introductory literature exists, and the interested reader is referred for example (e.g.) to [Kir21, Gro93, FSU19].

Mathematical formulation. Information on the non-observable quantity of interest p, termed parameter, shall be acquired from experimental data $y \in \mathcal{Y}$, in some data space $\mathcal{Y}$. The map that assigns the corresponding data to a given parameter value is termed the forward operator and reads

$$F : \mathcal{A} \to \mathcal{Y}.$$

It is defined on the admissible set $\mathcal{A}$ that incorporates the shape as well as a-priori knowledge on the parameter p, and is also called the prior model in [Bal12]. In what follows, $\mathcal{A}$ and $\mathcal{Y}$ are considered subsets of normed vector spaces.

The inverse problem amounts in inverting F, i.e. in finding the parameter $p_\star \in \mathcal{A}$ such that the given data y, that was experimentally observed, is reflected well by the measurement of an experiment with underlying parameter $p_\star$, in short

$$\text{find} \quad p_\star \in \mathcal{A}_\star \quad \text{such that} \quad F(p_\star) = y, \tag{2.1}$$

where the equality might be changed for an approximate equality for instance in the context of noisy data. This holds contrast to the forward problem, where $F(\hat{p})$ is computed for a given parameter $\hat{p}$. In both cases, the forward operator F is known.

PDE Parameter Reconstruction. Parameter identification problems for differential equations have a long history as they appear in many applications. Probably the most famous example is the Calderón problem from electrical impedance tomography that stimulated research in the area of inverse problems [FSU19] since its proposal in 1980 [Cal80]. Because they are often based on the observation of some real world phenomenon, these inverse problems frequently share a specific structure that shall be introduced in the following, together with some nomenclature.

Data Generation. The data is generated by real world experiments on the phenomenon which are initiated by imposing forcing data ϕ on the observable quantity, for instance an initial or boundary configuration. The evolution of the observable quantity is observed through detectors, with potentially variable profiles μ, yielding the data $y_{\phi,\mu}$. An experiment is thus described by the choice of the so-called input data $(\phi, \mu) = s$ which we will also call an experimental setup s in the following.

To conduct the inverse problem, this experiment is translated to the mathematical level, where the evolution of the observable quantity u is modeled by the PDE model involving the unknown parameter p, i.e. $u = u_p^\phi$ denotes the solution to the PDE with parameter p and forcing data ϕ. The measurement by the detector is mimicked mathematically by a measurement operator M_μ. In total, the forward map reads

$$F_s(p) = F_{\phi,\mu}(p) = M_\mu(u_p^\phi).$$

A graphical description is given in Figure 2.1. The inverse problem seeks to find that parameter value $p \in \mathcal{A}$ for which the synthetic data $F_s(p)$ approximates the experimentally observed data y_s well. In many contexts, forcing data ϕ and measurement specifications μ can be influenced by experimenters. Then testing with different input data can enrich the data set and frequently ameliorates the inverse problem, when different aspects of the parameter can be triggered by different input data or measured by different measurement specifications. Moreover, generalizability of the found parameter to generic input data may be enhanced, if experimental setups are chosen appropriately. Collecting all conducted sub-experiments, each of which following its own experimental setup $s_i = (\phi_i, \mu_i)$ for i in some index set $I \neq \varnothing$, into one overarching experimental design $D = \{s_i\}_{i \in I}$, one obtains the full inverse problem

$$\text{find} \quad p \in \mathcal{A} \quad \text{such that} \quad F(p) = F_D(p) = (F_s(p))_{s \in D} = (y_s)_{s \in D} = y. \tag{2.2}$$

By denoting the set of all accessible experimental setups by $\mathcal{D}$, then the set of all possible experimental designs is given by its power set $\mathcal{P}(\mathcal{D})$.

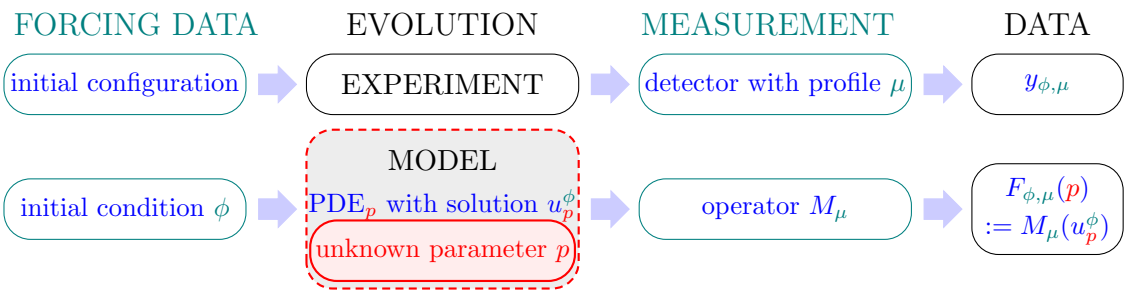


Figure 2.1.: The structure of a PDE parameter identification inverse problem, where one imposes the initial condition ϕ. Unknown quantities are depicted in red, known quantities in blue, and controllable (known) quantities in teal. As only the structure of the PDE, but not the precise value of the parameter p is known, it is depicted in gray. The inverse problem seeks to align these two rows so they produce the same data.

The question how to select experimental setups that generate informative data leads to the research area of (optimal) experimental design (now referring to the designing process). Because this should typically be decided before real world experimentation, the selection is based on the mathematical input-to-output map

$$\mathrm{ItO}_p : \mathcal{D} \to \mathbb{R}, \qquad s \mapsto F_s(p),$$

that can be used to study the influence of variations in the experimental setup $s \in \mathcal{D}$ on the synthetic data $F_s(p)$ under a fixed ground truth parameter value $p \in \mathcal{A}$ and thus mimics the assumption of a fixed data generating parameter value in real world experiment which is necessary for meaningful reconstruction. It characterizes the experimental setting, i.e. the degree of freedom in the input data and its relation to the data - not to be confused with experimental setups.

Noise. In real contexts, data is mostly polluted by noise that stems from detector inaccuracy in the measurement procedure, numerical round-offs or imperfect modeling of the phenomenon by the forward model. This noise is typically modeled as a stochastic process, and a frequently encountered example is the additive centered Gaussian noise model, which, for a finite dimensional data space $\mathcal{Y} \subset \mathbb{R}^L$ reads

$$y = F(p) + \eta \qquad \text{with} \qquad \eta \sim \mathcal{N}(0, \Gamma_{\mathrm{n}}) \tag{2.3}$$

with symmetric positive definite covariance matrix $\Gamma_{\mathrm{n}} \in \mathbb{R}^{L \times L}$.

In this work, we mostly exclude noise from our considerations, in order to the focus is on investigating the informative quality of the collected data w.r.t. the parameter in terms of the data type or the experimental design. We stress, however, that this is an important and much investigated part of inverse problems. Theories to deal with noise such as regularization or Bayesian inversion are for instance presented in [Kir21, Gro93, Stu10, KS06].

Well- and Ill-posedness. Hadamard's concept of well-posedness [Had02] applies to inverse problems as well. Continuity of the solution w.r.t. the data is investigated through so called stability estimates of the form

$$\|p_1 - p_2\|_{\mathcal{A}} \leq \omega(\|F(p_1) - F(p_2)\|_{\mathcal{Y}}) \tag{2.4}$$

that characterize continuity of F^{-1} through a modulus of continuity $\omega : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ with $\omega(0) = 0$.

Definition 2.1 ([Bal12])**.** Inverse problem (2.1) is called (conditionally) well-posed, if the following properties hold:

(i) Identifiability: For every $y \in F(\mathcal{A}) \subset \mathcal{Y}$, there exists exactly one $p \in \mathcal{A}$ for which $y = F(p)$, i.e. F is injective.

(ii) Stability: F satisfies (2.4) and the modulus of continuity ω of the F^{-1} grows at an acceptable rate, typically $\omega(x) \leq Cx$ for some $C > 0$.

If the inverse problem is not well-posed, it is called ill-posed.

The first property states existence and uniqueness of the reconstruction through bijectivity of $F : \mathcal{A} \to F(\mathcal{A})$, the key requirement for inversion. Stability depends on the choice of norms and is subjective, since it depends on the context whether a particular rate is considered acceptable or not. For $\omega(x) \leq Cx$, this formulation of stability, as also used in [FSU19], requires Lipschitz continuity and is thus stronger than the pure requirement of continuity of F^{-1} as used in [Kir21, Gro93]. In case of linear operators as considered in [Kir21], both formulations are equivalent. Stability is particularly important when noisy data are considered, as it bounds the amplification of data noise in the reconstruction. But also identifiability can be affected by noise, in particular when y is driven out of the image of F or coinciding data for two different parameters are generated.

Well-posedness can be understood as a proper balance between a-priori knowledge in the admissible set $\mathcal{A}$ and the information compression on the parameter through the forward map F. In case of ill-posedness, additional information has to be gathered, either by changing the measuring technique, or by introducing more a-priori knowledge. The latter refers to regularizing strategies which comprise higher a-priori smoothness or structural assumptions on the parameter p, penalization methods to promote smallness or sparsity, or introduction of prior information of p in the Bayesian framework. More details can be found in standard literature such as [Kir21, Bal12].

In order to treat inverse problem (2.1) numerically, it is typically rephrased in one of the following two ways. Precise computational methods to solve the inverse problems are the subject of several textbooks, for instance [Vog02].

2.2. Cost Minimization

Motivated by the fact that noise might prevent a perfect fit in (2.1), a relaxed notion of solutions to the inverse problem can be considered, when a variational perspective is taken

and the inverse problem is rephrased as a cost minimization problem [Kir21, Gro93]

$$p_\star = \arg\min_p \mathcal{C}_d(p), \tag{2.5}$$

with cost function $\mathcal{C}_d : \mathcal{A} \to \mathbb{R}_0^+$ that measures the error between the observed data y and the synthetic data $F(p)$ produced by the considered parameter value p. A typical choice is $\mathcal{C}(p) = \|y - F(p)\|_\mathcal{Y}$. Hence, the solution $p_\star$ does not necessarily fit the data perfectly, but in an optimal way as measured by the cost function.

For ill-posed problems, regularization theory suggests to add a penalization term to the cost functional, for instance $\alpha\|p - p_0\|_\mathcal{A}$ for some $\alpha > 0$, that biases the solution towards a reference solution $p_0 \in \mathcal{A}$ and thus restores a unique and stable reconstruction through introduction of the additional a-priori knowledge that p is supposed to be close to p_0 [Kir21, Gro93, TA77].

Given the large variety of well developed optimization algorithms, as summarized for instance in [Ren10, ABT18, KS06], this framework is well suited for numerical reconstruction. The level of difficulty of this optimization task depends on the choice of the cost function, as well as the properties of the forward map, and is therefore very problem dependent.

It is particularly desirable to work with strongly convex cost functions, because these uniquely determine the solution of the inverse problem as their unique minimum, and a successful reconstruction is guaranteed even for even simple optimization methods such as the vanilla gradient descent method described below. Establishing the strong convexity property, however, might be challenging due to non-linearity of the forward map F, and dependence on the chosen experimental design. A lack of convexity calls for techniques from non-convex optimization which typically require more sophisticated methods and bigger computational effort, see for instance [JK+17, HL04], or regularization to restore convexity.

2.2.1. Numerical Optimization by Gradient Descend

Consider the problem of minimizing a C^2 cost function $\mathcal{C} : \mathbb{R}^p \to \mathbb{R}_0^+$ on $\mathbb{R}^d$:

$$\arg\min_{p\in\mathbb{R}^p} \mathcal{C}(p).$$

Gradient descend is a standard iterative minimization scheme that updates an initial guess $p^{(0)} \in \mathbb{R}^p$ according to

$$p^{(n+1)} = p^{(n)} - \eta_n \nabla_p \mathcal{C}(p^{(n)}), \tag{2.6}$$

for a suitable step size $\eta_n \in \mathbb{R}^+$ [WR22, Pol63] . Given that the gradient is the direction of steepest descend in the Euclidean norm, this decreases the value of $\mathcal{C}$, if η_n is chosen sufficiently small, as can be seen from the Taylor expansion

$$\mathcal{C}(p^{(n+1)}) = \mathcal{C}(p^{(n)}) - \eta_n \|\nabla_p \mathcal{C}(p^{(n)})\|^2 + \mathcal{O}(\eta_n^2) \leq \mathcal{C}(p^{(n)}),$$

assuming uniform boundedness of the Hessian $H_p\mathcal{C}(p) \prec C_H$. Because the monotonously decreasing sequence $(\mathcal{C}(p^{(n)}))_n$ is bounded due to $\mathcal{C} \geq 0$, it converges. In general, this does not necessarily guarantee convergence of $(p^{(n)})_n$. Further requirements on $\mathcal{C}$ are necessary.

Definition 2.2. A C^2 function $\mathcal{C} : \mathbb{R}^p \to \mathbb{R}$ is called c_H-strongly convex, if $H_p\mathcal{C}(p) \succeq c_H > 0$ for all p. It is called C_H-smooth for a constant $C_H > 0$, if $C_H \succeq H_p\mathcal{C}(p) \succeq -C_H$ for all p.

Matrix inequalities are understood in the sense of eigenvalues. C_H-smoothness ensures convergence of the sequence $\mathcal{C}(p^{(n)})$ by Taylor, as explained above, while strong convexity implies the existence of a unique minimum point $p_\star$. In combination, both properties provide that the limit is the minimum of $\mathcal{C}$ and that convergence translates to the parameters.

Proposition 2.3 ([Pol63])**.** *Let $\mathcal{C}$ is c_H-strongly convex and C_H-smooth, then the gradient descend scheme (2.6) with $\eta_n \equiv \eta \leq \frac{2}{C_H}$ converges, in the Euclidean norm, to the unique minimum point $p_\star \in \mathbb{R}^p$ of $\mathcal{C}$, with rate*

$$\|p^{(n)} - p_\star\| \leq r^n \|p^{(0)} - p_\star\| \qquad \textit{with} \qquad r := (\max(|1 - \eta c_H|, |1 - \eta C_H|)).$$

Optimal convergence is obtained for $\eta = \frac{2}{C_H + c_H}$.

A proof can be found in [Pol63]. Numerous variants of gradient descend exist that improve convergence or computational cost, for instance line search chooses the optimal η_n that minimizes $\mathcal{C}(p^{(n)} - \eta_n \nabla_p \mathcal{C}(p^{(n)}))$, the heavy ball method includes a momentum term in the update to improve convergence and avoid saddle points of non-convex functions, Newton's method includes Hessian information to obtain a second-order optimization method, and the stochastic gradient descend method decreases computational complexity when $\nabla_p \mathcal{C}(p)$ consists of a sum of many sub-gradients. The interested reader is referred to standard literature such as [WR22, BV04].

2.3. The Bayesian Approach

Another approach to deal with ill-posedness and noise is proposed by the Bayesian framework, where the inverse problem is lifted to a stochastic level. This approach accounts for uncertainty in the parameter reconstruction based on uncertainty in the concrete value of noise. Introductory literature, on which this subsection is based, may be found in [Stu10, KS06].

The Bayesian approach assumes a stochastic noise model. For notational convenience, the method is presented for a the finite dimensional additive Gaussian noise model (2.3) and a separable Banach space $\mathcal{A}$. More general frameworks, for instance with non-Gaussian noise or infinite dimensional data spaces [Stu10, DS17] or multiplicative and mixed noise models [Dun19], are available in literature.

As the model parameter p is also modeled stochastically, a probability distribution $\pi_0 : \mathcal{A} \to \mathbb{R}_0^+$ for p, termed prior distribution, encodes the prior model on the parameter space $\mathcal{A}$. Moreover, p is assumed independent of the data noise η. Then the conditional distribution π^y of $p|y$, i.e. p conditioned on the observed data y, is called posterior distribution and is regarded as the solution to the Bayesian inverse problem. Bayes' theorem explains its form:

Theorem 2.4 (Bayes' theorem; [Stu10])**.** *If $F : \mathcal{A} \to \mathcal{Y}$ is continuous, then $p|y$ is distributed according to a measure π^y that is absolutely continuous w.r.t. π_0 and attains a normalized version of the likelihood $\ell(y|p) = \exp(-\frac{1}{2}(y - F(p))^T \Gamma_n^{-1}(y - F(p)))$ of observing data y given parameter p under the noise model (2.3) as Radon-Nikodym derivative, i.e.*

$$\pi^y(p) = \frac{1}{Z}\ell(y|p)\pi_0(p), \tag{2.7}$$

where $Z := \int_{\mathcal{A}} \ell(y|p)\,\mathrm{d}\pi_0(p)$ is a normalization constant.

This technique can also be applied when data is not very informative about the parameter. In this case prior knowledge in the prior distribution leverages the missing information in the data [Bal12, Stu10], see also Section 2.3.2. By equipping the space of posterior distributions with the Hellinger metric, the concept of well-posedness can be transferred to the Bayesian inverse framework [Stu10] and it has been shown to hold under very mild conditions [Lat23, DS17]. Bayesian inversion can thus be regarded a regularization technique.

On the other hand, the posterior distribution offers richer insight into the structure of the solution to the inverse problem than the deterministic value attained from cost minimization. It allows sampling, identification of multiple modes, and opens opportunities for uncertainty quantification [DS17].

The numerical treatment of Bayesian inversion, however, poses remarkable challenges, as mostly no closed form of the posterior distribution is available. Different strategies have been developed, that are briefly summarized below, following the review in [Stu10]:

- Sampling: Drawing samples from the posterior distribution allows exploration of the full posterior density without imposing additional structural requirements. However, sampling from non standard distributions, as arising for non linear inverse problem or non Gaussian prior or noise distributions, is challenging and requires specific algorithms. More details can be found in the subsequent Section 2.3.1.

- Variational methods refer either to computing the maximum a-posteriori point (MAP) $p_{\mathrm{MAP}} \in \mathcal{A}$, the point that maximizes the posterior probability density function. Information on the posterior is then compressed to only one point value. For Gaussian priors μ_0, this results in maximizing the Lebesgue density of the posterior which can be rephrased as a Tikhonov regularized cost function minimization as in Section 2.2. The same optimization techniques [Ren10, KS06] apply, with low computational cost compared to sampling.

  Alternatively, they also refer to techniques that approximate the posterior distribution by a simpler distribution, for instance a Gaussian in Laplace approximation, which can then be accessed explicitly or sampled by easier methods [Ada03, GRI21].

- Filtering techniques impose additional structure on the unknown parameter p and data y which they exploit when building the posterior distribution incrementally by adding more and more data. The hope is that single update steps can be performed more efficiently, as the distance between an update and a previous state is not as

far as between posterior and prior. These methods naturally find application in online settings, where new data is constantly acquired. In high dimensions, filtering requires high computational effort and frequently particle approximations are used, e.g. in the ensemble Kalman filter [BVLE98].

2.3.1. Bayesian Posterior Sampling

Sampling from non standard distributions as arising in non Gaussian Bayesian inverse problems is challenging. Typically, an a-priori computation of the full posterior distribution $\pi^y(p) = Z^{-1}\ell(y|p)\pi_0(p)$ or even the normalization constant Z is not feasible, as each evaluation of the likelihood requires computing the forward map, which is often expensive for PDE models.

Samplers have been developed on the basis of Markov chains whose distribution converges to the target sampling distribution, for which these difficulties are circumvented through building their dynamics such that it

(P1) is invariant w.r.t. scaling of the sampling distribution, which allows us to avoid computing the normalization constant Z and directly work with $\ell(y|p)\pi_0(p) \propto \pi^y(p)$ - this is often achieved working with quotients or (approximations of) the log density gradient - and

(P2) utilizes a minimal number of expensive sampling distribution evaluations - typically only local values of the (not normalized) sampling distribution at previously sampled or the newly proposed states are required to update the Markov chain.

Several methods have been proposed that fulfill these criteria:

- The most classical methods are Markov Chain Monte Carlo (MCMC) methods that construct a Markov chain, in which sample members are Members are evolved in time according to a random walk, that is biased so to converge to the target distribution. Once this equilibrium is reached, a subset of members of the remaining chain constitutes the sample. Typical methods comprise the Metropolis Hastings algorithm [MRR+53, Has70], Langevin Monte Carlo, Hamiltonian Monte Carlo, Metropolis-Hastings LMC, and so on [DK19, RT96, CCBJ18, CFG14, MS21, BRH13, DCWY19].

  Though very generally applicable, these methods suffer from high computational cost: gradient free methods often exhibit very slow convergence to the equilibrium [CV02], leading to a long burning time that consumes many posterior evaluations, whereas the gradient computation raises the cost of gradient based methods. Moreover, the sequential sample generation leads to autocorrelation, which can be mitigated by thinning, entailing additional computational cost to generate the same size of a sample.

- Recently, the idea to evolve the sample simultaneously as an interacting Markov chain, originating in data assimilation [Rei11, EVVL22], has been transferred to

sampling tasks [GIHLS20, CHSV22, SWZ23, DNS23, HV19]. Particles can communicate and explore the distribution landscape more efficiently. Furthermore, their interaction can be designed to entail further properties such as affine invariance, or gradient free approximations, allowing a very efficient application. As these methods are an active area of research, their non-asymptotic convergence behaviour is not fully understood yet.

Two ensemble based methods will be explained in the following in more detail: the Ensemble Kalman Sampler (EKS) [GIHLS20], and the Consensus Based Sampler (CBS) [CHSV22]. Both methods are very efficient, as they can be used in a gradient-free manner.

Let the admissible $\mathcal{A} = \mathbb{R}^Q$ be finite dimensional and denote the target distribution by $\pi(p)$ and assume absolute continuity w.r.t. the Lebesgue measure. Then its positivity often allows us to rewrite it by means of a potential Φ as

$$\pi(p) \propto e^{-\Phi(p)}. \tag{2.8}$$

EKS [GIHLS20]. The EKS is developed as an ensemble version of the overdamped Langevin dynamics that, in its original form, updates single samples p_n independently according to a noisy gradient flow in Φ, i.e.

$$\mathrm{d}p_n = -\nabla_p \Phi(p_n)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_n, \qquad n = 1, ..., N, \tag{2.9}$$

where the W_n denote independent and identically distributed (i.i.d.) Brownian motions on $\mathbb{R}^Q$. Introducing a preconditioner can speed up convergence to the target distribution (2.8), and by choosing the preconditioner as the empirical particle covariance matrix $\mathrm{Cov}[p] = N^{-1}\sum_n (p_n - \bar{p}) \otimes (p_n - \bar{p})$, where $\bar{p} = N^{-1}\sum_{n'} p_{n'}$ denotes the ensemble mean, then the full ensemble $\{p_n\}_{n=1,...,N}$ is evolved in an interactive manner according to the stochastic differential equation (SDE)

$$\mathrm{d}p_n = -\mathrm{Cov}[p]\nabla_p \Phi(p_n)\,\mathrm{d}t + \sqrt{2\mathrm{Cov}[p]}\,\mathrm{d}W_n, \qquad n = 1, ..., N. \tag{2.10}$$

Assuming a certain square norm structure of Φ, gradient free dynamics can be derived: as a simple extension of [GIHLS20], let $\Phi(p) = h_2(\|h_1(p)\|^2)$ be a C^1 function $h_2 : \mathbb{R} \to \mathbb{R}$ of the square norm of a mildly non-linear C^1 function $h_1 : \mathcal{A} \to \mathbb{R}^k$ for some $k \in \mathbb{N}$. Then the drift term can be approximated by a gradient free difference term, leading to the gradient free implementation of the sampling scheme:

$$\begin{aligned}\mathrm{Cov}[p]\nabla_p \Phi(p_n) &= \frac{2h_2'(\|h_1(p_n)\|^2)}{N}\sum_{n'}(p_{n'} - \bar{p})(p_{n'} - \bar{p})^T (D_p h_1(p_n))^T h_1(p_n) \\ &\approx 2h_2'(\|h_1(p_n)\|^2)\frac{1}{N}\sum_{n'}(p_{n'} - \bar{p})(h_1(p_{n'}) - \bar{h}_1)^T h_1(p_n) \qquad =: \sum_{n'} C_{n,n'} p_{n'}.\end{aligned}$$

Note that the approximation $D_p h_1(p_n)(p_{n'} - \bar{p}) \approx h_1(p_{n'}) - \bar{h}_1$ in the last line, where $\bar{h}_1 = N^{-1}\sum_{n'} h_1(p_{n'})$ denotes the ensemble mean of $\{h_1(p_n)\}_n$, is justified by the fact that h_1 attains an almost constant Jacobi matrix $D_p h_1(p)$. Moreover, the coupling coefficient

between the ensemble members are summarize in $C_{n,n'} := 2h_2'(\|h_1(p_n)\|^2)N^{-1}(h_1(p_{n'}) - \bar{h}_1)^T h_1(p_n)$, and the missing summand involving $\bar{p}$ vanishes due to averaging in $\bar{h}_1$.

Asymptotic Sampling Behaviour. The mean field limit describes the evolution of the particle density ρ, as the many particle limit $N \to \infty$ of the empirical ensemble distribution

$$\rho_{(N)}(t) = \frac{1}{N}\sum_{n=1}^{N} \delta_{p_n(t)}, \tag{2.11}$$

and it gives insight into the behaviour of the particle system. For Lipschitz smooth Φ, the mean field limit for the gradient-based dynamics (2.10) has been derived in [DL21] and [Vae24] as:

$$\partial_t \rho = \nabla_p \cdot (\rho \mathrm{Cov}(\rho)\nabla_p \Phi) + \mathrm{tr}(\mathrm{Cov}(\rho) H_p \rho\,), \tag{2.12}$$

where $\mathrm{Cov}(\rho)$ denotes the macroscopic covariance

$$\mathrm{Cov}(\rho) = \int (q - m(\rho)) \otimes (q - m(\rho)) \rho(q)\,\mathrm{d}q, \quad \text{with macroscopic mean} \quad m(\rho) = \int q\,\rho(q)\,\mathrm{d}q. \tag{2.13}$$

By independence of $\mathrm{Cov}(\rho)$ from p, one easily sees that $\rho \propto e^{-\Phi}$ is a stationary state of (2.12). Exponential convergence to this steady state can be guaranteed under certain circumstances, for instance a well chosen initialization distribution π_{init}, and strong convexity of Φ while excluding ensemble concentration in one point, as shown in [GIHLS20, Prop. 3.1]. This means that for large N and t, the empirical distribution approximates the sampling distribution $\rho_{(N)}(t) \overset{N \text{ large}}{\approx} \rho(t) \overset{t \text{ large}}{\approx} \pi$, and the sample $\{p_n\}_n$ can be considered as a sample from π.

Numerical Algorithm. The numerical algorithm first initializes a sample $\{p_1^0, ..., p_N^0\}$ according to a suitable initial distribution π_{init}. To derive the numerical sample propagation law, the gradient free approximation to the continuous SDE (2.10) is discretized in time. Application of the classical Euler-Maruyama scheme, for instance, yields

$$p_n^{k+1} = p_n^k - \Delta t^k \sum_{n'} C_{n,n'}^k p_{n'}^k + \sqrt{2\Delta t^k \mathrm{Cov}[p^k]}\,\zeta_n^k,$$

with i.i.d. $\zeta_n^k \sim \mathcal{N}(0, I)$. An adaptive time step of $\Delta t^k = \frac{\Delta t^0}{\|C^k\|_F + \varepsilon}$ between the samples p_n^k and p_n^{k+1}, for some $\varepsilon > 0$, was proposed in [GIHLS20, KS19], where the matrix $\mathbb{R}^{N\times N} \ni C^k = (C_{n,n'}^k)_{n,n'}$ summarizes the coupling between all ensemble members. The above consideration of the asymptotic sampling behaviour justifies taking the output sample as $\{p_n^k\}_{n=1}^N$ for sufficiently large k, i.e. large evolution times t^k.

CBS Sampling [CHSV22]. The CBS leverages the Laplace method [SM95] for sampling. It is constructed such that its mean field limit converges in the long time limit to a Gaussian approximation of the target distribution around its maximum.

It evolves a set of N particles $\{p_n\}_{n=1}^N$ according to the gradient free dynamics given by the SDE

$$\mathrm{d}p_n = -(p_n - m_\beta(\rho_{(N)}(t)))\,\mathrm{d}t + \sqrt{2(1+\beta)\mathrm{Cov}_\beta(\rho_{(N)}(t))}\,\mathrm{d}W_n\,, \tag{2.14}$$

where $\rho_{(N)}$ denotes the empirical particle distribution from (2.11), and $m_\beta(\rho) := m(L_\beta\rho)$ and $\mathrm{Cov}_\beta(\rho) := \mathrm{Cov}(L_\beta\rho)$, for m and Cov as in (2.13), denote the mean and covariance of the reweighted distribution

$$L_\beta\rho = \frac{\rho e^{-\beta\Phi}}{\int \rho(q)e^{-\beta\Phi(q)}\,\mathrm{d}q}, \qquad \text{parametrized by some } \beta > 0.$$

The relaxation in the drift term in (2.14) drives particles towards their reweighted mean. On the other hand, the diffusion term on the right-hand side (RHS) of (2.14) introduces a stochastic deviations proportional to the reweighted ensemble covariance, which allows exploration of the distribution landscape.

Asymptotic Sampling Behaviour. As pointed out in [CHSV22], taking the mean field limit $N \to \infty$ of (2.14) shows that the particle density follows

$$\partial_t\rho = \nabla_p \cdot \left[(p - m_\beta(\rho))\,\rho + (1+\beta)\mathrm{Cov}_\beta(\rho)\nabla_p\rho \right].$$

For strongly convex, smooth potentials Φ that are bounded by the reciprocal of a Gaussian, the mean field limit admits a Gaussian steady state, that is close to the Laplace approximation of the target distribution at the MAP for large enough $\beta > 0$, and ρ converges to this steady state exponentially fast as $t \to \infty$, if initialized by a non-degenerate Gaussian.

Approximation at the MAP can be anticipated by taking the limit as $\beta \to \infty$: then $L_\beta\rho$ converges to a Dirac delta that is located at the global minimum of Φ over the support of ρ, i.e. the MAP if it lies in the support. The mystery why the gradient free dynamics (2.14) succeeds to optimize the distribution by finding the MAP has been unveiled in [RKGF23], where the authors linked the process to a gradient free relaxation of the Langevin dynamics (2.9). Moreover, extensions of the CBS to Gaussian mixture models have been propose in [BRW24] by introducing a kernelized form of the mean and covariance.

Numerical Algorithm. The algorithms starts by initializing the sample $\{p_n^0\}_{n=1}^N$ by a non-degenerate Gaussian distribution and choosing a sufficiently large β. A temporal discretization of the continuous time dynamics (2.14) by an exponential integrator, as proposed in [CHSV22], then yields the numerical propagation

$$p_n^{k+1} = e^{-\Delta t}p_n^k + (1 - e^{-\Delta t})m_\beta(\rho_{(N)}^k) + \sqrt{(1 - e^{-2\Delta t})(1+\beta)\mathrm{Cov}_\beta(\rho_{(N)}^k)}\zeta_n^k,$$

with $\zeta_n^k \sim^{i.i.d.} \mathcal{N}(0, I)$, and the output sample is again taken as $\{p_n^k\}_{n=1}^N$ after sufficiently many time steps k.

2.3.2. Connection between Bayesian and Cost Minimization

A link between Bayesian and the constrained cost optimization framework is established through the evaluation of the Bayesian MAP, which, for Gaussian priors and noise, turns out to be a Tikhonov regularized cost function minimization. In this setting, otherwise arbitrary or heuristic choices of the cost function $\mathcal{C}$, as well as the norm and reference

solution in the penalty term $\alpha\|p-p_0\|_{\mathcal{A}}$ are motivated by the noise and the prior distribution, respectively, see [Stu10, Sec.2.2], [Bal12, Sec.11.3.1]. This underpins the regularizing nature of the Bayesian framework.

Another connection is established through the small noise limit, that was studied in [Stu10, HP07, NP08]. As the variance of a centered Gaussian noise vanishes, the Bayesian solution to a uniquely solvable inverse problem converges weakly to a Dirac measure at the true solution [Stu10], i.e. Bayesian inversion returns a point value that coincides with the true solution. For ill-conditioned linear inverse problems with Gaussian prior and noise, uncertainty from the prior information is pertained in some directions in the small noise limit [Stu10], which demonstrates that the prior information compliments the missing information in the data and thus highlights the regularizing effect of Bayesian inversion.

2.4. Identifiability analysis

Being rooted in control theory [BÅ70, GDI85] and applications [Tho19, Koo49, Fis61], identifiability analysis provides a different perspective onto well-posedness of inverse problems. It sheds light onto the question whether given data contains enough information to uniquely identify the model parameter. If the data is of bad quality in the sense that it is not informative enough, checking for sources of non-identifiability allows a more precise diagnosis and the proposal of suitable remedies, for instance changes in the model, the admissible set or the experimental design, [GJML+19, WHR+21] to cure such ill-posedness and lead to better stability in reconstructions. The framework thus takes a holistic perspective onto the inverse problem. As the identification of redundant or non-informative data, the optimal use of information to reduce uncertainty in the parameter reconstruction and the understanding of remaining uncertainty due to modelling limitations provide important insight into cost-effective data collection procedures, it is also referred to as qualitative experimental design [WP96].

In the remainder of this section, a general framework for identifiability analysis is introduced in adaptation of [LG94, WL82, VC99].

Sources of Non-Identifiability. Non-identifiability can be caused by either of the three subsequent sources [GJML+19, DBP+12]:

(S.I) A mismatch in the structure of the model and the type of collected data might prevent parameter identification, even for perfectly chosen data. In this case, the full input-to-output map ItO_p, i.e. the experiment in total, is not informative about p.

(S.II) A poor selection of the particular experimental design D may lead to non-informative data, even if the type of data is suitable. Poor measurement specifications may provoke an information loss from the measurement procedure, whereas poorly selected input data might not trigger all aspects of the parameter, or a too small

number of experiments might not allow to collect all necessary information on the parameter. In this case, the forward map F_D for this design is not informative, calling for experimental design methods.

(S.III) Finally, noise might induce parameter uncertainty such that the reconstruction cannot be trusted anymore. For instance, the most plausible parameter, e.g. the MAP, might not be unique, pointing towards non-identifiability.

The following example demonstrates the three sources for an easy toy model.

Example 2.5. Consider the determination of the constant speed $c > 0$ of a moving object in the two-dimensional plane $\mathbb{R}^2$. The object is known to start at the origin $(0, 0)$ at time $t = 0$ and move in direction $(1, 0)$. Imagine for instance a bicycle driving on a straight street. The location of the object at time t is $(l(t), 0)$, with first coordinate $l(t) = ct$. An observer, e.g. a camera, is located at a position $x \in \mathbb{R}^2 \backslash \{(\lambda, 0)\}_{\lambda \in \mathbb{R}_0^+}$ and a direction of attention $v \in \mathbb{S}^1$, in order not to interfere with the potential trajectories of the object. The observer has a 180° field of vision and collects data on the location of the moving object, if it is in its field of vision. This data shall be used to recover the speed of the moving object.

(S.I) If the observer only collects the locations of the object without noting the time at which the object reaches this location, reconstruction of its speed will not be possible, as all locations $(\mathbb{R}_0^+, 0)$ will eventually be reached by the object as long as its speed is positive and the observation time is not limited. This type of data is not informative about the speed, no matter which observation position x and direction of attention v are chosen.

(S.II) Assume instead, that the observer collects data on the location $l(T)$ of the moving object for a fixed $T > 0$. This is only possible if $l(T)$ is in its field of vision $\{y \in \mathbb{R}^2 \mid \langle y - x, v\rangle \geq 0\}$. For a suitable choice of x and v, for instance $x = (1, -1)$ and $v = (0, 1)$, the observer can perfectly track $l(T) = cT$ and the speed $c = l(T)/T$ can uniquely be recovered. If x and v are chosen poorly, however, e.g. $x = (1, -1)$ and $\hat{v} = (0, -1)$, then the final location cannot be observed as it lies outside the observer's field of vision and a reconstruction of the speed is not possible. Because x and v refer to properties of the observer, they represent measurement specifications.

(S.III) Assume that the observer is located at the well suited position $x = (1, -1)$ and $v = (0, 1)$, but it has a blurred focus and can only determine the final location within an interval of length 1, i.e. $l(T) \in [a, a + 1]$ for some $a \in \mathbb{R}$. Then the correct speed of the object is not unique. It is only possible to narrow it down to the interval $c \in [a/T, (a + 1)/T]$. In some cases, this estimation will be too coarse to be useful in practice. In this case, even though reconstruction with the clean data would work, observation noise prevents a meaningful reconstruction.

◇

2.4.1. Structural identifiability

Structural identifiability aims to exclude non-identifiability of source (I) and was introduced in [BÅ70]. Different definitions exist [ACDVM20, CDr80], and the following will be used throughout this work:

Definition 2.6 (Structural identifiability, persistent excitation; adapted from [SAD03, LG94])**.** The inverse problem corresponding (2.2) is called (globally) structurally identifiable, if for any ground truth parameter $p_\star \in \mathcal{A}$ there exists an experimental design $D_\star = D_\star(p_\star) \in \mathcal{P}(\mathcal{D})$, such that $F_{D_\star}(p) = F_{D_\star}(p_\star)$ implies $p = p_\star$ for all $p \in \mathcal{A}$. In this case, $D_\star$ is called persistently exciting w.r.t. $F, \mathcal{A}$ and $p_\star$.

Local versions of structural identifiability exist, where uniqueness of $p_\star$ is required to hold only in a neighbourhood $U_{p_\star} \subset \mathcal{A}$ [ACDVM20]. Moreover, it is possible to include non-observable and non-controllable parameters that shall not be reconstructed[LG94, SAD03].

Structural identifiability is weaker than identifiability according to Definition 2.1i, since the experimental design $D_\star$ is not fixed, but can be chosen in dependence on the true parameter $p_\star$, and coinciding values $F_{D_\star}(\hat{p}) = F_{D_\star}(\bar{p})$ are allowed for $\hat{p} \neq \bar{p}$ as long as they are distinct from $p_\star$. Structural identifiability should be examined before experiments are conducted to make sure that the correct type of data is collected and the experiment is informative. In the proofs, a noise-free setting is considered to exclude non-identifiability effects of source (III), and access to the full input-to-output map $\mathrm{ItO}_{p_\star}$ is assumed [WL82, LG94]. The method of investigation of structural identifiability heavily relies on the model and data structure. A brief overview over the rich methodology developed for structural ODE identifiability analysis can be found in the literature review at the end of this subsection.

Persistent excitation [WP96, LG94] refers to the fact that the chosen experimental design $D_\star$ is suitable to discriminate between the ground truth $p_\star$ and other parameter values in the measurement and thus anticipates identifiability of type (II). Given that the true parameter is typically unknown a-priori, in the most favourable case, an experimental design $D_\star$ can be found that is persistently exciting for all $p_\star \in \mathcal{A}$. Unfortunately, this is mostly not possible.

2.4.2. Cost Function and Sensitivity Based Identifiability

To study the persistent excitation property of an experimental design $D_\star$, and thus the existence or lack of sources (I) and (II), identifiability can be combined with the cost minimization reconstruction framework in Section 2.2 [BÅ70, GJML+19, RKM+09], or studied in a linearized setting, for which source (III) is excluded by consideration of noise-free data.

Cost Function Identifiability.

Definition 2.7 (cost function identifiability; adapted from [BÅ70, CDr80])**.** Consider a fixed ground truth parameter $p_\star \in \mathcal{A}$ and a prescribed experimental design $D_\star \in$

$\mathcal{P}(\mathcal{D})$. The inverse problem in the cost minimization framework (2.5) is called (locally) ℓ-identifiable w.r.t. a loss function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_0^+$ at $p_\star, D_\star$, if the cost $\mathcal{C}(p) := \ell(y_\star, F_{D_\star}(p))$ associated to noise free synthetic data $y_\star = F_{D_\star}(p_\star)$ has an isolated local minimum at $p = p_\star$. If $p_\star$ is the unique global minimum, then the inverse problem is globally ℓ-identifiable w.r.t. $p_\star, D_\star$.

For a suitable class of loss functions, such as the standard $\ell(y_1, y_2) = \|y_1 - y_2\|_{\mathcal{Y}}$ for some norm on $\mathcal{Y}$, (local/global) cost function identifiability at a point $p_\star, D_\star$ is equivalent to the fact that $D_\star$ is (locally/globally) persistently exciting w.r.t. $p_\star$.

Remark 2.8. The Bayesian setting by definition does not generate uniqueness of the parameter. Investigating identifiability from sources (I) and (II) is still beneficial to learn whether uncertainty in the parameter arises only from uncertain noise or also from other sources. Statistical approaches to identifiability can be applied, for instance identifiability can be defined through comparison of the shape of the full (log)-likelihoods almost everywhere (a.e.w.) in the parameter space as in [LHL10] or uniqueness of the maximum of the log-likelihood [RL19]. This is connected to cost function identifiability with the log likelihood as a cost function, in analogy to the considerations in Section 2.3.2.

Sensitivity analysis. Sensitivity analysis is another tool frequently used to investigate identifiability issues of types (I) and (II) [MXPW11, LDM22, GVB17]. The sensitivity expresses how much the output measurement changes if the parameter is varied. For notational convenience, a finite dimensional parameter $p \in \mathcal{A} \subset \mathbb{R}^Q$ and a finite design

$$D = \{s_1, ..., s_L\} \in \mathcal{P}(\mathcal{D})$$

are considered in the following, while pointing out that extensions to infinite designs are studied in Chapter 8. Assuming sufficient regularity of the forward map, the sensitivity matrix for the finite experimental design D refers to the Jacobi matrix $J_p F_D(p_\star)$ of the measurement w.r.t. the parameter values, evaluated at a prescribed ground truth parameter $p_\star \in \mathcal{A}$,

$$J = J_p F_D(p_\star) \quad \in \mathbb{R}^{L \times Q}, \quad \text{with rows} \quad J_{l:} = (\nabla_K M_l(f_{K_\star}))^T, \qquad l = 1, ..., L, \quad (2.15)$$

and represents the linearized inverse problem with linearization at $p_\star$

$$0 = J(p - p_\star).$$

A lack of sensitivity of all measurements w.r.t. one parameter - i.e. a vanishing column of J, a high correlation between a pair of columns, collinearity of a group of columns or rank deficiency of the Hessian $J^T J$, all allow for non-vanishing solutions to the linearized inverse problem and therefore indicate non-identifiability of the considered (pair or group of) parameter(s) or the full set of parameters; an overview over several sensitivity based methods may be found in [MXPW11]. In this work, sensitivity based identifiability will be evaluated by the eigenvalue criterion in terms of rank deficiency of $J^T J$.

Definition 2.9 ([MXPW11])**.** The inverse problem (2.2) under the experimental design D is called sensitivity based identifiable at $p_\star$, if $J^T J$ is positive definite.

Remark 2.10. Sensitivity based and cost function identifiability are qualitative criteria. They do not account for poor conditioning of J^TJ or equivalently a flat cost landscape of $\mathcal{C}$, both of with leading to major difficulties in numerical reconstruction through high computational complexity fo finding the cost minimum, even for noise free data. In these cases, noise, by slightly distorting the data, can easily render the inverse problem non-identifiability. To avoid this, the sensitivity based identifiability criterion can for instance be exchanged by a threshold for the minimum eigenvalue or conditioning of J^TJ that is not be exceeded.

Connection between Cost Function and Sensitivity Based Identifiability. The connection between sensitivity and identifiability is given by the quadratic cost function for this design

$$\mathcal{C}(p) = \mathcal{C}_D(p) = \frac{1}{2L}\|y_\star - F_D(p)\|_2^2 = \frac{1}{2L}\sum_{s\in D}(y_s - F_s(p))^2$$

with noise free data $y_\star = F_D(p_\star)$ - this exclude source (III). Linearization of the forward map F_D at $p_\star$ provides $\mathcal{C}_{\text{lin}}(p) = L^{-1}(p-p_\star)J^TJ(p-p_\star)$, i.e. the Hessian of the linearized cost function reads $H_p\mathcal{C}_{\text{lin}}(p) = L^{-1}J^TJ$. A full rank structure of J^TJ reveals that $\mathcal{C}_{\text{lin}}(p)$ is strongly convex and has a unique global minimum at $p_\star$, whereas in case of singularity of J^TJ, the linearized inverse problem contains a non-identifiable parameter [WP01].

By the implicit linearization, sensitivity analysis can in general only provide local guidance on identifiability of the non-linear inverse problem. Under sufficient regularity of the forward map, however, identifiability results are locally transferable:

Proposition 2.11 (adapted from [HKLT25])**.** *Let a finite design $D \in \mathcal{P}(\mathcal{D})$ be given and assume that the forward map $F_s(p)$ is twice continuously differentiable in p and its Hessian $H_pF_s(p) \in \mathbb{R}^{Q\times Q}$ is Lipschitz continuous w.r.t. a norm $\|\cdot\|_{\mathcal{A}}$ on the admissible set and the Frobenius norm on $\mathbb{R}^{Q\times Q}$, in a neighbourhood around the ground truth parameter $p_\star$, for all $s \in D$. Then sensitivity based identifiability according to Definition 2.9 is equivalent to local strong convexity of the quadratic cost $\mathcal{C}$, and implies local square loss identifiability.*

The proposition connects sensitivity based identifiability with local strong convexity of the quadratic cost function, a very favourable property that facilitates the numerical optimization of the cost function. Note that a global characterization of cost function convexity is in general hard for non-linear forward maps, because the Hessian depends on the considered parameter value.

The regularity assumption on F is rather mild and holds true in many applications, in particular in many PDE parameter identification problems, where well posedness of the forward model typically comprises a certain regularity of the solution w.r.t. the parameters and initial data.

Remark 2.12. Proposition 2.11 can be extended to infinite experimental designs, as studied in Chapter 8.

To proof the proposition, recall that local strong convexity is equivalent to positive

definiteness of the Hessian of the quadratic cost function

$$\mathbb{R}^{Q\times Q} \quad \ni \quad H_p\mathcal{C}(p) \quad = \quad L^{-1}\sum_{s\in D}\left(\nabla_p F_s(p)\otimes\nabla_p F_s(p) + (y_s - F_s(p))H_pF_s(p)\right) \tag{2.16}$$

for all p in the respective domain, according to Definition 2.2. Noting that the Hessian of the quadratic cost $\mathcal{C}$ evaluated at the ground truth parameter $p_\star$ reads

$$H_p\mathcal{C}(p_\star) = L^{-1}\sum_{s\in D}\nabla_p F_s(p_\star)\otimes\nabla_p F_s(p_\star) = L^{-1}J^TJ, \tag{2.17}$$

the proof is established by extending its positive definiteness to a neighbourhood of $p_\star$ through Lipschitz continuity of the Hessian of the cost function, established through Lipschitz continuity of all terms appearing in (2.16) by assumption.

Proof. By the above characterization of local strong convexity in terms of positivity of the Hessian eigenvalues, it is clear that it implies sensitivity based identifiability. Vice versa, matrix perturbation theory [HJ85, Corr.6.3.8] establishes continuity of the minimal eigenvalue w.r.t. perturbation in the matrix. Moreover, by representation (2.16), local Lipschitz continuity of the Hessian forward map H_pF_s in p translates to the quadratic cost Hessian $H_p\mathcal{C}$, and, therefore, one has for all p in a neighbourhood U of $p_\star$ that

$$|\lambda_{\min}(H_p\mathcal{C}(p_\star)) - \lambda_{\min}(H_p\mathcal{C}(p))| \leq \|H_p\mathcal{C}(p_\star) - H_p\mathcal{C}(p)\|_F \leq C\|p_\star - p\|_{\mathcal{A}}, \tag{2.18}$$

for a constant $C > 0$. Hence, if $\lambda_{\min}(H_p\mathcal{C}(p_\star)) = \lambda_{\min}(J^TJ) > 0$, then one can find a $\|\cdot\|_{\mathcal{A}}$-neighbourhood $U_{p_\star} \subset U$ of $p_\star$ for which $\lambda_{\min}(H_p\mathcal{C}(p)) > 0$ for all $p \in U_{p_\star}$, i.e. in which $\mathcal{C}$ is strongly convex. □

Minimal Number of Data. As a simple consequence of sensitivity analysis, a minimum size of an experimental design can be derived that is capable of rendering the inverse problem sensitivity based identifiable. Intuitively, it makes sense, that this number has to be related to the dimension of the parameter p: if only few data is collected, it is unlikely that this data is sufficient to reconstruct a high dimensional parameter, and the inverse problem is typically underdetermined. The following proposition suggests collecting at least as many data points as the dimension of the parameter.

Proposition 2.13. *Every (finite) experimental design D for which the corresponding inverse problem (2.2) is sensitivity based identifiable consists of at least Q experimental setups s_l. If $|D| > Q$, then a reduced design with Q experimental setups exists for which the corresponding inverse problem is also sensitivity based identifiable.*

This is not surprising, given that sensitivity analysis studies identifiability of the linearized inverse problem, i.e. linear system of equations with Q unknowns. A combination with Proposition 2.11, however, allows drawing conclusions for the nonlinear inverse problem and local strong convexity of its quadratic cost function.

Proof. If $|D| = L < Q$, then the Hessian $H_p\mathcal{C}(p_\star)$, as a sum (2.17) of L tensor products summands of rank 1, attains a rank of at most L and is thus rank deficient. On the other

hand, if a design provides more data than the dimension of $p \in \mathcal{A} \subset \mathbb{R}^Q$, i.e. $|D| > Q$, and leads to a sensitivity based identifiable inverse problem, then one can select a subset of these experimental setups for which identifiability still holds: Because $J^T J \in \mathbb{R}^{Q \times Q}$ is of full rank Q by assumption, so is $J \in \mathbb{R}^{L \times Q}$ and one can select Q experimental setups s_l for which the corresponding rows of J are linearly independent, and thus the corresponding Hessian at ground truth is still of full rank. □

Again, the Lemma can be generalized to infinite experimental designs, under the setting considered in Chapter 8.

2.4.3. Practical identifiability

Even if an experiment can theoretically provide sufficient information, i.e. the inverse problem is structurally identifiable, parameter identification can fail in practice due to bad quality of experimental data. One speaks of practical non-identifiability, when the variance of parameter estimates corresponding to noisy data is 'too high' compared to the data noise level [LDM22]. Because the notion of a 'too large variance' is very problem and method dependent, no unified definition of practical identifiability exists throughout literature [LDM22]. Practical identifiability is conceptually related to the stability requirement 2.1(ii).

It might be a consequence of structural non-identifiability (I) or a poor choice of experimental specification (II), both resulting in a lack of information also in the noise-free data, or distortion of the data by noise (III). It manifests in (near-) flatness of the cost function, the likelihood or the Bayesian posterior density landscape. Detection techniques established in ODE models comprise Monte Carlo simulation of noisy data in order to evaluate the spread of the corresponding parameter reconstruction, the consideration of confidence regions based on the profile likelihood function [WHR+21, SM23] or the Hessian matrix of least squares [RKM+11, RKM+09] as well as Bayesian methods [SM23]. A review over these and some more methods can be found in [MXPW11, LDM22].

2.4.4. Dealing with Non-Identifiability.

The way to deal with non-identifiability depends on the purpose of the reconstruction [WHR+21, RKM+11]: if the model is supposed to be used for prediction only in the specific setting of the experiment and predictions are not affected by the concrete choice of an optimal parameter, then non-identifiability can be tolerated. If, however, the parameter itself is the quantity of interest, identifiability of this parameter should be improved. The following methods are proposed in literature [WHR+21, RKM+11]:

- Non-identifiability of source (I) can only be remedied by a change of model structure, i.e. either another type of data is observed or the complexity of the parameter is reduced, for instance by fixation of non-identifiable parts, prescription of a parametric form of the parameter, reparametrization through parameter pooling or introduction

of a-priori knowledge.

- A poor choice of the experimental specification (II) calls for improvement through experimental design.
- Data noise (III) might be reduced by improved measurement quality. Otherwise, prior knowledge on the parameter could be incorporated through regularization techniques [GJML+19].

2.4.5. Literature Survey

Historical development. Identifiability issues in applications such as econometrics [Koo49] or psychology [Tho19] shaded light on this problem early on. The development of a systematic framework for identifiability analysis in a statistical setting started in the 1950s, in [Koo49, KR50] and literature therein, with a focus on linear systems and rank conditions. This was picked up well by the econometrics community [Fis61] and further criteria such as the information matrix criterion were developed [Rot71].

The focus shifted when control theory and systems identification for dynamical ODE systems provided another access to identifiability analysis [BÅ70, GDI85]. This ansatz was widely adopted in subsequent literature and a rich methodology was developed as several review papers suggest [GJML+19, WP96, MXPW11, WHR+21, RKM+11, CBBC11b, LDM22].

Applications. Applications prompted the development of different notions [ACDVM20] and techniques for identifiability analysis, recurring to different requirements in the respective fields such as disease modeling [MXPW11], water quality modeling [Bec87] or environmental studies [GJML+19, BRK01], for instance. Another large area of application is biology, where a limited insight into systems provided by restricted measurement capabilities often faces complex models with numerous potential influence factors. The question of identifiability was studied particularly well in systems biology, for instance for enzyme kinetics [RKM+11, CBBC11b], signaling pathways [RKM+09, GVB17], metabolic models [GVB17, CBBC11b, SCM19], circadian clock related genetic networks [GVB17, CBBC11b], pharmacokinetic models [CBBC11b, RL19], reaction networks [CBBC11b], bacterial growth models [RL19] or oncology models [ST18], just to mention some.

Structural ODE identifiability. A variety of methods has been developed to test for structural identifiability for ODE models. Typical techniques involve

- analytical [DVJB00] or numerical [WBJK04] direct test,
- Laplace transform [RKM+11] for linear models as proposed by [BÅ70],
- generating series expansion (and identifiability tableaus)[WL82, CBBC11a],
- Taylor series expansion [Poh78, WP96],
- similarity transform and local state isomorphism theorem [WL82, DVJB00, VRWL89],
- differential algebra manipulations [LG94, BSAD07, MRCW01],

- implicit function theorem [XM03, WZMP08].

Several software implementations based on a subset of these techniques exist. An overview comprising a critical comparison is available under [RBV23].

PDE identifiability analysis. Despite the large number of PDE inverse problem applications, the focus of most existing literature lies in the sole verification of falsification of the identifiability property in Definition 2.1(i) and the development of regularization strategies to compensate non-identifiability or instability. Not many works consider the inverse problem from an identifiability analysis point of view in order to obtain a more detailed insight into the sources of (non-)identifiability. A brief overview over existing literature and methods is given in the following.

The first consideration of PDE identifiability analysis dates back to the late 1970s [KN77] and instigated a line of structural identifiability research that has emerged in an academic context [TW82, TW85, KS86, Kra88, ZK23, Nak93, Nak97, Les00, LEI00, Suz83, GH07, Giu91, HM97, Lah87, LC89, CKK97, Kob80, HG13, OB00, CG08, FGBC22, CR80, KW86, DuC13]. As applications prompted the further development of ODE identifiability analysis concept and techniques, the topic reemerged in the 2000s in the context of PDE applications, most of which can be found in biology [MMS24] such as epidemiology [RKE22, PLCT11, PR16, TGCM16, GQMT22, ZVDVK18, VMZDV20] or the modelling of cell invasion [SBVM20, LSM+24], cell cycle [BTFB24, FCCB24, SMM24], cell motion [HHT13], morphogenesis [CRY19], protein dynamics [CDM+24] or blood-tissue exchange [VC99]. Further applications were found in information propagation in social networks [KA21b], food processing [RFBCEB07] or environmental sciences [VDVJB06].

PDE practical identifiability. Given that practical and cost function identifiability does not directly rely on the underlying model, established methods readily extend to the PDE context. As such, methods based on sensitivity analysis [KA21b, CDM+24, GQMT22, ZK23, RKE22, DuC13] as well as Monte Carlo methods [TGCM16], Bayesian MCMC [SBVM20, CDM+24, FCCB24] or the profile likelihoods [RKE22, SBVM20, CDM+24, LSM+24, SMM24, HHT13, MMS24] were successfully applied to different types of PDEs that emerge in applications, as Table 2.1 illustrates.

Another approach incorporates the PDE into the cost functional in the cost minimization framework 2.2, which then entails specific PDE based methods for investigation. An example of this idea is given in [LL23] for the interaction potential reconstruction in the McKean-Vlasov equation and, more generally, for PDEs [GLL24] with linear parameter dependence with gradient flows as an example.

PDE structural identifiability. Structural identifiability heavily relies on the model and the data structure. It is thus not obvious which previously mentioned techniques for structural ODE identifiability analysis generalizes to which type of PDE model [RKE22]. Up to this point, no unified framework is available [LSM+24, BHO+24] and the development of general PDE structural identifiability techniques is a subject of current research endeavors [BHO+24]. Literature so far concentrated on the following approaches:

- *Identifiability under approximation.* The so-called identifiability analysis under approximation studies identifiability of discretized versions of the parameter emerging from spatial discretization of the PDE by different methods such as orthogonal

	Sensitivity	Monte Carlo	Bayesian MCMC	Profile likelihood
parabolic PDE	1D [KA21b], multiD [DuC13]			1D [SMM24] 2D [LSM+24, HHT13]
reaction diffusion system	[CDM+24]		[SBVM20, CDM+24, FCCB24]	[SBVM20, CDM+24]
transport PDE system	[GQMT22, RKE22]	[TGCM16]		[RKE22]
1D quasi-hyperbolic PDE	[ZK23]			

Table 2.1.: Practical identifiability techniques applied to PDE models.

polynomials, spectral expansions, finite element (FEM) discretization or finite difference (FD) schemes. The emerging system of ODEs can be investigated by established methods for ODE identifiability analysis such as the differential algebra approach [Lah87, LC89, VDVJB06], the Laplace transform [KW86], direct methods [CKK97, CR80, FGBC22] or the Taylor series expansion [RFBCEB07]. This method has been applied to inverse problems for one- [Lah87, LC89, CKK97, CR80, KW86, VDVJB06] or multi-dimensional [FGBC22] linear parabolic PDE models as well as a one-dimensional nonlinear second order PDE model in [RFBCEB07]. Table 2.2 summarizes which techniques have been applied to which PDE type in literature.

	differential algebra	Laplace transform	direct method	Taylor expansion
1D heat equation		FEM [CR80, KW86]	spectral expansion [CKK97], FEM [CR80]	
1D linear parabolic PDE	orthogonal polynomials [Lah87, LC89], FD [VDVJB06]			
multiD linear parabolic PDE			FEM [FGBC22]	
1D non-linear 2nd order PDE				FD [RFBCEB07]

Table 2.2.: Structural identifiability under approximation: techniques, types of PDE models and spatial discretization.

- *Differential algebra approach.* As the differential algebra approach provides a systematic framework for ODE identifiability analysis studies, where it even allows for software implementation [BSAD07], there has been considerable attempt to apply it to PDE models. Several types of PDEs have been investigated in this framework, for instance models in one space dimension for multi-species linear [CRY19, PLCT11] and semi-linear transport reaction systems [RKE22] emerging from age structured models, the heat equation [Giu91], a linear parabolic PDE [KN77], linear reaction diffusion systems such as the semi-linear autochemotactic Keller-Segel system [CDM$^+$24, BTFB24], and a diffusion equation with logistic growth [BTFB24]. Identifiability of one-dimensional quasi-linear parabolic PDEs and reaction diffusion systems as well as a linear elliptic equation and the wave equation was studied by this method in [BHO$^+$24], as it was for the Euler-Bernoulli beam equation in [Les00]. In multiple space dimensions, specific non-linear reaction diffusion systems in epidemiology were studied in [ZVDVK18, VMZDV20]. A first attempt to generalize the technique to one-dimensional problems of quasi-linear elliptic, parabolic and hyperbolic type PDE models was undertaken in [HM97], before general higher order one-dimensional quasi-linear equations of a certain form were considered in [LEI00]. The authors of the preprint [BHO$^+$24] then extended it to generic non-linear spatio-temporal evolution systems of very general form. It should be mentioned, that the size of the algebraic system grows cubically in the number of state variables [BTFB24] and linearly in the number of unknown parameters, making this approach often computationally infeasible for large systems. Furthermore, the method requires derivatives of the data which might not be accessible, especially for point wise data.

- *Model specific methods.* Other proofs in the realm of PDE identifiability rely on model specific methods to transform the identifiability problem in a more tractable shape. These comprise explicit solution formulas for linear transport equation models [PR16, VC99] or a Laplace transformation to study solutions to linear two-species advection-reaction system [VC99], both in one space dimension. Identifiability for a quasi-linear heat equation was studied through construction of a suitable test function in [Kra88]. This is conceptually similar to the Fredholm integral based adjoint method applied to a n-dimensional parabolic PDE in [DuC13], which is also applied to a coupled parabolic system of chemotaxis in a moving fluid environment in combination with a variation method in [LL24]. A projection method was introduced for general linear PDEs and applied to a one dimensional heat equation in [TW85] as well as linear parabolic PDEs in one space dimension in [TW82, TW85], for which transformation into Fourier and Laplace domain proved useful. Both transformations were combined with a spectral method to study an identifiability problem for the Klein-Gordon equation, a linear hyperbolic PDE, in one space dimension in [HG13]. Spectral expansions also proved useful for linear parabolic PDE models [CPA81, KS86, Suz83, GH07, Nak97, KN77] or the wave equation [CG08], both in one space dimension, as well as a multi-dimensional heat equation [Nak97]. The same technique was applied to a generic multi-dimensional linear hyperbolic equation as well as the Euler-Bernoulli beam equation, a linear fourth order equation, in one space

dimension [Kob80]. A review over early Japanese works in this line of research can be found in [Nak93]. Table 2.3 summarizes the techniques that were used in literature to prove structural identifiability for different types of PDEs.

	explicit solution	testing / adjoint	projection method	Fourier / Laplace domain	spectral expansion
1D linear transport	[PR16, VC99]				
1D advection reaction system				[VC99]	
1D linear parabolic PDE			[TW82, TW85]	[TW82, TW85]	[CPA81, KS86, Suz83, GH07, Nak97, KN77]
1D quasi-linear heat equation		[Kra88]			
multiD heat equation					[Nak97]
multiD parabolic equation / system		[DuC13] / [LL24] (+ variation method)			
1D linear hyperbolic equation				Klein-Gordon eq. [HG13]	wave eq. [CG08], Klein-Gordon eq. [HG13]
1D Euler-Bernoulli beam equation					[Kob80]
multiD linear hyperbolic PDE					[Kob80]
multiD linear PDE			[TW85]		

Table 2.3.: Further methods used to prove structural PDE identifiability analysis.

- Another criterion was proposed in [RS10], where identifiability is connected to group theoretic approaches to establish local injectivity of the parameter to output map.

The concrete choice of method also depends heavily on available measurement information, see e.g. [KN77] where a differential algebra approach was used for distributed measurements, available over the full space domain, and a spectral method was used for point wise measurements. Finally, it should be noted that the application of identifiability

methods to other types of models, such as SDEs is an evolving field [LLM+21, PHF22, BWB+20].

2.5. Optimal Experimental Design

The limitations in experimental resources that practitioners face in real world inverse problems require carefully selecting a finite experimental design to be conducted in reality. As suggested by source (II) of non-identifiability, its choice can have a dramatic effect on identifiability and stability of an inverse problem, and a poor choice might entail collection of redundant or sub-optimal data and inaccurate parameter reconstructions. The mathematical area of optimal experimental design is thus devoted to the development of techniques that find experimental designs whose data is most expressive w.r.t. the parameter, leading to well behaved inverse problems. In the following, we give an introductory overview into this area on the basis of the review articles [Ale21, HJM24]. For a deeper dive, readers are referred to these or the many other available review articles, e.g. [Puk06, Das96, CV95, RDMP16], and references therein.

2.5.1. The Basic Idea

In noisy data settings, the stochastic behaviour of the data provokes uncertainty in the parameter estimate $\hat{p}$, as observed for instance in the Bayesian framework. An experimental design $D \in \mathcal{P}(\mathcal{D})$ is thus regarded as optimal, if it minimizes uncertainty of the parameter estimate $\hat{p}$ associated to the inverse problem with this design D. Different criteria $\phi : \mathcal{P}(\mathcal{D}) \to \mathbb{R}_0^+$, some of which are introduced in the subsequent paragraph, have been proposed to measure this uncertainty, mostly based on notions of variance of the estimate $\hat{p}$. The experimental design optimization problem then reads

$$\arg \min_{D \in \mathcal{P}(\mathcal{D})} \phi(D).$$

Note that the notion of a design here is rather general: the space of experimental setups $\mathcal{D}$ can be discrete or continuous, or, more generally, it can even be infinite dimensional and attain the form of a functional space. A particular example is the space of measures over a domain $X \subset \mathbb{R}^n$ that might describe a measure valued sensor placement [Wal19, NPVW19]. We present the following design criteria for finite designs D, yielding a finite dimensional data space $y \in \mathcal{Y} \subset \mathbb{R}^L$, for sake of simplicity, keeping in mind that extensions to other settings are readily available. Moreover, we assume that the parameter p is finite dimensional $p \in \mathcal{A} \subset \mathbb{R}^Q$ for simplicity of the presentation, even though the theory is clearly not limited to this case and many criteria can readily be extended to the infinite dimensional case under mild adaptations [Ale21].

2.5.2. Optimality Criteria.

Optimality criteria are often equivalently defined for both, the non-Bayesian and Bayesian formulation [HJM24].

i) In the non-Bayesian regime, they mostly refer to an approximation of the covariance matrix of the maximum likelihood estimate $\hat{p}$ in the non Bayesian setting, that is given through the inverse Fisher information matrix $\mathrm{FIM}^{-1}(d, D)$ which - under sufficient regularity - reads

$$\mathrm{FIM}(p, D) = \mathbb{E}_{y|p,D}[\nabla_p \log \ell(y|p, D) \otimes \nabla_p \log \ell(y|p, D)] \quad \in \mathbb{R}^{Q\times Q},$$

where $\ell(y|p, D)$ denotes the likelihood of observing the data y under the parameter value p and experimental design D, and the expectation is taken over this conditional distribution.

ii) In the Bayesian framework, criteria utilize the Bayesian posterior covariance matrix. In a slight abuse of notation, we will also denote the Bayesian parameter estimate, i.e. the random variable distributed according to the posterior distribution π^y, by $\hat{p}$.

Linear Gaussian Regime. Under a linear forward map $F_D(p) = F_D \cdot p$ and an additive Gaussian noise model as in (2.3), as well as a Gaussian prior distribution $\pi_0 = \mathcal{N}(m_{\text{prior}}, \Gamma_{\text{prior}})$ for the Bayesian formulation, then the parameter estimate $\hat{p}$ is also Gaussian. Its covariance matrix $\mathrm{Cov}_D[p]$ attains an explicit form that is independent of the ground truth parameter, and given by

i) either the inverse Fisher information matrix $\mathrm{FIM}^{-1}(p, D) = (F_D^T \Gamma_{\text{n}}^{-1} F_D)^{-1}$ in the non-Bayesian setting,

ii) or the Bayesian posterior covariance matrix $\Gamma_{\text{post}} = (F_D^T \Gamma_{\text{n}}^{-1} F_D + \Gamma_{\text{prior}})^{-1}$.

It can thus be used as a measures of uncertainty for the parameter estimate $\hat{p}$ independently of an unknown ground truth $p_\star$, and many optimality criteria only differ in choice of 'scalarization' of this covariance [Kie59, Mit00, BSE+09, PKG+18, AC22]:

- A-optimality minimizes the average variance over all parameter entries, which translates to the trace $\phi_A(D) = \mathrm{tr}(\mathrm{Cov}_D[p])$.
- D-optimality minimizes the determinant $\phi_D(D) = \det(\mathrm{Cov}_D[p])$ which attains an information theoretic interpretation as the expected information gain from the prior distribution to the posterior distribution in the Bayesian setting. Together with A-optimality, these are the most common criteria.
- E-optimality seeks to minimize the maximum eigenvalue $\lambda_{\max}$ of the covariance matrix $\phi_E(D) = \lambda_{\max}(\mathrm{Cov}_D[p])$.
- K-optimality minimizes the condition number $\phi_K(D) = c(\mathrm{Cov}_D[p]) = \frac{\lambda_{\max}(\mathrm{Cov}_D[p])}{\lambda_{\min}(\mathrm{Cov}_D[p])}$.
- c-optimality is a so-called 'goal oriented' design criterion, that seeks to minimize the covariance of a linear combination $c^T p$ of the parameter entries, which might for instance describe a linear predictor of a quantity of interest, i.e. $\phi_c(D) = c^T \mathrm{Cov}_D[p] c$.

This list is by far not exhaustive and many more so-called 'alphabetical' optimality criteria have been developed for different purposes [HJM24, Puk06].

Remark 2.14. Seldom, optimal experimental design is considered on the basis of sensitivity analysis where optimality criteria based on the Hessian of a cost function $\mathcal{C}$ evaluated at the ground truth parameter $p_\star$, which characterizes the size of the ε indifference region of the cost function $\{p \in \mathcal{A} \mid |\mathcal{C}(p) - \mathcal{C}(p_\star)| < \varepsilon\}$ for some small $\varepsilon > 0$ [LS82, RFKPS07]. In the common case where $\mathcal{C}$ is the quadratic cost function and under linearity of the forward map F, this represents a special case of the Fisher information matrix FIM, where additive noise follows an i.i.d. normal distribution $\mathcal{N}(0, \sigma^2 I)$ and thus

$$H_p\mathcal{C}(p_\star) = L^{-1}J^TJ = L^{-1}F^TF = \frac{\sigma^2}{L}\text{FIM}. \tag{2.19}$$

Non-Gaussian Regime. Non-Gaussian distributions of $\hat{p}$, occurring for instance for non-linear forward maps, give rise to additional difficulties in the definition of design criteria: there exists in general no closed form of the Bayesian posterior covariance matrix and the inverse Fischer information matrix only locally approximates the maximum likelihood estimator covariance, and both matrices vary in the data generating parameter $p_\star$. Different strategies have been suggested to face these challenges [HJM24, Ale21], for instance

- designing a 'local' design through implicit linearization by considering the Fisher information matrix at a guess of the ground truth parameter - for instance the prior mean or the MAP point p_{MAP} [BTGMS13],
- equivalently, a Laplace approximation of the Bayesian posterior distribution approximates the posterior by a Gaussian and corresponds to the at the MAP linearized inverse problem [LSTW13, LMT15, APSG16],
- a minimax formulation gets rid of the p dependence by assuming the worst case parameter value $\arg\min_D \max_p \phi(D, p)$, but leads to a difficult optimization problem,
- an averaging approach averages the design criteria over all unknown values, i.e. the parameter p, as well as the data y, which might for instance be required for computing the MAP - possibly weighted by a prior distribution. An example is the well-known Bayes risk in the Bayesian framework, that measures the averaged quadratic deviation of the MAP point from the data generating parameter p through

  $$\phi_{\text{Bayes}}(D) = \iint \|p - p_{MAP}(y, D)\|^2 \ell(y|D, p)\,\mathrm{d}y\,\mathrm{d}\pi_0(p),$$

  where $\ell(y|D, p)$ denotes the likelihood of observing data y under parameter p and design D.

It should be noted that the optimal experimental design framework is flexible include other design criteria that leverage different utility functions [CV95], for instance to tailor the optimization to the specific needs of the considered inversion framework.

Computational Complexity. The numerical computation of optimality criteria, as well as their optimization is often challenging due to high computational complexity arising from high dimensional objects, the need for repeated inversion within the optimization procedure, and costly PDE forward solvers. Moreover, external restrictions have to be taken into account, for instance a limited number of sensors to be placed in the domain that calls for sparsity promoting penalty functions, whose design is non trivial. The development of suitable computational strategies and methods is thus an own area of research, with considerable contributions over the past decades, as summarized for instance in [HJM24, RDMP16].

2.5.3. Sequential Optimal Experimental Design.

If experiments are conducted sequentially one after another, experimental design can benefit from designing the next experiment adapted to the information, for instance the parameter reconstruction, gained in previous experiments [HJM24, RFIBS24]. Two strategies can be distinguished: Greedy approaches design the next experiment based on feedback from previous experiments and thus allow application of the previously derived techniques to continuously updated prior distributions, which reduces computational complexity. In contrast to that, 'fully' sequential approaches also take subsequent future experiments into consideration for planning, and build the design to maximize the total information collected in all remaining experiments. These approaches are instead based on Markov decision processes and therefore share similarities to reinforcement learning.

Part I.

Relaxation of Theory Approach

3

The Inverse Problem for Chemotaxis

Evolution created biological systems of an immense diversity, and with it an incredible complexity, expressed for instance in redundancies of systems [NBCS97, OL02]. Many phenomena are not fully understood up to today [SI19]. This complexity, together with a lack of fundamental principles, poses a major difficulty in quantitative modelling in biology. Models often represent simplifications, and model parameters have to be fitted to observed experimental data, which leads to the realm of inverse problems.

In this first part of this work, we study one particular inverse problem from mathematical biology, that seeks to reconstruct the coefficient in a PDE that describes the directed motion of bacteria. We assume synthetic data of sub-optimal type, as often encountered in real applications due to experimental restrictions or insufficient measurement procedures, and study suitability of this experimental setting for parameter identification. The constructive proof of structural identifiability that is developed in the next chapter will allow the application of the 'relaxation of theory' approach of experimental design, that is then carried out in Chapter 5, turning this part into an example for this approach.

This chapter serves as an introduction to the model and the associated inverse problem.

3.1. Modelling Chemotaxis

In biology, the ability to move autonomously and react to external stimuli represent two characteristics of life. Since they constitute an evolutionary advantage for many organisms in many situations, a large variety of distinct reaction systems [Adl75] and motility patterns have evolved among different species [WB22, Hen72]. Chemotaxis combines both phenomena and describes the motion of organisms in response to an external (chemical) stimulus, which occurs, for instance, in the context of food procurement, communication between cells through self-produced chemicals or avoidance of dangers such as environmental toxins. Applications can be found e.g. in motility of bacteria [Ber04] or synthetic micro-swimmers [LL18] for instance in the context of bioreactors [SO08] or the spread and prevention of diseases [JS02], cancer metastasis [LNJ18], biofilm formation [OFD16],

tissue engineering [IMB+06], immune response [OY05] or pest control [LTD+15], just to mention some.

3.1.1. Biological Background.

Bacterial chemotaxis is particularly well studied for Escherichia coli (E.coli) bacteria, which serves as a model system to study the biochemical signalling pathway [WA04] as well as for active matter modelling [SLAJ+16]; see for instance [Adl75] and references therein or [Ber04, Ber75] for a summary of the history of investigation of bacterial motion.

Bacterial Motility: Run-and-Tumble. Swimming motility of E.coli bacteria is driven by propellers, the so-called flagella. These organs are long, helical filaments that sit on the cell surface and can be rotated in two directions by the motors on their roots, steered by intracellular signals. Counter-clockwise rotation allows the flagella to align and form a bundle that rotates smoothly, strongly propelling the cell forward along a straight line. This phase is called 'running'. A change of rotation direction of one or more flagella from counter-clockwise to clockwise leads to a noisy motion of the bacterium termed 'tumbling'. It is induced by the change of the rotation direction itself as well as the resulting untangling of the previously formed flagella bundle. In this process, the bacterium changes orientation. Switching back to counter-clockwise rotation of all flagella initiates the next running phase. Since the new direction is chosen at random, the resulting motion describes a random walk, that can be idealized as non-correlated alternating phases of running and tumbling. A more detailed presentation can be found in [Ber04] and references therein.

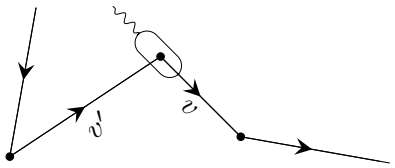


Figure 3.1.: Run-and-tumble motion of a single bacterium.

Chemoattractant sensing. E.coli reacts to a variety of different chemicals, that either attract it (sugars, Aspartate, Serine, et cetera (e.t.c.)) or provoke a repulsive reaction (Acetate, alcohols, low or high pH value, Benzoate, e.t.c.) [Adl75]. Those chemicals are detected by sensors on the cell surface. Because bacteria are too small to sense gradients along their length, a temporal mechanism involving an excitation and adaptation processes in the internal pathway system is used to memorize a previous concentration of the chemical stimulus and compare it to the current one [HH83, MBBVO03].

Chemotactic response. In this way, cells can determine whether they ascend or descend the chemical gradient in the running phase. This information is passed to the flagellar motor. In case of a chemoattractant, the tumbling frequency is decreased when the bacterium is moving up the gradient, and increased when it is moving down the gradient

[BB72] (and vice versa for chemorepellents). This biases the random walk towards directions of increasing gradient of the chemoattractant concentration and permits a directed motility.

Modelling assumptions. Typical approximations in modelling, that will also be adopted in our framework, are:

(M1) All bacteria run with the same constant speed. It can be observed experimentally [BB72], that the speeds of tracked cells stabilize quickly to an almost identical value after every tumbling event - for wild type E.coli in a homogeneous medium, this speed is 14.2μm/s.

(M2) The change in velocity due to tumbling happens instantaneously. This assumption is justified by the discrepancy in size of run and tumbling times, which are exponentially distributed with means 0.86s and 0.14s, respectively, for wild type E.coli in a homogeneous medium [BB72].

(M3) Runs are straight lines. In fact, collisions with surrounding water molecules result in random displacement and rotation of the cells. The effect, however, is manageable for short run lengths, which explains why cells prefer those [Ber04].

(M4) A full space setting is considered $x \in \mathbb{R}^d$ of dimension $d \in \{1, 2, 3\}$. This corresponds to the assumption that the experimental domain, e.g. a petri dish, is infinitely wide, and is frequently assumed in the context of chemotaxis modelling [CMPS04, ODA88]. It avoids the necessity to deal with boundary conditions, that might vary between different experimental setups and species. Suitable boundary conditions could for instance be of reflective [Alt80], Neumann (no flux) [OH02] or no-slip [ZLCZ15] type. For the models in this work, the simplification to full space is justified by the local nature of the considered reconstruction strategies with short experimental time and compactly localized initial data, in combination with a constant speed of propagation according to Assumption (M1), which bounds the data away from the boundary.

(M5) Birth-death effects are neglected due for a short observation time. In fact, birth usually does not happen while cells are in motion [OH02] and the cell division time of around 20min under optimal conditions [Ber04] is large compared to the chemotaxis time scales as given in (M2).

Moreover, we restrict ourselves to models that satisfy the following additional assumption:

(M6) The chemoattractant concentration is stationary in time, and bacteria do not interact with the chemoattractant. For short experimental times, as will be considered in our inversion framework, this can be considered an approximation of models with bacterial interactions, e.g. through auto-chemotactic behaviour or consumption of the chemical.

In this setting, many perspectives can be taken to model chemotactic behaviour, and we will focus on the mesoscopic scale. A brief overview over other types of models is given afterwards in Sections 3.1.3 and 3.1.4. To properly introduce the model, we make use of the following notation.

Notation. The velocity space V collects velocities v that the bacteria can move into, and is set to $V := \mathbb{S}^{d-1}$, the unit sphere in $\mathbb{R}^d$, to reflect assumption (M1) of a constant bacterial speed. The convention $\mathbb{S}^0 := \{\pm 1\}$ in 1D is used. Accordingly, integrals $\int_V h(v)\,\mathrm{d}v$ of a function h over V are considered surface integrals over the sphere $\mathbb{S}^2$ in dimension $d = 3$, line integrals along the unit circle $\mathbb{S}^1$ in dimension $d = 2$ and sums $h(+1) + h(-1)$ in dimension $d = 1$. For $k \in \mathbb{N}_0 \cup \{\infty\}$ and $p > 0$, Lebesgue spaces L^p, the spaces of p-times continuously differentiable functions C^p, as well as Sobolev spaces $W^{k,p}$ with the usual convention $H^k = W^{k,2}$, and the corresponding Bochner spaces, as well as their corresponding norms on $[0,T]$, $\mathbb{R}^d$, V or their Carthesian products, are defined and equipped with norms in the typical manner, where $T > 0$ denotes a final experimental time horizon. A subscript on a function class indicates that a subset is considered, where $+$ stands for non-negativity, c for compact support and b for boundedness of the considered functions.

3.1.2. Kinetic Chemotaxis Model

Mesoscopic models study the collective population dynamics of many particle systems while preserving individual velocity information, allowing the model to capture fine scale motions. For chemotaxis, the bacteria density $f(t,x,v)$ at time $t \in [0,T]$ is modelled as a function on the phase space $(x,v) \in \mathbb{R}^d \times V$, and denotes the density of bacteria that moves into direction v at location x at time t. Starting with initial condition $\phi \in L^1_+ \cap L^\infty(\mathbb{R}^d \times V)$, that is non-negative a.e.w. in V, the further evolution of f can be described on a statistical level using a linear kinetic PDE [Alt80, ODA88, OH02, CMPS04, EO04]

$$\partial_t f + v \cdot \nabla_x f = \mathcal{K}(f) := \mathcal{L}(f) - \sigma f\,, \tag{Ch}$$
$$f(t=0,x,v) = \phi(x,v). \tag{iCh}$$

This PDE resembles the run-and-tumble behaviour that was described above: The left-hand side (LHS) of (Ch) describes a movement along a straight line in direction v - this corresponds to the 'run' phase. The RHS, on the other hand, accounts for the velocity jump due to reorientation during tumbling and exhibits a diffusive nature. It is composed of the two terms given by

$$\mathcal{L}(f)(t,x,v) = \int_V K(x,v,v') f(t,x,v')\,\mathrm{d}v' \quad \text{and} \quad \sigma(x,v) = \int_V K(x,v',v)\,\mathrm{d}v'\,, \tag{3.1}$$

which are determined by the so-called tumbling kernel $K(x,v,v')$ that encodes the transition probability of individuals to changes their velocity from $v' \in V$ to $v \in V$ at location $x \in \mathbb{R}^d$. As such, the first term $\mathcal{L}(f)$ represents a statistical gain to $f(t,x,v)$ from bacteria

from other velocities $v' \in V$, sometimes denoted by the short notation

$$f' := f(t, x, v'), \tag{3.2}$$

that tumbled and attained velocity v afterwards, and the second term $-\sigma f$ describes a statistical loss of bacteria from $f(t, x, v)$ that tumbled into any other velocity v'. As suggested by the individual chemotactic response behaviour, the effect of the chemical stimulus enters into the equation through K by changing the tumbling frequency.

Remark 3.1. Equation (Ch) is a so-called kinetic PDE. These models are often handy when modelling many particle systems, as they live on the mesoscopic scale, between microscopic models that trace trajectories of single individuals, and the macroscopic level population models that lack any individual information. They allow observation of certain microscopic structures of the motion, while avoiding the computational complexity under which microscopic models suffer for a large particle number. These models find applications in neutron [DS58], photon or electron [RL86] transport, rarefied gas dynamics [Cer12], animal migration [TKLRC14] or even opinion formation [ACD24, CFRT09, CLP24, MT14, Tos06]. In comparison to standard kinetic models, such as the Vlasov equation for plasma [Vla68] or Boltzmann equation in gas dynamics [Bol72], our model does not contain a force term, because bacteria are assumed to run independently of any external force acting to them, and the tumbling operator on the RHS is linear in f, resembling the fact that the velocity change is an individual decision of the bacteria that does not require interaction with a second individual, as for the collision operator in gas dynamics. There exists a rich literature on kinetic models, tools for their analysis and numerics, and the interested reader is referred to [Cer12, Per04, DP14, SK13].

Remark 3.2. The one-dimensional setting $d = 1$ corresponds for instance to scenarios where experiments are conducted in tubes [GAM+15, SCB+11, SCB+10]. In this setting, (Ch) is sometimes regarded as a two species model for the species of forward and backward moving individuals $f^{\pm}(t, x) = f(t, x, \pm 1)$, whose reaction term corresponds to tumbling [OH02, HS00, ODA88, HRL01]. The two species system is also known as the telegraph or Goldstein-Kac process [McK67, Gol51, Kac74].

Admissible set for K. Throughout this work, we assume the tumbling kernel to be independent of time, which is a consequence of a stationary chemical stimulus by assumption (M6), which, as argued earlier, is a valid approximation of the non-stationary cases, such as the auto-chemotaxis case, under short observation times. Similarly, temporal changes in K due to other potential influence factors are assumed to happen on a slower time scale than the chemotactic behaviour, also allowing for a stationary approximation.

The interpretation as a turning probability motivates the standard assumptions of non-negativity and uniform boundedness by some constant $C_K > 0$. Since tumbles with coinciding incoming and outgoing velocity cannot be distinguished from straight running by assumption (M2) and have no influence on the solution, the value $K(x, v, v) = 0$ is prescribed. In summary, K is supposed to belong to the admissible set

$$\mathcal{A}_K = \{K \in L^\infty(\mathbb{R}^d \times V \times V) \mid \quad K \geq 0, \quad \|K\|_\infty \leq C_K, \quad \text{and} \quad K(\cdot, v, v) \equiv 0 \ \forall v \in V\}. \tag{3.3}$$

Existence of solutions. Existence of solutions to initial value problem (IVP) (Ch)–(iCh) is well-established. In literature, the linear case is treated in [HO00] by semigroup theory. Mostly the more convoluted case of auto-chemotaxis is studied, where (Ch) is coupled with an elliptic or parabolic equation that describes diffusion of the chemoattractant and production by the bacteria. The emerging non-linearity is treated by the vanishing viscosity method in [HRL01, HKS06], or a contraction principle in [HS00] for the 1D models. In 2D or 3D, semigroup theory and a-priori bounds were exploited to provide the respective results in [CMPS04, HKS05, BCGP08]. All previously mentioned references also provide preservation of non-negativity in the initial data. This list merely points towards some relevant literature and has no intent to be exhaustive. A review on existence and blow-up results for the auto-chemotaxis system can be found in [BC10].

Because the thus considered model (Ch)–(iCh) is linear, such sophisticated methods are not necessary and standard arguments suffice: Semigroup theory directly provides existence of solutions, and preservation of non-negativity of initial data can be shown by source iteration, in analogy to [Maj97, HW07]. According to the hyperbolic nature of (Ch), compactness of the x-support of initial data is preserved for $f(t)$. These findings are summarized in the following proposition and a detailed proof can be found in Appendix A.1 for sake of completeness.

Proposition 3.3. *Let $T > 0$, $K \in \mathcal{A}_K$ and the initial condition $\phi \in L^1_+ \cap L^\infty(\mathbb{R}^d \times V)$ be non-negative for almost every (a.e.) $v \in V$. Then IVP (Ch)–(iCh) attains a unique mild solution $f \in C^0([0,T]; L^1_+ \cap L^\infty(\mathbb{R}^d \times V))$, that is non-negative $f(t,x,v) \geq 0$ for all t, x, for a.e. $v \in V$ and bounded by*

$$\|f(t)\|_{L^1 \cap L^\infty(\mathbb{R}^d \times V)} \leq e^{|V| C_K t} \|\phi\|_{L^1 \cap L^\infty(\mathbb{R}^d \times V)}. \tag{3.4}$$

Moreover, if the initial data ϕ has compact x-support a.e.w. in V, in a sense that there exists a compact set $S \subset \mathbb{R}^d$ such that for almost all (a.a.) v one has $\phi(x,v) = 0$ for all $x \in \mathbb{R}^d \backslash S$, then $f(t)$ has compact support for all $t \in [0,T]$.

Moreover, the solution to (Ch) admits regularity with respect to the tumbling kernel, which can be established by consideration of difference equations, as explained in Section A.1.1.3.

Lemma 3.4. *The solution f of (Ch)–(iCh) is twice continuously differentiable in $K \in \mathcal{A}_K$, and Lipschitz continuity holds for f and its derivatives holds under the L^∞ norm on $\mathcal{A}_K$, i.e. there exists a constant $C = C(T, C_K, \|\phi\|_{L^1 \cap L^\infty(\mathbb{R}^d \times V)})$ independent of K such that, for instance,*

$$\|f_K - f_{\tilde{K}}\|_{C([0,T]; L^1 \cap L^\infty(\mathbb{R}^d \times V))} \leq C \|K - \tilde{K}\|_\infty.$$

3.1.3. Chemotaxis Models across the Scales

Like many physical phenomena that describe particle dynamics, chemotaxis can be modelled on different physical space-time scales [Hor03]. Figure 3.2 summarizes these models,

as well as their relations, and serves as an overview over the subsequent paragraphs that give a brief insight into different modelling regimes for chemotactic motion and the key ideas how these are related. Rigorous formulations and results can be found in the referenced literature.

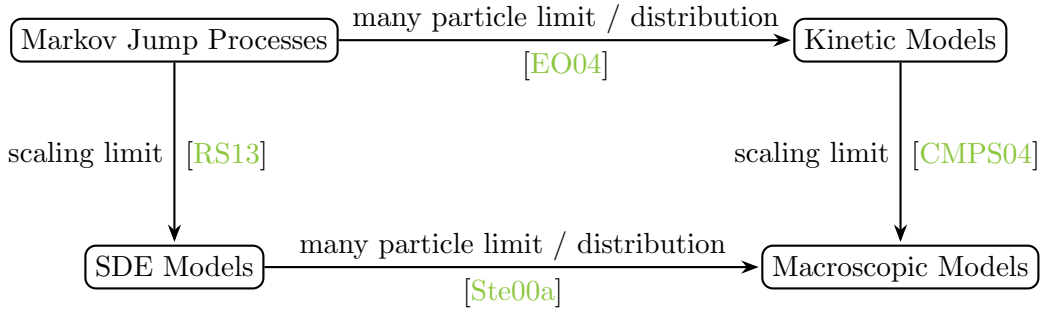


Figure 3.2.: Relation between particle models on different scales.

Individual Based Models

Individual based models, being placed on the microscopic scale where individual cells can be distinguished, provide very detailed information of potentially very fine motions. They model the single trajectories $x_n(t) \in \mathbb{R}^d$, $n = 1, ..., N$, of a finite, but possibly large number $N \in \mathbb{N}$ of individual cells.

Markov Jump Processes. Markov jump processes track the evolution of the current state of each individual, described in our case by its current location $x_n(t)$ and velocity $v_n(t) \in V$. They are the models closest to the microscopic description of chemotaxis in Section 3.1.1, and explicitly describe the running phase with velocity $v_n(t)$ in (3.5), and the random change to a new velocity ν_k at a randomly Poisson distributed jump time T_k in the tumbling phase in (3.7). New velocities ν_k are adopted according to a probability measure m on V. One of the simplest model of this type is the following, as proposed in [RS13, PHA+17, ODA88]:

$$\mathrm{d}x_n(t) = v_n(t), \tag{3.5}$$

$$\int_{T_k}^{T_{k+1}} \sigma(x_n(t), v_n(t))\,\mathrm{d}t = \zeta_{k+1}, \tag{3.6}$$

$$v_n(t) = \nu_k \sim m, \qquad \text{for } t \in [T_k, T_{k+1}], \tag{3.7}$$

where ζ_k are i.i.d. random variables with normalized exponential distribution, such that (3.6) generates jump times $(T_k)_k$ according to a Poisson process with space and velocity dependent rate σ.

SDE Models. A parabolic scaling can be introduced to place the above Markov jump process the long time or equivalent frequent jump regime, in which the model converges

in the diffusion limit to an SDE, describing the trajectories $x_n(t)$ as biased random walks of the form

$$\mathrm{d}x_n(t) = \gamma(t, x_n(t))\,\mathrm{d}t + \theta(t, x_n(t))\,\mathrm{d}W_t^{(n)}, \qquad n = 1, ..., N \tag{3.8}$$

where the $W_\cdot^{(n)}$ denote independent $\mathbb{R}^d$-valued standard Brownian motions, and γ and θ describe the bias (drift) and random parts (diffusion) of the motion [Ste00a, Ste00b]. Chemoattractant information enters into the drift term γ, to bias the motion towards high concentration regions. Note that the state variable is now $x_n(t)$, and velocity information has been lost, as the model is placed in the frequent tumble limit where the velocity constantly changes.

Population Models.

Because the degree of detail that individual based models provide might not be necessary in many contexts, and to mitigate the enormous computational cost of running these models for a large number N of particles, population based models were developed that summarize the motion of a population by the propagation of its particle density, under the approximation of an infinite number of infinitesimally small particles. These models emerge as many particle limits $N \to \infty$ of individual based models. Despite the approximation, population based models are well capable of capturing significant characteristics of the motion and show good alignment with experimental data [EGB+16, SCB+10, SCB+11], if parameters are chosen appropriately.

Kinetic Models. Kinetic models lift the Markov particle models to a statistical level and describe the temporal propagation of the probability density of the state of the model, in our case denoting the current location and velocity $(x(t), v(t))$. The kinetic chemotaxis model (Ch) thus emerges as the Kolmogorov forward equation from Markov jump process (3.5)–(3.7), when choosing the velocity turning probability at a given (x, v) as the rescaled tumbling kernel $m = (\sigma^{-1}K)(x, v, v')\,\mathrm{d}v'$ [RS13, EO04]. The state density f can be regarded as the many particle limit of the empirical distribution function of the particles $(x_n, v_n)_n$ by the law of large numbers, i.e.

$$\rho^{\mathrm{kin}}_{(N)}(t, x, v) = \frac{1}{N}\sum_{n=1}^{N} \delta_{x_n(t)}(x)\delta_{v_n(t)}(v) \qquad \xrightarrow{N\to\infty} \qquad f(t, x, v).$$

Macroscopic Models. Macroscopic PDE models can be derived either as space-time scaling limits from kinetic equations [CMPS04], or as so-called mean-field limits of SDE models [Ste00a], i.e. as Lebesgue densities of the measures that arise as limits $N \to \infty$ of the empirical density

$$\rho_{(N)}(t, x) = \frac{1}{N}\sum_{n=1}^{N} \delta_{x_n(t)}(x)$$

of particles $(x_n(t))_n$ satisfying (3.8). The most well-established model for bacterial chemotaxis, the Patlak-Keller-Segel (PKS) model [Pat53, KS70, KS71]

$$\partial_t \rho - \nabla \cdot (D \cdot \nabla \rho) + \nabla \cdot (\rho \Gamma) = 0, \tag{3.9}$$

represents a parabolic PDE and emerges as the diffusion limit of a parabolic space-time scaling [CMPS04]. It describes an advection-diffusion behaviour with drift term Γ and diffusion coefficient D, that are linked to the kinetic tumbling kernel K, or equivalently the turning velocity measure m and time rate σ in the Markov process (3.5)–(3.7), or the drift γ and diffusion θ in the SDE model (3.8).

By emphasizing the transport part of the kinetic model (Ch) instead of the tumbling, hyperbolic scaling limits allow to derive macroscopic hyperbolic PDE models for chemotactic behaviour [HS00, FLP05, ST16], which are better suited in certain situations, for instance when travelling waves occur in large chemoattractant gradient environments.

3.1.4. Refined Models

The modelling of chemotaxis, and related questions have constituted and still constitute an active area of research. In the following, we mention some directions, without any intention to cover the field exhaustively. Instead, we direct the interested reader to one of the many review articles, such as [Hor03, HP09, TMPA08, AT21].

Auto-Chemotaxis. A frequently studied scenario is auto-chemotactic motion, where bacteria themselves secrete the chemoattractant and thus exhibit self-attracting behaviour [TKL18]. In this case, the models of bacterial motion are coupled to a model describing the propagation of the chemoattractant and its production by the bacteria, most commonly a Poisson equation or parabolic PDE [CMPS04, BCGP08, KS70]. The chemoattractant sensing in the chemotaxis equation then renders the coupled model non-linear, and the bacterial self-attraction provokes interesting behaviour, such as finite time blow-up [BC09, DP04, BDP06], travelling pulses or waves [SCB$^+$11, SCB$^+$10], as well as pattern formation [PY18].

Signalling Pathway. The intracellular signalling pathway describes the protein interactions that govern chemotactic behaviour within one cell, by steering the motors of the flagella and inducing the adaptation process in gradient sensing. The pathway is particularly well understood for E.coli bacteria, as summarized in the overview article [TPM$^+$08]. By incorporating an additional internal state variable, the adaption behaviour can be taken into account in the chemotaxis models [EO04, XO09], leading to pathway based Markov velocity jump processes [RS13], pathway based SDE models [ST16] or pathway based kinetic models [STY14, PTV16]. Under large chemoattractant gradients, these models explain observed average drift velocities by a hyperbolic macroscopic limit [ST16], and the consideration of noise in the pathway leads to macroscopic fractional diffusion models [PST18, XT21] that align with the Levy flights that can be observed experimentally in certain regimes.

Interaction of Bacteria and Fluid Environment. A relatively new field studies chemotaxis in a dynamic fluid environment, with interactions in both directions: the fluid velocity drives the bacteria and chemoattractant, but bacterial motion also affects the fluid. This intricate relationship is typically modelled by coupling an auto-chemotactic PKS model with a Navier-Stokes equation [GH21, ZZZ21, LL24] which can be shown to suppress blow-ups of the bacterial concentration under certain conditions [ZZZ21].

Adding more Physics. Finally, some efforts have been undertaken to adjust the modelling as closely to reality as possible, and avoid some of the simplifying assumptions (M1)– (M6) at the end of Section 3.1.1:

- unphysical overcrowding in blow-ups under auto-chemotactic behaviour can be avoided by introducing volume filling or quorum sensing strategies [PH02, CR06],
- birth/death effects can be incorporated in the model [ODA88, OH02],
- non-local sensing of the chemoattractant is often introduced to account for the adaptation process [OH02, CMPS04],
- non-instantaneous tumbling can be modelled via introduction of a resting phase, often combined with birth during the resting phase [ODA88, OH02],
- the consideration of multi species models allows studying e.g. competition scenarios [BW16, ESV09].

3.2. Inverse Problem for Chemotaxis

In order to run realistic simulations and derive implications on the real world, a realistic choice of chemotaxis model parameters is essential. These parameters can be influenced by a plethora of potential factors beyond the chemoattractant, for instance the geometry of the domain [RLS+19] or the temperature [MISO76] as well as internal factors such as the type of considered bacteria [BB72, SCF+19]. A lack of general, first-principle-based descriptions on how the previously mentioned factors interact and translate to concrete values of model parameters, together with the fact that these parameters are typically not directly observable, provoke adoption of the inverse perspective in which model parameters are estimated on the basis of experimental data, and explains the ongoing interest of biologists and practitioners in model fitting [RHL86, TZL88, FL91, SCF+19, BDM+22, SLS17, GAM+15].

In this work, we focus on the mesoscopic chemotaxis model (Ch), where the tumbling kernel K uniquely determines the law of bacterial motion. Our goal is to study the inverse parameter identification problem that amounts to identifying K from experimental data on the bacteria density f as an observable quantity. The underlying experiment is described below and presented graphically in Figure 3.3.

3.2.1. Experimental Setting

Forcing data. The lab experiment is started by placing bacteria in an environment with a fixed stationary concentration of the chemical stimulus. This initial bacterial configuration is assumed to be tightly controllable by the practitioner and shall be prescribed by a compactly supported, non-negative initial density $\phi \in L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V)$ in (iCh).

Experiment and Model. During the experimental time horizon $[0,T]$ motility takes place. Mathematically, this is described by the linear kinetic model (Ch) and its solution for a given initial data ϕ is denoted by $f_K^\phi \in C([0,T]; L^1_+ \cap L^\infty(\mathbb{R}^d \times V))$ according to Proposition 3.3. The sub- and superscripts, explicitly expressing the value K of the tumbling kernel and the initial data, may be neglected in cases where they are clear from the context. Adopting the full space setting despite the confinement in real experiments is justified by the discussion under assumption (M4), which also suggests that the presented results extend to finite domain problems.

Measurements. Obtaining velocity dependent data on the bacteria density $f(t,x,v)$ is technically challenging if a large number of individuals is considered, because it would require tracking of the single cell trajectories in time through video recordings which requires specifically designed equipment [JJH+19, GT21, ZBFS10, MBC+12], entails high computational cost and might even distort the data through photo-toxicity [BMDB09]. Additionally, such elaborate techniques are barely feasible in some contexts. Instead, our measurement will be based on the velocity-averaged, macroscopic bacteria density

$$\langle f \rangle := \int_V f \, \mathrm{d}v.$$

This density is mostly accessible and can be obtained either from measurements on an individual level, when reading off a photo by counting individual cells on a grid to obtain a histogram [JDBC12], or on a population level through optical density measurements [MBJPM85, BFTea20].

Localized interior domain data is collected in time by detectors, which are characterized by their profile, a space-time test function $\mu(t,x) \in L^1([0,T] \times \mathbb{R}^d)$, that we assume to be controllable by the practitioner. The measurement operator modelling this detector $M_\mu : C^0([0,T]; L^1_+ \cap L^\infty(\mathbb{R}^d \times V)) \to \mathbb{R}$ is then given by the duality product

$$M_\mu(f) = \int_0^T \int_{\mathbb{R}^d} \langle f \rangle (x,t) \; \mu(t,x) \; \mathrm{d}x \; \mathrm{d}t \quad = \int_0^T \int_{\mathbb{R}^d} f(t,x,v) \, \mathrm{d}v \, \mu(t,x) \; \mathrm{d}t \, \mathrm{d}x \, . \tag{M}$$

Remark 3.5. A typical example for the choice of a measurement test functions would be a characteristic function on a squared domain and a temporal interval. This corresponds to a pixel reading from a photo with a given exposure time.

Inverse Problem. In summary, the single experimental setup forward operator $F_{\phi,\mu} : \mathcal{A}_K \to \mathbb{R}$ for this experiment takes the form

$$F_{\phi,\mu}(K) = M_\mu(f_K^\phi) = \int_0^T \int_{\mathbb{R}^d} \int_V f_K^\phi(t,x,v) \, \mathrm{d}v \, \mu(t,x) \, \mathrm{d}x \, \mathrm{d}t. \tag{F}$$

To simplify notation, dependencies on ϕ, μ in previously defined operators may be surpressed when the context is clear. As initial data ϕ and the measurement test function μ are supposed to be controllable, the input-to-output map for a fixed value $K \in \mathcal{A}_K$ reads

$$\mathrm{ItO}_K : \quad L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V) \times L^1([0,T] \times \mathbb{R}^d) \quad \to \quad \mathbb{R}, \qquad (\phi, \mu) \mapsto F_{\mu,\phi}(K). \quad \text{(ItO)}$$

After fixing a design $D \subset L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V) \times L^1([0,T] \times \mathbb{R}^d)$ and an admissible set $\mathcal{A}'_K \subset \mathcal{A}_K$, the inverse problem amounts to

$$\text{find} \qquad K \in \mathcal{A}'_K \qquad \text{such that} \qquad y_D = F_D(K), \qquad (\mathrm{Ch}^{-1})$$

where y_D is the experimentally observed data under this design D. The inverse problem is non-linear by non-linearity of the forward map in K, given that the solution f_K^ϕ depends non-linearly on K. Note however, that the input-to-output map is linear in initial data ϕ and measurement test function μ. Depending on the choice of the admissible set $\mathcal{A}'_K$ for K, the problem can be either framed as a non-parametric or a parametric inversion.

Remark 3.6. If several experimental setups $s = (\phi, \mu) \in D$ are conducted with the same initial data ϕ, this means that these data can be collected from the same experiment by using multiple detectors.

Remark 3.7. The difficulty to measure velocity dependent data might seem contradictory to the possibility to freely prescribe initial data $\phi(x, v)$, also in velocity. But generating an initial configuration and data measurement are distinct tasks, each bearing its own challenges. In fact, experimental apparatuses have been developed that can even realize almost singular in x and v configurations, e.g. micro-confinement in a thin tube was used with E.coli bacteria [SSW13] and artificial micro-swimmers [LZWZ16]. Euglena gracilis algae were manipulated by polarized light in [YHZZ21] and the references therein. This justifies the use of such initial data.

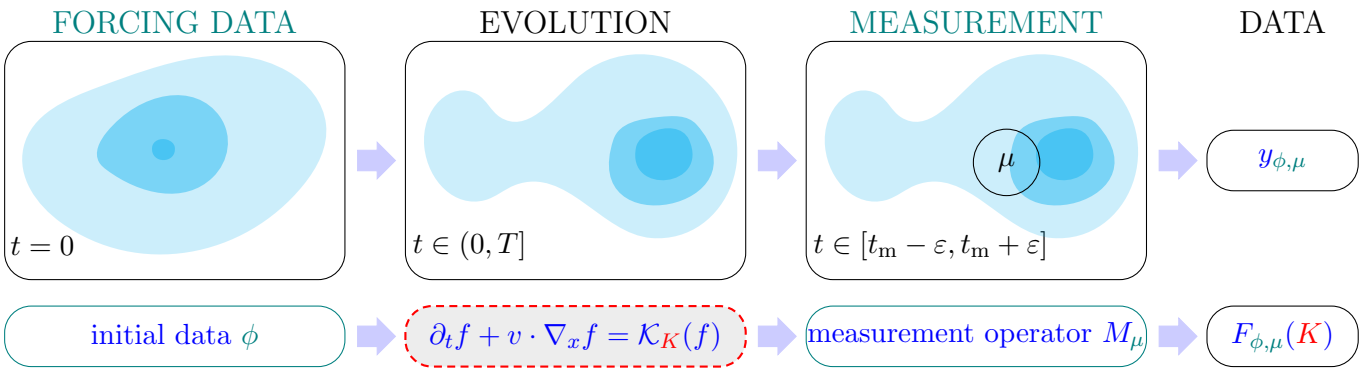


Figure 3.3.: Experimental setup, using the same structure and colour coding as in Figure 2.1: The initial configuration of bacteria is imposed at $t = 0$. Chemotactic behaviour takes place in the time interval $(0, T]$ and the final configuration of bacteria is measured by a space-time measurement test function μ, for instance compactly supported in $[t_m - \varepsilon, t_m + \varepsilon]$ in time and the circle in space.

3.2.2. Research question.

A closer look at (Ch^{-1}) exhibits an intrinsic challenge: The parameter $K(x, v, v')$ to be inferred lives on the microscopic level. It describes the probability of a velocity jump and thus depends on the jump location x, as well as incoming and outgoing velocities v' and v. In contrast to that, the data is only collected from the velocity-averaged bacteria density $\langle f \rangle$ on the coarser macroscopic level that lacks velocity information, i.e. it is non-local in the velocity. Intuitively, it is not clear how such data can reflect the fine characteristics of the velocity jump coefficient K. This brings us to the following question :

Can data on the velocity averaged kinetic bacteria density $\langle f \rangle$ contain enough information to allow a unique reconstruction of the kinetic tumbling kernel $K(x, v, v')$? (Q)

In fact, the non-locality reduces available information in the data, which lives on a lower dimensional space compared to velocity dependent data, and was shown to introduce major difficulties that hinder unique reconstruction in a related context, for the inverse problem of recovering of the scattering coefficient K in the kinetic radiative transport equation from angularly averaged data [Bal09, BLM08, Lan08, BJ09, BJ10, BJJ10]. On the other hand, data on the interior of the domain $[0, T] \times \mathbb{R}^d$ is richer than, for instance, frequently considered boundary data, and the possibility to control the initial condition ϕ and space-time measurement test function μ of the single experiments offers a detailed insight into the chemotactic behaviour of the macroscopic bacteria density. Each of these aspects on its own already ameliorates the reconstruction, as the previously mentioned literature suggests, but none could fully remediate the degradation in information due to velocity averaging. A detailed review of this literature can be found in the literature survey of Chapter 4.

Before deriving a suitable experimental design, it is thus necessary to study whether the experimental setting described in Section 3.2.1, combining all these ameliorations, is at all suitable for the purpose of inversion for K. This naturally calls for identifiability analysis, and the two major aspects that will be addressed in this work are

(Q.i) suitability of the experimental setting, i.e. of the macroscopic data under the freedom of choosing initial and measurement data, and

(Q.ii) construction of a suitable experimental design.

This information is fundamental to the inverse problem and should be studied before inversion. In case (Q) is answered with a 'no', the principle of 'no free lunch', or as stated by Lanczos [Lan96]

"A lack of information cannot be remedied by any mathematical trickery.",

points out that even regularization methods that can render an inverse problem well-posed will only generate a result that might not be very informative about the ground truth, unless the introduced additional information is in good alignment with the properties of the ground truth parameter.

In addressing (Q.ii) we will lay out the "relaxation of theory" approach to qualitative experimental design, which will be based on the theoretical considerations under (Q.i).

3.2.3. Literature on Inverse Problems related to Chemotaxis.

In the subsequent literature survey, we focus on chemotaxis models of the types described in Section 3.1.3. We thus exclude inverse or model fitting problems dealing with the signalling pathway modelling [Tu13] or other forms of modelling chemotactic behaviour, for instance through cell boundary deformation [CEL+15].

Most literature on inverse problems for chemotactic behaviour refers to the PKS model, or reads off microscopic tumbling statistics from microscopic trajectory data.

Parametric Inference from Microscopic Statistics. In [PHA+17, SABS18], for instance, the authors used the microscopic statistics to infer the parameters of a Markov velocity jump process model similar to (3.5)–(3.7). A generalization to renewal theory models, that need not neccessarily be Markovian, can be found in [ZKZ+24, KZZ+24]. To test accuracy of a kinetic model to simulate auto-chemotactic behaviour, the authors in [SCB+11] adopted a parametric form of the tumbling kernel, based on microscopic running statistics, which could be read off from microscopic trajectory data. Similarly, a 1D macroscopic PKS equation has been fitted to real data with this technique in [ZWWC+14, LF01]. The strategy is extended to a PKS model with two chemoattractants in [WECW97] with approximations for shallow chemoattractant gradients. More recently, this methodology was applied to fit other types of stimulus induced motility in cells such as aerotactic behaviour in Burkholderia contaminans bacteria [BDM+22] or electrotaxis of fibroblast cells [SLS17] with application in wound healing.

These examples shed light on the fact that a discrepancy in scale of the measured quantity and the to be inferred parameter often occurs in biological frameworks. In the above examples, the collected microscopic data lived on a finer scale than the mesoscopic or macroscopic parameters, allowing a derivation of their parametric forms from the limiting consideration. [KA21a] represents an exception where the authors estimated a parameter of a microscopic SDE model (3.8) by minimizing the difference of its macroscopic distribution to observed data 1D. The following literature now treats the case where both the data and the parameter live on the same macroscopic scale.

Parametric inversion for PKS. Many authors focus on fitting the PKS models to various settings, where the main goal lies in the reproduction of experimental data. The inverse problem is regularized by complexity reduction through proposing parametrized expressions for the model parameters with very few specification parameters. The expressions are obtained heuristically. Reconstruction of the low dimensional specification parameters from experimental data represents an efficient approach when the proposed expression is valid. Whereas [PHA+17] used data on the chemoattractant decay to fit these parameters, most frequently, bacterial density data is used. Examples can be found in [RHL86, TZL88, FL91], where data on the chemotactic behaviour of E.coli is collected in different gradient generating chambers and parametrized forms of the drift and diffusion coefficient are fitted. Different domain geometry and different heuristic formulas are used in [SCF+19]. In [GAM+15] the performance of different potential formulas for the chemotactic drift term in the PKS model, that are fitted to data on the phototactic behaviour of

Euglena gracilis, is compared. Together with the historical development for the heuristic formula for the drift term as displayed in [FL91], the last example elucidates that heuristic formulas are not always available and might not hold in certain regimes or for different species of organisms. In such cases the results under this approach are questionable.

Non-parametric inversion for PKS.
Auto-Chemotaxis. A non-parametric setting is considered in [FM08, EPS15] for the autochemotactic PKS model (in [EPS15] with volume filling), where the chemotactic drift term is reconstructed as a function for which a-priori only mild regularity and structural assumptions are prescribed. In this rather general setting, the inverse perspective for the reconstruction of the drift is adopted and unique identifiability is studied theoretically. Numerical simulations are based on a Tikhonov regularized least squares setting to mitigate stability issues and use a finite element discretization of the drift coefficient with a large number of degrees of freedom, compared to previously mentioned model fitting approaches. Unique identifiablity of a solution dependent diffusion coefficient of this model is discussed in [EPS15].

The authors in [HLH21], instead, point towards the possibility to recover an arbitrary discretization of the unknown chemotactic drift term as a trainable variable in a neural network through training. A different access to parameter reconstruction through optimal control is provided in [FM03] for identification of a death or equivalently harvesting rate of cells including numerical evidence by finite differences, or in [DP19] for identification of boundary data on the chemoattractant equation for autochemotaxis.

Moreover, simultaneous structural identifiability of the per capita consumption rate of the chemoattractant by the bacteria and the gravitational potential of the fluid in an autochemotactic PKS Navier Stokes system was shown in [LL24] by a variation method.

Chemotaxis for a fixed stimulus. An application in cancer research lead to the development of numerical algorithms for the ill-posed reconstruction problem of the chemotactic diffusion and the chemotactic drift in a finite difference discretization of the PKS model in [BBN22]. Another methodology [HRP+21] exploits assumed radially symmetric geometry in the PKS model for drift and diffusion coefficient reconstruction by means of a linear relation to a specific function and functional regression.

Experimental Design for Biological Models. Biological experiments can be very costly, for instance due to long waiting times for the model to evolve, high maintenance cost or complex apparatuses. In addition to that, found models and parameter estimates are not always trusted due to the complexity of certain biological models and the incapability of modelling all influence factors, which explains an initial hesitation towards optimal experimental design methods, as well as the desire to balance optimality against robustness [FAJ05, HBY10]. Instigated by the development of identifiability analysis methods in systems biology, nowadays, optimal experimental design techniques have been applied and proposed in a number of publications, e.g. [CH12, BCAB08, BSE+09, LFKS13].

As pointed out in more detail in the paragraph on related literature in Chapter 5, however, experimental design for chemotaxis is so far restricted to the development of lab

apparatuses and protocols to realize experiments.

Inverse problems for kinetic PDEs. The parameters in kinetic PDEs often describe interaction between particles or, as in the case of chemotaxis, decisions of single individuals, or material properties of the background. It is often challenging to observe these parameters directly, calling for inverse problems. The most prominent example is optical tomography in medical imaging, where boundary data is used to infer optical properties, and hence interior structures, of the body. A rich body of literature on the mathematical treatment of this topic exists, and here we only guide the interested reader to review articles on theoretical results [BJ09] and their numerical counterparts [AS09, Ren10].

3.2.4. Novelty of this Work and Outline of Part I.

This is the first work to investigate the inverse problem (Ch^{-1}) of reconstructing the kinetic tumbling kernel from macroscopic data, where the coarser scale of the data in comparison to the parameter is challenging. On the other hand, this setting allows greater flexibility in designing the experiments than in optical tomography motivated radiative transport inverse problems in [BJ09] and references therein.

In Chapter 4, based on the publication [HKLT24], we give a positive answer to (Q.i) by proving structural identifiability of the tumbling kernel from velocity-averaged data for the first time, exploiting all degrees of freedom that this experimental setting provides: time dependent, interior domain data and the possibility to tightly control the initial data. The proof borrows a technique developed in the realm of optical tomography that explicitly constructs a suitable design which is capable of triggering microscopic information. This framing of singular decomposition results on kinetic inverse problems under the structural identifiability perspective is also novel. It emphasizes clearly the theoretical nature of the result, requiring access to the full input-to-output map ItO_K.

The remaining two chapters of this part are devoted to study (Q.ii) and describes the 'relaxation of theory approach' of experimental design on the example of chemotaxis. It originate from the publication [HKLT25]. The investigation of experimental designs for their suitability in the reconstruction context is also novel for chemotactic phenomena. By consideration of macroscopic density data instead of microscopic trajectory data, the question becomes relevant and can help raising reconstruction quality or decreasing experimental cost, by improving the sensitivity of data w.r.t. the parameter. Chapter 5 builds up the sensitivity based identifiability framework, that we operate in, and analytically investigates suitability of two different designs, one of which is designed by the 'relaxation of theory' approach, which demonstrates the coexistence of identifiability and non-identifiability in dependence on the design. The current part of this work is then concluded by numerical studies on the suitability of different designs in Section 8.3.2 in one and two spatial dimensions, contrasting the effectiveness of well chosen designs by the degradation of reconstructions under decreasing data diversity.

4

Structural identifiability: Experimental Setting

This section aims to answer the question (Q.i) of suitability of the experimental setting as described in Section 3.2.1 for the parameter reconstruction (Ch^{-1}). It utilizes structural identifiability analysis.

4.1. Setting and Main Results.

By postulating continuity of the tumbling kernel K on $\mathbb{R}^d \times W$, where $W := \{(v, v') \in V \times V \mid v \neq v'\}$, in this chapter, we mildly restrict the admissible set to

$$\mathcal{A}_K^{\mathrm{cont}} := \{K \in \mathcal{A}_K \quad | \quad K\,|_{\mathbb{R}^d \times W} \in C_+(\mathbb{R}^d \times W)\}. \tag{4.1}$$

Imposing this additional regularity of K has a regularizing effect on the inverse problem (Ch^{-1}). However, we regard the assumption of continuity as rather mild in practical settings.

On the other hand, the choice of the experimental setup $(\phi, \mu) \in L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V) \times L^1([0,T] \times \mathbb{R}^d)$ will not be restricted, i.e. we assume access to the full, infinite dimensional input-to-output map, and the non-parametric inverse problem of searching K in the infinite dimensional function space subset $\mathcal{A}_K^{\mathrm{cont}}$ can be described as

$$\text{find} \qquad K \in \mathcal{A}_K^{\mathrm{cont}} \qquad \text{such that} \qquad \mathrm{ItO}_K(\phi, \mu) = y(\phi, \mu) \qquad \forall (\phi, \mu) \in \mathcal{D}, \tag{4.2}$$

where the data y is a function $\mathcal{D} := L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V) \times L^1([0,T] \times \mathbb{R}^d) \to \mathbb{R}$.

Identifiability of this problem, and with it suitability of the given experimental setting described in Section 3.2.1, was proven in the publication [HKLT24] on which this chapter is based.

Theorem 4.1 (Structural identifiability; [HKLT24]). *The tumbling kernel $K \in \mathcal{A}_K^{\mathrm{cont}}$ can be uniquely recovered from the input-to-output map ItO_K. In particular, for every*

$(x, v, v') \in \mathbb{R}^d \times W$, *a proper choice of* ϕ, μ *allows to explicitly read off the point value* $K(x, v, v')$ *from the measured data* $M_\mu(f_K^\phi)$.

This theoretical results provides identifiability in the sense of well-posedness in Definition 2.1i) for the infinite dimensional forward map $F : K \mapsto \mathrm{ItO}_K$. Given that the parameter space $\mathcal{A}_K^{\mathrm{cont}}$ is infinite dimensional, it makes sense that the data has to share this property, as suggested also by Proposition 2.13. Together with the fact that access to noise free measurements $y = M_\mu(f_K^\phi)$ is considered, this avoids source (II) and (III) of non-identifiability, and allows sole investigation of source (I). Considered experimental designs will be constructed independently of the a-priori unknown value of K, which means they are persistent exciting w.r.t. ItO_K and $\mathcal{A}_K^{\mathrm{cont}}$ uniformly in K. This is a consequence of transport nature of model (Ch) and the fact that K does not affect the spatial propagation, but only the value of the density in considered locations.

The proof is laid out in the subsequent subsections 4.2 for dimensions $\mathrm{d} \in \{2, 3\}$ and in 4.3 for $d = 1$, where the construction can be simplified due to the discrete velocity space. It relies on a mathematical machinery termed singular decomposition [CS96b, LS20] that was specifically designed to trigger microscopic information in kinetic equations by a detailed construction of an experimental design. This technique exploits the transport nature of kinetic equations using the following key ingredients:

a) The solution is artificially decomposed in terms of regularity.

b) The propagation of induced singular initial data can then be traced in the more singular terms in this decomposition.

c) Compatible measurement specifications isolate that term in the decomposition that shall be used for the reconstruction.

d) The assumption that tumbling is rare underlines the transport nature of (Ch) and explains why higher regularity terms are negligible.

Intuitively, this methods exploits the fact that the kinetic model describes the distribution of an individual based Markov jump model: the singular initial condition can be regarded as a particle approximation to the initial density with just one particle. The propagation with the kinetic equation describes the motion of this one particle in a statistical manner. A compatible measurement procedure filters out the information on those parts of the solution that correspond to one specific tumbling event and thus allows the direct read off of the tumbling kernel - encoding the probability of this tumbling event - from the measurement.

Theorem 4.1 is rigorously proven by an explicit construction of the previously described experiment through specification of ϕ, μ. The rare tumbling assumption (d) is implemented via a small measurement time, defined by the temporal component of μ, that leaves no time for more than one tumbling. It has to be balanced with the singularity in the initial data to avoid that the measurement only captures the initial data. Then the experimental geometry enforces that all measured particles must have undergone exactly one tumbling at a prescribed location and with prescribed outgoing direction. This restores

singularity in velocity in the measurement, as previously destroyed by velocity averaging in the measurement.

The small time requirement is in line with the kernel reconstruction for the fragmentation equation [DET24] and the use of high frequency data in the radiative transfer equation kernel reconstruction from velocity averaged data in [BJLM11, BM12].

The results presented in this chapter emerge from joint work with Christian Klingenberg, Qin Li and Min Tang that was published in [HKLT24].

Overview over related literature.

We give a brief overview over literature regarding singular decomposition and structural identifiability for transport equations, before presenting the novelties of this work.

Singular Decomposition. Singular decomposition is conceptually related to van Neumann decomposition of solutions to kinetic PDEs [CZ63] and the Born series expansion for the wave equation [Wei63]. Driven by applications in optical tomography, it was traditionally used for inverse problems in radiative transport or photon transport through tissue, with boundary data. In the following, we give a brief overview over existing literature. A comprehensive review can be found in [Bal09].
Angularly resolved data: In 1996, the technique was originally developed to study existence and uniqueness of the reconstruction of the absorption coefficient and the scattering kernel in the radiative transfer equation from outgoing boundary data [CS96b, CS99, CS96a]. In the stationary case [CS96b, CS99], results showed non-uniqueness, as the coefficients are determined only up to a specific gauge transform [CS96b, ST09]. Uniqueness can be restored in sufficiently high dimensions by prescribing simplified dependencies of the parameters on the variables, e.g. symmetry of the absorption rate in v. The technique was extended to provide stability results in the cases where uniqueness holds [SU03, BJ08], or otherwise under consideration of the gauge transformation equivalence classes [MST10, BJ18]. Access to time-dependent measurements improves these results: they additionally hold for a lower spatial dimension [CS96a, BJ10].

Non-linear models were considered in [LS20] by introducing a generic, potentially non-linear function of the solution to the RHS of a stationary transport equation. This result was applied to provide uniqueness of velocity independent absorption coefficient and the scattering coefficient in the linear case of stationary radiative transport. A quadratically non-linear absorption was reconstructed in an iterative procedure in [SZ22]. Another example is provided in [LUY21, LO23], where uniqueness of the collision kernel reconstruction from boundary measurements for the stationary [LUY21] or time-dependent [LO23] Boltzmann equation is shown through linearization and application of singular decomposition under certain structural assumptions on the kernel.

Angularly Averaged Data. In certain applications, angularly resolved measurements are not available. This inspired the consideration of angularly averaged data on the outgoing boundary for the inversion, mostly the outgoing flux. These macroscopic type measurements reduce the dimensionality of the data and hinder the selection of the desired decom-

position part, which introduces difficulties to the reconstruction [Bal09]. In the stationary setting and with a simplified variable dependence of the parameters, isotropic sources only allowed regularized reconstructions of the parameter up to a certain error [BLM08], whereas uniqueness and stability of these parameters can be shown if the induced singular source is angularly resolved [Lan08]. Under isotropic sources, the situation is also improved, if time-dependent measurements are available [BJ09, BJ10], and uniqueness and stability of integral forms of the absorption σ and scattering kernel K can be shown. A similar effect was observed in the stationary case when interior domain data was available [BJJ10].

For radiative transport in the frequency domain, the asymptotic behaviour of the singular decomposition parts of angularly averaged boundary data is studied in [BJLM11, BM12], which suggests better identifiability of the coefficients σ, K for high frequencies and translates to small times in the time domain.

Multi-scale behaviour: Furthermore, the technique was used to study the multi-scale behaviour of the inversion: the instability introduced by a diffusive scaling was studied through tracing it in the stability estimates for the reconstruction of the absorption coefficient and the scattering cross-section in the stationary radiative transport equation [LLU19] and explicit, scaling parameter dependent reconstruction formulas were given in [CLL20].

Structural Identifiability for transport equations. Available literature on structural identifiability analysis focuses on inverse problems related to advection equations with fixed velocities. Driven by concrete applications, the spatial variable is chosen one dimensional and often attains a different interpretation, such as the concentration of a catalyst for chemical reaction dynamics [OH98], or an additional age structure of biological agents that undergo some reproduction process [RKE22, GQMT22, CRY19, PLCT11, PR16]. Frequently, multi-species systems of transport equations (potentially coupled with ODEs [GQMT22, OH98]) are considered [PLCT11, VC99, RKE22, CRY19]. Structural identifiability of parameters such as the transport speed [VC99, OH98], the decay [PR16, VC99, RKE22, CRY19] or reproduction rate [CRY19], or the reaction parameter that model the changes between the species [PLCT11, VC99, RKE22] is studied either by a leveraging explicit solution formulas [PR16], by a generalized differential algebra approach [PLCT11, RKE22, CRY19] or a Laplace expansion [VC99, OH98].

Another branch of literature considers the question from a more classical inverse problem perspective and studies identifiability in the sense of well-posedness Definition 2.1i). In [KP08], the kinetic model (Ch) is considered on a bounded domain and the reconstruction of the absorption coefficient from observation of the time evolution of boundary data is studied. (Structural) identifiability is shown under mild assumptions on the initial data and symmetry assumptions on the absorption coefficient through Carleman estimates.

Novelty of this work.

The framework of identifiability analysis is applied to a kinetic model for the first time. Compared to the transport models in identifiability literature, our model (Ch) couples the bacteria densities $f(t,x,v)$ for all velocities $v \in V$ through the term $\mathcal{L}(f)$. For $d \geq 2$, the model does not deteriorate to a multi-species model and thus falls out of the realm of available results from identifiability analysis [PLCT11, VC99, RKE22]. For $d = 1$, the consideration of averaged measurements prohibits the use of the differential approach as in [RKE22, PLCT11]. This work shares some properties with techniques used in [VC99]: it also makes use of a series representation of the measurement, in response to a singular initial condition, in order to attain insight into identifiability. The way to obtain this series representation, however, is fundamentally different: In [VC99], the PDE is solved by a Laplace transform. For the 1D chemotaxis system, the more complex structure of the equation, with space and velocity dependent coupling coefficients $K(x,v,v')$ and transport in both equations, does not permit this simple solving strategy and we have to rely on the decomposition according to regularity. This work thus uses the singular decomposition technique that exploits the characteristic properties of kinetic models and can thus be regarded as a model specific method. The classification of the results obtained by singular decomposition in the identifiability framework has not been considered in previous literature.

Furthermore, angularly averaged measurements are considered, which reduces the information in the data and worsens parameter identifiability [Lan08, BLM08, Bal09, CLW18a, ZZ19]. In this work, this is balanced against the improvement through access to interior measurements [BJJ10] with time dependence [BJ10], generated by an angularly resolved induced singularity [Lan08]. The combination of all these remedies has not been studied in literature, where previous results only included one improvement mechanism. As a consequence, available results can only guarantee identifiability of the parameters with simplified dependence on the variables [BJJ10, BJ10, Lan08], whereas similar restrictions are not required in our result. This is important in the setting of chemotaxis, where different chemoattractant concentrations generate different tumbling behaviour and thus different shapes of K.

To provide this detailed picture of K, an additional mechanism is introduced to control the contribution of the most regular parts: small measurement time reflects the assumption that tumbling events are rare. It has been shown in [DET24], that the kernel reconstruction for the fragmentation equation benefited greatly from the use of short time data. In a radiative transport setting, this was also observed in the frequency domain for high frequency, velocity averaged boundary flux measurements [BJLM11, BM12]. In this work interior domain data is available, which improves the previous reconstruction results in the sense that K can be directly read off the measurement. To achieve this, our result requires a delicate balance between the short measurement time and the singularity in the initial data, as described earlier.

The described setting makes sense for the chemotaxis inverse problem: especially for $d \in \{1,2\}$, the bacterial motion in the whole interior domain can be observed from above. A relatively low bacterial travel speed allows time dependent measurements, and experi-

mental techniques have been developed to control bacteria or micro-swimmers, as described in Remark 3.7.

4.2. Proof of Theorem 4.1 in Dimension $d = 2, 3$

In dimension $d \geq 2$, the velocity space V is continuous, and the tumbling kernel $K(x, v, v')$ lives on a $d(d-1)^2$ dimensional manifold. Its reconstruction requires a carefully crafted experimental design, as the following formal motivation shows.

Formal Intuition. A formal consideration demonstrates required properties of the design: smallness of measurement time, singular measurement and initial data and an adapted experimental geometry. Subsequently, an adaptation of the intuition in [DET24] to the chemotaxis model (Ch) is given. Assume a small measurement time $t_\mathrm{m} > 0$ and sufficient regularity of f. An evaluation of (Ch) at point $(t-s, x-vs, v)$ gives a formula for

$$-\partial_s f(t-s, x-vs, v) = -\sigma(x-vs, v) f(t-s, x-vs, v) + \mathcal{L}(f)(t-s, x-vs, v),$$

also referred to as integration along characteristics. A twofold application leads to the approximation

$$\begin{aligned} f(t_\mathrm{m}, x, v) \approx f(0, x-vt_\mathrm{m}, v) - \int_0^{t_\mathrm{m}} \sigma(x-vs, v)\,\mathrm{d}s\; f(0, x-vt_\mathrm{m}, v) \\ + \int_0^{t_\mathrm{m}} \int_V K(x-vs, v, v') f(0, x-vs-v'(t_\mathrm{m}-s), v')\,\mathrm{d}v'\,\mathrm{d}s \quad + o(t_\mathrm{m}), \end{aligned} \tag{4.3}$$

where the usual Landau notation is used. Now consider singular initial data $f(t = 0, x, v) = \delta_{x_\mathrm{in}}(x)\delta_{v_\mathrm{in}}(v)$ that is given in terms of Dirac delta functions $\delta_{x_\mathrm{in}}, \delta_{v_\mathrm{in}}$ which concentrate at some initial location $x_\mathrm{in} \in \mathbb{R}^d$ and velocity $v_\mathrm{in} \in V$. Assuming that the remainder term is still of order $o(t_\mathrm{m})$, this provides

$$\begin{aligned} \int_V f(t_\mathrm{m}, x, v)\,\mathrm{d}v \approx \delta_{x_\mathrm{in}}(x - v_\mathrm{in}t_\mathrm{m}) \left(1 - \int_0^{t_\mathrm{m}} \sigma(x - v_\mathrm{in}s, v_\mathrm{in})\,\mathrm{d}s\right) + o(t_\mathrm{m}) \\ + \int_V \int_0^{t_\mathrm{m}} K(x_\mathrm{in} + v_\mathrm{in}(t_\mathrm{m}-s), v, v_\mathrm{in})\; \delta_{x_\mathrm{in}}(x - vs - v_\mathrm{in}(t_\mathrm{m}-s))\,\mathrm{d}s\,\mathrm{d}v. \end{aligned} \tag{4.4}$$

It will turn out that the Dirac delta in the integral can be resolved, and for a suitable choice of the measurement location x_m with $\|x_\mathrm{m} - x_\mathrm{in}\| < t_\mathrm{m}$, one finds a unique tumbling time $t_\mathrm{tumb} \in (0, t_\mathrm{m})$ and a unique outgoing direction $v_\mathrm{out} \in V \backslash \{v_\mathrm{in}\}$ with $x_\mathrm{m} = x_\mathrm{in} + v_\mathrm{in}t_\mathrm{tumb} + v_\mathrm{out}(t_\mathrm{m} - t_\mathrm{tumb})$ such that

$$\int_V f(t_\mathrm{m}, x_\mathrm{m}, v)\,\mathrm{d}v \approx K(x_\mathrm{in} + v_\mathrm{in}t_\mathrm{tumb}, v_\mathrm{out}, v_\mathrm{in}) + o(t_\mathrm{m}).$$

This shows that a point-measurement of $\int_V f\,\mathrm{d}v$, taken at small measurement time t_m at a specifically chosen measurement location x_m, approximately reflects the value of the

tumbling kernel at the so-called tumbling location $x_{\text{tumb}} := x_{\text{in}} + v_{\text{in}} t_{\text{tumb}}$, with incoming velocity v_{in} and outgoing velocity v_{out}. Only information of this one precise travel path is contained in the measurement.

The goal of the remaining section is to prove this intuition rigorously.

Remark 4.2. Resolving the Dirac delta in the integral in (4.4) can be understood geometrically: For particles that have tumbled exactly once, as the form of the integral suggests, the tumbling location x_{tumb} and time t_{tumb}, and velocity v_{out} after the tumble are uniquely determined by their initial location x_{in} and velocity v_{in} and the measurement time t_{m} and location x_{m} with $\|x_{\text{in}} - x_{\text{m}}\| < t_{\text{m}}$, as Figure 4.1 illustrates. The point-wise measurement of propagated singular initial data thus only contains information of the tumbling kernel at one location, namely $(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}})$. This allows the point-wise reconstruction.

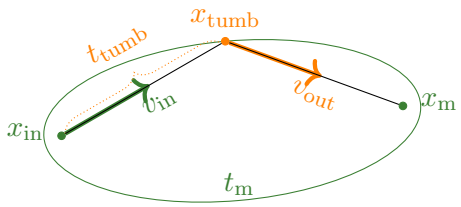


Figure 4.1.: The ellipse with given focal points $x_{\text{in}}, x_{\text{m}}$ and fixed radius t_{m} determines all points x with distance $\|x - x_{\text{in}}\| + \|x - x_{\text{m}}\| = t_{\text{m}}$. These are all points where particles that have started in x_{in} and tumbled exactly once before they reach x_{m} at time t_{m} could potentially have tumbled. As v_{in} is given, the tumbling point x_{tumb} is the unique intersection of the half line, starting at x_{in} in direction v_{in}, with this ellipse. Due to the constant speed of motion, the tumbling time t_{tumb} is determined by the length of the line from x_{in} to x_{tumb} and v_{out} by the direction of the line from x_{tumb} to x_{m}. Color olive green stands for prescribed quantities that uniquely determine orange quantities.

Construction of Initial Data and Measurement Test Function. Explicit design choices are laid out in this subsection. Let $(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}}) \in \mathbb{R}^d \times W$ be the point at which the value $K(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}})$ shall be reconstructed.

Geometry. Let the measurement time

$$t_{\text{m}} := \varepsilon^{\alpha}, \qquad \text{for an } \alpha \in \left(\tfrac{3}{4}, 1\right) \text{ and a scaling parameter } 0 < \varepsilon < T^{\frac{1}{\alpha}}, \tag{4.5}$$

become small as $\varepsilon \to 0$. Then set the initial and measurement location as in Figure 4.2 to

$$x_{\text{in}} := x_{\text{in}}(t_{\text{m}}) := x_{\text{tumb}} - v_{\text{in}} \frac{t_{\text{m}}}{2} \qquad \text{and} \qquad x_{\text{m}} := x_{\text{m}}(t_{\text{m}}) := x_{\text{tumb}} + v_{\text{out}} \frac{t_{\text{m}}}{2}. \tag{4.6}$$

Both are pushed towards the tumbling location x_{tumb} as measurement time t_{m} shrinks with ε. It is important to note that the geometry in Figure 4.2 is sustained, while the experimental domain shrinks.

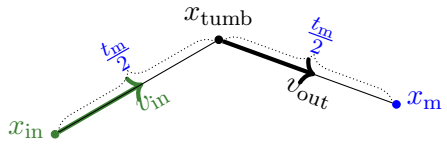


Figure 4.2.: Geometry of the experiment. The quantities t_{m} and x_{m} that belong to the measurement are depicted in blue and v_{in} and x_{in} that describe the initial data in Olive-Green.

Singularity in initial data and measurement. As Dirac delta initial conditions are not covered by the existence theory for the forward equation in Proposition 3.3, mollifications are constructed subsequently. The previously defined geometry describes the concentration points of initial data $\phi^\varepsilon \in C_c^\infty(\mathbb{R}^d \times V)$ and measurement test functions $\mu^\varepsilon \in C_c^\infty(\mathbb{R}^d \times [0,T])$ defined as

$$\phi^\varepsilon(x,v) = \begin{cases} \frac{1}{\varepsilon^d \varepsilon^{\beta(d-1)}} \phi_x\left(\frac{x-x_{\text{in}}}{\varepsilon}\right) \phi_v\left(\frac{\mathbb{P}_{-v_{\text{in}}}(v)}{\varepsilon^\beta}\right) j_{\mathbb{P}_{-v_{\text{in}}}}(v), & x \in \mathbb{R}^d, v \in V \backslash \{-v_{\text{in}}\} \\ 0, & x \in \mathbb{R}^d, v = -v_{\text{in}} \end{cases} \tag{4.7}$$

$$\mu^\varepsilon(t,x) = \left(\frac{t_{\text{m}}}{2}\right)^{d-1} \frac{1}{\varepsilon^{\beta d} \varepsilon^\beta} \mu_x\left(\frac{x-x_{\text{m}}}{\varepsilon^\beta}\right) \mu_t\left(\frac{t-t_{\text{m}}}{\varepsilon^\beta}\right) C_{v_{\text{in}},v_{\text{out}}} \tag{4.8}$$

for the same scaling parameters $\varepsilon > 0$ as used in (4.5) for the measurement time, and some rate $\beta > 1$. The test functions $\phi_x, \mu_x \in \Xi^d$, $\phi_v \in \Xi^{d-1}$ and $\mu_t \in \Xi^1$ shall belong to the spaces

$$\Xi^n = \{h \in C_c^\infty(\mathbb{R}^n; \mathbb{R}) \mid \operatorname{supp}(h) \subset B^n(0,1), \quad 0 \le h \le 1, \quad h(0) = 1 \text{ and } \int_{\mathbb{R}^n} h(s)\,\mathrm{d}s = 1\} \tag{4.9}$$

of non-negative smooth functions on $\mathbb{R}^n$ that are compactly supported in the unit balls $B^n(0,1) \subset \mathbb{R}^n$ of dimension $n \in \{d, d-1, 1\}$, bounded by 1, and attain this value at 0, and integrate to 1. Furthermore, the constant $C_{v_{\text{in}},v_{\text{out}}}$ is defined as

$$C_{v_{\text{in}},v_{\text{out}}} := (1 - \langle v_{\text{in}}, v_{\text{out}} \rangle), \tag{4.10}$$

and $\mathbb{P}_{-v_{\text{in}}} : \mathbb{S}^{d-1} \backslash \{-v_{\text{in}}\} \to \mathbb{R}^{d-1}$ denotes the stereographic projection on the sphere with direction $-v_{\text{in}}$ and $j_{\mathbb{P}_{-v_{\text{in}}}}(v) = (1 + \langle v, v_{\text{in}} \rangle)^{-(d-1)}$ its absolute Jacobi determinant. The projection is described in more detail in Appendix B.1.

Note that the factor $(t_{\text{m}}/2)^{d-1}$ in μ^ε vanishes asymptotically as $\varepsilon \to 0$. This is necessary to balance the singularities generated by the test functions as $\varepsilon \to 0$, in the sense that it keeps the value of the measurement finite, as will become more obvious in Section 4.2.2.

Measurement time vanishes at a slower rate $\alpha \in (\frac{3}{4}, 1)$ than singularities in initial data and measurement are generated. This means that singularity is produced before measurement time and with it the experimental domain is shrunk. As a consequence, the experiment does not deteriorate to simply 'measuring the initial data' and the experimental geometry is sustained in the limit.

Remark 4.3 (Comparison to [HKLT24]). In this work, only one scaling parameter is used with different rates, opposed to the setting in [HKLT24] with chosen, where independent scaling parameters $\nu, \eta, \delta, \varepsilon$ for measurement location and time, initial velocity and location, respectively, were used. We believe this to be closer to practical settings, where singularity is never fully achieved in experiments.

The different rates of the scaling parameter ε determine the relative speed of the generation of singularities and smallness of measurement time and this order coincides with with the one used in [HKLT24], where the singularity limit in initial velocity, measurement time and location preceded the coupled singularity in initial location and smallness of measurement time limit. This order and coupling were crucial ingredients of the proof in [HKLT24].

Mixed settings are also possible, where the singularities in measurement location or time, or initial velocity are generated by independent scalings ν, η, δ, as long as the singularity in these quantities is generated before the singularity in initial location and smallness of measurement time. Additionally, individual rates $\varepsilon^{\beta_1}, \varepsilon^{\beta_2}, \varepsilon^{\beta_3}$ for $\beta_i > 1$ for these singularities could be considered.

The following lemma collects properties of initial data and measurement test function that will be used multiple times throughout the proof of Theorem 4.1.

Lemma 4.4. *Let $\varepsilon > 0$ be sufficiently small.*

a) The functions $\phi^\varepsilon, \mu^\varepsilon$ as constructed above in (4.7)–(4.8) are non-negative, smooth and compactly supported, to be precise

- *$\phi^\varepsilon \in C_c^\infty(\mathbb{R}^d \times V)$ with $\operatorname{supp} \phi^\varepsilon \subset B^d(x_{\mathrm{in}}, \varepsilon) \times (V \cap B^d(v_{\mathrm{in}}, 2\varepsilon^\beta))$, and*
- *$\mu^\varepsilon \in C_c^\infty([0,T] \times \mathbb{R}^d)$ with $\operatorname{supp} \mu^\varepsilon \subset B^1(t_{\mathrm{m}}, \varepsilon^\beta) \times B^d(x_{\mathrm{m}}, \varepsilon^\beta)$.*

b) A change of variables makes the concentration effect on functions $h : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ without v dependence and $g : [0,T] \times \mathbb{R}^d \times V \to \mathbb{R}$ with v dependence obvious:

$$\int_0^T \int_{\mathbb{R}^d} h(t,x)\mu^\varepsilon(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_{-1}^1 \int_{B^d(0,1)} \tfrac{C_{v_{\mathrm{in}},v_{\mathrm{out}}} t_{\mathrm{m}}^{d-1}}{2^{d-1}} h(t_{\mathrm{m}} + \varepsilon^\beta \tilde{t}, x_{\mathrm{m}} + \varepsilon^\beta \tilde{x}) \mu_x(\tilde{x}) \mu_t(\tilde{t})\,\mathrm{d}x\,\mathrm{d}t,$$

$$\int_0^T \int_{\mathbb{R}^d} \int_V g(t,x,v) \frac{1}{\varepsilon^{\beta(d-1)}} \phi_v\left(\tfrac{\mathbb{P}_{-v_{\mathrm{in}}}(v)}{\varepsilon^\beta}\right) j_{\mathbb{P}_{-v_{\mathrm{in}}}}(v)\,\mathrm{d}v\ \mu^\varepsilon(t,x)\,\mathrm{d}x\,\mathrm{d}t$$

$$= \int_{-1}^1 \int_{B^d(0,1)} \int_{B^{d-1}(0,1)} \tfrac{C_{v_{\mathrm{in}},v_{\mathrm{out}}} t_{\mathrm{m}}^{d-1}}{2^{d-1}} g\big(t_{\mathrm{m}} + \varepsilon^\beta \tilde{t}, x_{\mathrm{m}} + \varepsilon^\beta \tilde{x}, \mathbb{P}_{-v_{\mathrm{in}}}^{-1}(\varepsilon^\beta y)\big) \phi_x(y)\,\mathrm{d}y\ \mu_x(\tilde{x})\mu_t(\tilde{t})\,\mathrm{d}x\,\mathrm{d}t.$$

Proof. a) Non-negativity, smoothness and the support condition for μ^ε directly follow from the definition. Non-negativity as well as the support of ϕ^ε in x is also clear. To specify the support in v, observe that $\phi_v\left(\mathbb{P}_{-v_{\mathrm{in}}}(v)/\varepsilon^\beta\right) \neq 0$ is only possible if $\varepsilon^{2\beta} \geq \|\mathbb{P}_{-v_{\mathrm{in}}}(v)\|^2 = (1 + \langle v, -v_{\mathrm{in}}\rangle)/(1 - \langle v, -v_{\mathrm{in}}\rangle)$ by Lemma B.1.1a, which is equivalent to $\langle v, v_{\mathrm{in}}\rangle \geq (1 - \varepsilon^{2\beta})/(1 + \varepsilon^{2\beta})$. For such v, one has in particular $\|v - v_{\mathrm{in}}\|^2 = 2(1 - \langle v, v_{\mathrm{in}}\rangle) \leq 4\varepsilon^{2\beta}/(1 + \varepsilon^{2\beta}) \leq 4\varepsilon^{2\beta}$. Smoothness in v follows from smoothness of ϕ_v and

$\mathbb{P}_{-v_{\text{in}}}$ in $V\backslash\{-v_{\text{in}}\}$ and the fact that a neighbourhood of $-v_{\text{in}}$ is excluded from the support of ϕ^ε by the above considerations for sufficiently small ε.

b) The form of the integrand after a change of variables to $\tilde{t} = \frac{t-t_{\text{m}}}{\varepsilon^\beta}$, $y = \frac{\mathbb{P}_{-v_{\text{in}}}(v)}{\varepsilon^\beta}$ and $\tilde{x} = \frac{x-x_{\text{m}}}{\varepsilon}$ is clear, given that $\mathbb{P}_{-v_{\text{in}}}$ is bijective on its support according to Lemma B.1.1a and $-v_{\text{in}} \notin \operatorname{supp}\phi^\varepsilon$ for small enough ε. It remains to verify the form of the integration domain. These can be reduced in consideration of the supports $\operatorname{supp}\mu_x \subset B^d(0,1)$, $\operatorname{supp}\phi_v \subset B^{d-1}(0,1)$, and $\operatorname{supp}\mu_t \subset [-1,1]$, that are contained in the domains after the change of variables $\tilde{x} \in \mathbb{R}^d, y \in \mathbb{R}^{d-1}, \tilde{t} \in [-t_{\text{m}}/\varepsilon^\beta, (T-t_{\text{m}})/\varepsilon^\beta]$ for sufficiently small ε, by the fact that $t_{\text{m}}/\varepsilon^\beta = \varepsilon^{\alpha-\beta} \to \infty$ as $\varepsilon \to 0$ due to $\alpha < 1 < \beta$ and $T > 0$ fixed by construction.

□

Decomposition. In the following, the dependence of f on its initial data ϕ will be expressed explicitly by writing $f = f^\phi$. It is decomposed into three parts

$$f^\phi = f_0^\phi + f_1^\phi + f_{\geq 2}^\phi. \tag{4.11}$$

The most singular part is the ballistic part f_0^ϕ of f^ϕ that collects the bacteria that did not tumble up to time t. Starting to count the tumbling at time $t = 0$, all particles of initial condition ϕ fully belong to f_0^ϕ. The run-and-tumble behaviour from section 3.1 suggests that they run along a straight line in direction v until they tumble for the first time and statistically disappear from f_0^ϕ. In summary, f_0^ϕ solves the following transport problem with absorption

$$\begin{aligned} \partial_t f_0^\phi + v \cdot \nabla f_0^\phi &= -\sigma f_0^\phi, \\ f_0^\phi(t=0,x,v) &= \phi. \end{aligned} \tag{T_0}$$

The particles that tumbled for the first time reappear as a source term $\mathcal{L}(f_0^\phi)$ for f_1^ϕ, the fraction of particles of f^ϕ that tumbled exactly once up to time t. A statistical loss is constituted by particles that tumbled for a second time, and the propagation of f_1^ϕ can be described by the following transport equation

$$\begin{aligned} \partial_t f_1^\phi + v \cdot \nabla f_1^\phi &= -\sigma f_1^\phi + \mathcal{L}(f_0^\phi), \\ f_1^\phi(t=0,x,v) &= 0. \end{aligned} \tag{T_1}$$

All remaining particles, i.e. those that tumbled twice or more often, are collected in the most regular part $f_{\geq 2}^\phi$. It solves (Ch) with zero initial condition and an additional source term to account for the particles that were formerly in f_1^ϕ and then tumbled for a second time

$$\begin{aligned} \partial_t f_{\geq 2}^\phi + v \cdot \nabla f_{\geq 2}^\phi &= -\sigma f_{\geq 2}^\phi + \mathcal{L}(f_1^\phi + f_{\geq 2}^\phi), \\ f_{\geq 2}^\phi(t=0,x,v) &= 0. \end{aligned} \tag{$\text{T}_{\geq 2}$}$$

Existence of solutions for (T_0), (T_1) and ($\text{T}_{\geq 2}$) is assured by standard semigroup theory, as laid out in Corollary A.1.3.

Lemma 4.5. *Let $\phi \in L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V)$ be non-negative for a.a. v and let $K \in \mathcal{A}_K$. Then (T_0), (T_1) and ($T_{\geq 2}$) attain unique mild solutions f_0^ϕ, f_1^ϕ and $f_{\geq 2}^\phi$ in the space $C^0([0,T]; L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V))$ that are non-negative in a.a. v. Explicit formulas for f_0^ϕ and f_1^ϕ exist and read*

$$f_0^\phi(t,x,v) = e^{-\int_0^t \sigma(x-vs,v)\,\mathrm{d}s}\phi(x-vt,v), \tag{4.12}$$

$$f_1^\phi(t,x,v) = \int_0^t \int_V e^{-\int_0^s \sigma(x-v\tau,v)\,\mathrm{d}\tau} K(x-vs,v,v') e^{-\int_0^{t-s} \sigma(x-vs-v'\tau,v')\,\mathrm{d}\tau} \cdot \phi(x-vs-v'(t-s),v')\,\mathrm{d}v'\,\mathrm{d}s\,, \tag{4.13}$$

and $f_{\geq 2}^\phi$ is bounded by its source term f_1^ϕ through

$$\|f_{\geq 2}^\phi(t)\|_\infty \leq e^{|V|C_K t} C_K |V| \int_0^t \|f_1^\phi(s)\|_\infty \,\mathrm{d}s. \tag{4.14}$$

Proof. Existence and positivity for f_0^ϕ are direct consequences of Corollary A.1.3 and Lemma A.1.5a. The same results provide the assertion for f_1^ϕ, given that the source term $\mathcal{L}(f_0^\phi) \in C^0([0,T]; L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V)) \subset L^1([0,T]; L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V))$ is non-negative, since $K \in \mathcal{A}_K^{\mathrm{cont}}$ is. Existence and non-negativity of $f_{\geq 2}^\phi$ follow analogously. The bound for its norm $\|f_{\geq 2}(t)\|_\infty$ is provided by (A.1.6), together with the boundedness of operator $\mathcal{L}$ by $\|\mathcal{L}\| \leq C_K|V|$. The explicit formulas for f_0^ϕ and f_1^ϕ emerge from a combination of (A.1.4) and (A.1.7). □

The information on K contained in the three parts has different quality: f_0^ϕ can only contain information on the integrated version $\sigma(x,v) = \int_V K(x,v',v)\,\mathrm{d}v'$ of K as suggested by (T_0). On the other hand, particles from $f_{\geq 2}^\phi$ were subject to two or more tumbles, whose influence cannot be separated. Both quantities do not seem suitable for a reconstruction of K, which will instead be based on f_1^ϕ that contains information on exactly one tumbling event.

Because tumbling has a diffusive effect, the three parts f_0^ϕ, f_1^ϕ and $f_{\geq 2}^\phi$ attain different regularity. This smoothing effect becomes obvious with singular initial data and measurement test functions. The aim of our construction is to exploit this behaviour so to obtain access to f_1^ϕ through the measurement

$$M_\mu(f^\phi) = M_\mu(f_0^\phi) + M_\mu(f_1^\phi) + M_\mu(f_{\geq 2}^\phi) \approx M_\mu(f_1^\phi), \qquad M_\mu(f_0^\phi) \approx 0 \approx M_\mu(f_{\geq 2}^\phi). \tag{4.15}$$

Since actual singularity is not realizable, neither theoretically by the smoothness assumptions on ϕ, μ, nor experimentally due to restrictions in measurement precision and preparation of initial data, the above will hold in the limit as initial data and measurement test function become more and more concentrated.

Behaviour of the Measurement.

Theorem 4.1 is a direct consequence of the following proposition.

Proposition 4.6. *Let $K \in \mathcal{A}_K^{\text{cont}}$ and consider a fixed point $(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}}) \in \mathbb{R}^d \times W$. Let $(\phi^\varepsilon, \mu^\varepsilon)$ be defined as in (4.7)-(4.8) with $x_{\text{in}}, x_{\text{m}}, t_{\text{m}}$ given in (4.5)–(4.6) for $\varepsilon > 0$. Then*

$$\lim_{\varepsilon \to 0} M_{\mu^\varepsilon}(f^{\phi^\varepsilon}) = K(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}})\,.$$

This proposition describes an explicit relationship between the (sequence of) measurement(s) and the parameter that shall be reconstructed. Its proof is divided into the subsequent three lemmas, each of which treats the limit for one of the parts of the decomposition of $M_{\mu^\varepsilon}(f^{\phi^\varepsilon})$ in (4.15).

Particles from f_0^ϕ travel along a straight line, so singular initial data will be propagated to the ballistic location

$$x_b := x_{\text{in}} + v_{\text{in}} t_{\text{m}} \neq x_{\text{m}} \tag{4.16}$$

at time t_{m} and will thus not appear in the measurement.

Lemma 4.7. *Let $K \in \mathcal{A}_K^{\text{cont}}$. Given a point $(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}}) \in \mathbb{R}^d \times W$, define $(\phi^\varepsilon, \mu^\varepsilon)$ for $\varepsilon > 0$ as in (4.7)-(4.8) for $x_{\text{in}}, x_{\text{m}}, t_{\text{m}}$ as in (4.5)–(4.6). Then $M_{\mu^\varepsilon}(f_0^{\phi^\varepsilon})$ vanishes in the limit*

$$\lim_{\varepsilon \to 0} M_{\mu^\varepsilon}(f_0^{\phi^\varepsilon}) = 0\,. \tag{4.17}$$

All particles from $f_1^{\phi^\varepsilon}$ in the measurement were subject to the one tumbling at the same location, time and with the same outgoing velocity, according to Figure 4.1. As $t_{\text{m}} \to 0$, this happens instantaneously at the beginning, and the measured particles are only subject to this tumbling, given that no time for further decay while running is left.

Lemma 4.8. *Let $K \in \mathcal{A}_K^{\text{cont}}$, and let a point $(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}}) \in \mathbb{R}^d \times W$ be given and construct initial data ϕ^ε and test functions μ^ε as in (4.7)-(4.8) for $x_{\text{in}}, x_{\text{m}}, t_{\text{m}}$ as in (4.5)–(4.6). Then the measurement $M_{\mu^\varepsilon}(f_1^{\phi^\varepsilon})$ reconstructs $K(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}})$, in the sense that*

$$\lim_{\varepsilon \to 0} M_{\mu^\varepsilon}(f_1^{\phi^\varepsilon}) = K(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}})\,. \tag{4.18}$$

The influence of higher regularity parts from $f_{\geq 2}^{\phi^\varepsilon}$ is controlled by a small time requirement $t_{\text{m}} = \varepsilon^\alpha \to 0$ as $\varepsilon \to 0$. Intuitively, it means that there is no time left for a second tumbling event.

Lemma 4.9. *Let $K \in \mathcal{A}_K^{\text{cont}}$, and define initial data ϕ^ε and test function μ^ε as in (4.7)-(4.8) for a given point $(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}}) \in \mathbb{R}^d \times W$ and $x_{\text{in}}, x_{\text{m}}, t_{\text{m}}$ as in (4.5)–(4.6). Then*

$$\lim_{\varepsilon \to 0} M_{\mu^\varepsilon}(f_{\geq 2}^{\phi^\varepsilon}) = 0. \tag{4.19}$$

The proofs of these lemmas are the subject of the remainder of this chapter. Together, they prove Proposition 4.6.

Proof of Proposition 4.6. Combining the previous lemmas, in the limit, equation (4.15) is obtained

$$\lim_{\varepsilon \to 0} M_{\mu^\varepsilon}(f^{\phi^\varepsilon}) = \lim_{\varepsilon \to 0} M_{\mu^\varepsilon}(f_1^{\phi^\varepsilon}) = K(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}}).$$

□

In the following subsections, the dependence of f on the initial data and the M on the measurement test function will be omitted, since it is clear that $\phi^\varepsilon, \mu^\varepsilon$ in (4.7)–(4.8) are considered.

Remark 4.10. Temporal changes in the chemoattractant might induce time dependence in K. Our strategy can be extended to this case, with some additional effort in the experimentation: All arguments in the proofs of the three lemmas are local-in-time and the setting $t_\mathrm{m} = \varepsilon^\alpha \to 0$ is considered. They can hence readily be extended to reconstruct time continuous $K(t, x, v, v')$, if the initial condition and measurement can be prepared around any time $\hat{t} > 0$. This means that the experiment starts delayed at time $\hat{t}$, and the same reconstruction strategy can be used.

4.2.1. The Ballistic Part f_0

The proof of Lemma 4.7 is based on a comparison of the spatial supports of measurement test function μ and f_0, given as balls $B(x_\mathrm{m}, \varepsilon^\beta)$ and $B(x_b, \varepsilon)$ around the measurement location x_m and the ballistic location x_b from (4.16). For sufficiently small supports as $\varepsilon \to 0$, there will be no overlap, leading to a vanishing measurement of f_0. The idea is summarized in Figure 4.3.

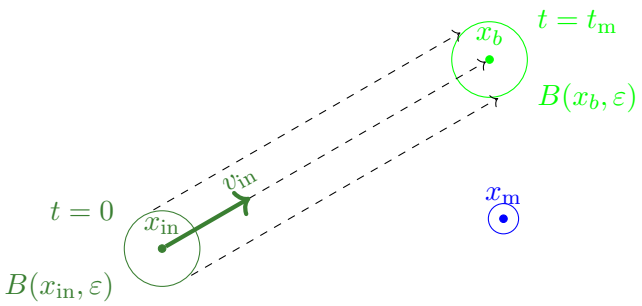


Figure 4.3.: Propagation of the spatial support of f_0^ϕ for initial data $\phi(x, v) = \tilde{\phi}(x)\delta_{v_\mathrm{in}}(v)$ from $B(x_\mathrm{in}, \varepsilon)$ at time 0 (OliveGreen) to $B(x_b, \varepsilon)$ at time t_m (green), and spatial support $B(x_\mathrm{m}, \varepsilon^\beta)$ of measurement test function $\mu = \tilde{\mu}_x \delta_{t_\mathrm{m}}$ (blue) for small ε.

Proof of Lemma 4.7. Using the explicit formula (4.12) for f_0, a change of variables makes the supports and their lack of overlap obvious, which leads to a vanishing measurement.

Step 1: Transformation of variables. Insert the explicit form (4.12) of f_0 in the measurement and conduct a change of variables as in Lemma 4.4b to obtain, for small

enough $\varepsilon > 0$,

$$M(f_0) = \int_0^T \int_{\mathbb{R}^d} \int_V f_0(t,x,v)\ \mathrm{d}v\, \mu^\varepsilon(t,x)\ \mathrm{d}x\ \mathrm{d}t$$
$$= \frac{C_{v_{\mathrm{in}},v_{\mathrm{out}}} t_{\mathrm{m}}^{d-1}}{2^{d-1}\varepsilon^d} \int_{-1}^{1} \int_{B^d(0,1)} \int_{B^{d-1}(0,1)} e^{-\int_0^{t_{\mathrm{m}}+\varepsilon^\beta \tilde{t}} \sigma(x_{\mathrm{m}}+\varepsilon^\beta \tilde{x} - \mathbb{P}^{-1}_{-v_{\mathrm{in}}}(\varepsilon^\beta y)s, \mathbb{P}^{-1}_{-v_{\mathrm{in}}}(\varepsilon^\beta y))\ \mathrm{d}s} \phi_v(y)\, \mu_x(\tilde{x}) \cdot$$
$$\mu_t(\tilde{t})\, \phi_x\left(\frac{x_{\mathrm{m}} + \varepsilon^\beta \tilde{x} - \mathbb{P}^{-1}_{-v_{\mathrm{in}}}(\varepsilon^\beta y)(t_{\mathrm{m}} + \varepsilon^\beta \tilde{t}) - x_{\mathrm{in}}}{\varepsilon}\right)\ \mathrm{d}y\ \mathrm{d}\tilde{x}\ \mathrm{d}\tilde{t}.$$

Step 2: Convergence. The above formula indicates that $M(f_0) = 0$ for small enough ε, because the argument of ϕ_x, denoted by

$$\mathfrak{a} := \frac{x_{\mathrm{m}} + \varepsilon^\beta \tilde{x} - \mathbb{P}^{-1}_{-v_{\mathrm{in}}}(\varepsilon^\beta y)(t_{\mathrm{m}} + \varepsilon^\beta \tilde{t}) - x_{\mathrm{in}}}{\varepsilon},$$

does not belong to its support $B^d(0,1)$. This can be seen, as by the reverse triangle inequality

$$\begin{aligned}\|\mathfrak{a}\| &\geq \varepsilon^{-1}\|x_{\mathrm{m}} - v_{\mathrm{in}} t_{\mathrm{m}} - x_{\mathrm{in}}\| - \tfrac{t_{\mathrm{m}}}{\varepsilon}\|v_{\mathrm{in}} - \mathbb{P}^{-1}_{-v_{\mathrm{in}}}(\varepsilon^\beta y)\| - \varepsilon^{\beta-1}\|\tilde{x}\| - \varepsilon^{\beta-1}\|\mathbb{P}^{-1}_{-v_{\mathrm{in}}}(\varepsilon^\beta y)\||\tilde{t}| \\ &\geq \varepsilon^{\alpha-1}\frac{1}{2}\|v_{\mathrm{out}} - v_{\mathrm{in}}\| - 2\varepsilon^{\beta+\alpha-1} - 2\varepsilon^{\beta-1}. \end{aligned} \tag{4.20}$$

In the second line, the definition of the measurement geometry (4.5)–(4.6) was exploited, together with boundedness of $\|\tilde{x}\|, |\tilde{t}| \leq 1$ and $\mathbb{P}^{-1}_{-v_{\mathrm{in}}}(\varepsilon^\beta y) \in \mathbb{S}^{d-1}$, and the fact that $\mathbb{P}^{-1}_{-v_{\mathrm{in}}}(\varepsilon^\beta y) = v_{\mathrm{in}} + \varepsilon^\beta v^{\varepsilon^\beta}(y)$ with $\|\varepsilon^\beta v^{\varepsilon^\beta}(y)\| \leq 2\varepsilon^\beta \|y\| \leq 2\varepsilon^\beta$ for $y \in B^{d-1}(0,1)$ by Lemma B.1.1b. Given $\beta > 1 > \alpha$ and $v_{\mathrm{in}} \neq v_{\mathrm{out}}$, the first summand becomes arbitrarily large as $\varepsilon \to 0$, whereas the remaining summands vanish. This shows $\|\mathfrak{a}\| > 1$ for ε small enough. □

4.2.2. The One Tumble Part f_1

The decay that the particle from f_1 undergo on their way from x_{in} to x_{m} due to too early tumbling before x_{tumb} or a second tumble after x_{tumb}, see Figure 4.4, vanishes with the measurement time $t_{\mathrm{m}} = \varepsilon^\alpha \to 0$ and $x_{\mathrm{in}}, x_{\mathrm{m}} \to x_{\mathrm{tumb}}$. Only the effect of the one tumbling event remains. The different scaling of the singularity in $\phi^\varepsilon, \mu^\varepsilon$ and the slower rate of decay of measurement time let the argumentation in Figure 4.1 stay valid, such that the increasing singularity in $\phi^\varepsilon, \mu^\varepsilon$ uniquely prescribes the tumbling location x_{tumb} and outgoing velocity v_{out}.

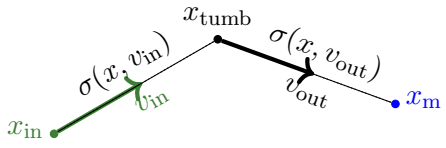


Figure 4.4.: Propagation of particles from f_1 from x_{in} to x_{m}, subject to decay at rate σ.

Proof of Lemma 4.8. The Proof of Lemma 4.8 exploits the availability of an explicit formula (4.13) for f_1. Convergence to the tumbling kernel evaluation is established through the dominated convergence theorem and heavily exploits the construction of initial data ϕ^ε and μ^ε in (4.7)–(4.8) and the experimental geometry (4.5)–(4.6). A transformation of variables is conducted and the integration domains are reduced to ε independent domains, for small enough ε, before the dominated convergence theorem can be applied. Some technical details are sourced out to Lemma 4.11.

Step 1: Transformation of variables and reduction of integration domain. After inserting the explicit formula (4.13) for f_1 and the definition (4.7) of ϕ^ε, the standard transformation of variables from Lemma 4.4b shows that for small enough $\varepsilon > 0$

$$
\begin{aligned}
M(f_1) &= \int_0^T \int_{\mathbb{R}^d} \int_V f_1(t,x,v)\, \mathrm{d}v\, \mu^\varepsilon(t,x)\, \mathrm{d}x\, \mathrm{d}t \\
&= \int_{-1}^{1} \int_{B^d(0,1)} \int_{[0,t_\mathrm{m}+\varepsilon^\beta \tilde{t}]\times V} \int_{B^{d-1}(0,1)} \frac{C_{v_\mathrm{in},v_\mathrm{out}} t_\mathrm{m}^{d-1}}{2^{d-1}\varepsilon^d} e^{-\int_0^s \sigma(x_\mathrm{m}+\varepsilon^\beta\tilde{x}-v\tau,v)\,\mathrm{d}\tau} \cdot \\
&\qquad K(x_\mathrm{m}+\varepsilon^\beta\tilde{x}-vs, v, \mathbb{P}^{-1}_{-v_\mathrm{in}}(\varepsilon^\beta y))\; e^{-\int_0^{t_\mathrm{m}+\varepsilon^\beta\tilde{t}-s} \sigma(x_\mathrm{m}+\varepsilon^\beta\tilde{x}-vs-\mathbb{P}^{-1}_{-v_\mathrm{in}}(\varepsilon^\beta y)\tau, \mathbb{P}^{-1}_{-v_\mathrm{in}}(\varepsilon^\beta y))\,\mathrm{d}\tau} \cdot \\
&\qquad \phi_x\left(\frac{x_\mathrm{m}+\varepsilon^\beta\tilde{x}-vs-\mathbb{P}^{-1}_{-v_\mathrm{in}}(\varepsilon^\beta y)(t_\mathrm{m}+\varepsilon^\beta\tilde{t}-s)-x_\mathrm{in}}{\varepsilon}\right) \phi_v(y)\, \mu_x(\tilde{x})\, \mu_t(\tilde{t})\, \mathrm{d}y\, \mathrm{d}(s,v)\, \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{t}\,.
\end{aligned}
$$

The concentration of ϕ_x as $\varepsilon \to 0$ shall be used to derive a unique tumbling point through specification of (s, v). This requires a change of variables. Write the argument of ϕ_x as

$$
\frac{x_\mathrm{m}+\varepsilon^\beta\tilde{x}-vs-\mathbb{P}^{-1}_{-v_\mathrm{in}}(\varepsilon^\beta y)(t_\mathrm{m}+\varepsilon^\beta\tilde{t}-s)-x_\mathrm{in}}{\varepsilon} = z + R(s),
$$

with $R(s)$ defined as in (4.24) and $z := \frac{\mathcal{T}_a^{v_\mathrm{in}}(s,v)}{\varepsilon}$ given by the transformation

$$
\mathcal{T}_a^{v_\mathrm{in}}(s,v) := a - s(v - v_\mathrm{in}) = x_\mathrm{m} - vs - v_\mathrm{in}(t_\mathrm{m} - s) - x_\mathrm{in} \in \mathbb{R}^d, \tag{4.21}
$$

$$
\text{where} \quad a := \frac{t_\mathrm{m}}{2}(v_\mathrm{out} - v_\mathrm{in}) = x_\mathrm{m} - x_b, \tag{4.22}
$$

where x_b denotes the ballistic location $x_\mathrm{in} + v_\mathrm{in} t_\mathrm{m}$ as in (4.16). A visualization of these quantities can be found in Figure 4.5 and the behaviour of $\mathcal{T}_a^{v_\mathrm{in}}$ is studied in Appendix B.2. Because $\mathcal{T}_a^{v_\mathrm{in}}$ cannot be injective in a, the (s, v) integration domain is reduced to $U := [c_1, t_\mathrm{m} + \varepsilon^\beta \tilde{t}] \times \{v \in V \mid \langle v, v_\mathrm{in}\rangle \le 1 - c_2\}$, with $c_1 := \frac{\|a\|}{4}$ and $c_2 := \frac{\|v_\mathrm{out} - v_\mathrm{in}\|^2}{2^7}$. This is possible, since the integrand vanishes outside U according to Lemma 4.11b. Then Lemma B.2.1a suggests that $\mathcal{T}_a^{v_\mathrm{in}} : U \to \mathcal{T}_a^{v_\mathrm{in}}(U), (s, v) \mapsto a - s(v - v_\mathrm{in})$ is bijective with inverse $(\mathcal{T}_a^{v_\mathrm{in}})^{-1}(z) = (\zeta(z), \omega(z))$ given in (B.2.1) and absolute Jacobi determinant $s^{d-1}(1 - \langle v_\mathrm{in}, v\rangle)$. Changing the variables by the full transformation $(s, v) \mapsto z$, and

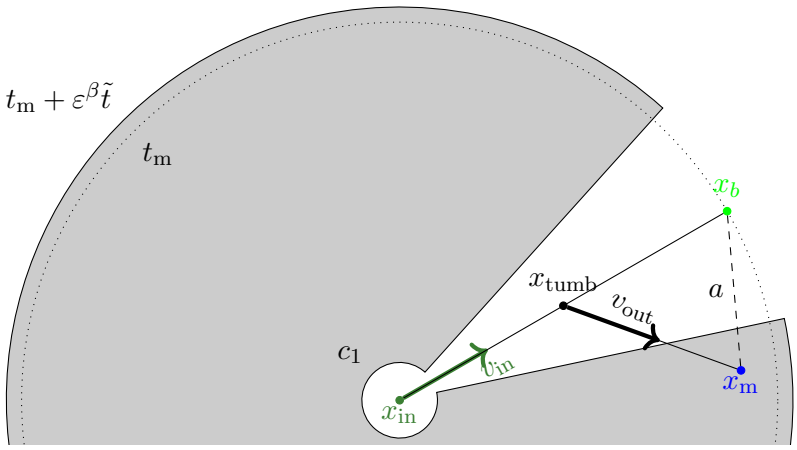


Figure 4.5.: Geometry and quantities used in the proof, displayed in 2D. The gray area depicts the sliced annulus $x_\mathrm{in} + vs$ for $(s, v) \in U$, which coincides with the area A in figure B.2.1a translated by x_in. The same colour coding as in Figure 4.3 is used.

inserting definition (4.10) of $C_{v_\mathrm{in}, v_\mathrm{out}}$, thus gives

$$M(f_1) = \int_{-1}^{1} \int_{B^d(0,1)} \int_{B^d(0,5)} \int_{B^{d-1}(0,1)} \frac{\left(\frac{t_\mathrm{m}}{2}\right)^{d-1}}{\zeta(\varepsilon z)^{d-1}} \frac{1 - \langle v_\mathrm{in}, v_\mathrm{out}\rangle}{1 - \langle v_\mathrm{in}, \omega(\varepsilon z)\rangle} e^{-\int_0^{\zeta(\varepsilon z)} \sigma(x_\mathrm{m} + \varepsilon^\beta \tilde{x} - \omega(\varepsilon z)\tau, \omega(\varepsilon z))\, \mathrm{d}\tau} \cdot$$
$$K(x_\mathrm{m} + \varepsilon^\beta \tilde{x} - \zeta(\varepsilon z)\omega(\varepsilon z), \omega(\varepsilon z), \mathbb{P}^{-1}_{-v_\mathrm{in}}(\varepsilon^\beta y)) \cdot \tag{4.23}$$
$$e^{-\int_0^{t_\mathrm{m} + \varepsilon^\beta \tilde{t} - \zeta(\varepsilon z)} \sigma(x_\mathrm{m} + \varepsilon^\beta \tilde{x} - \zeta(\varepsilon z)\omega(\varepsilon z) - \mathbb{P}^{-1}_{-v_\mathrm{in}}(\varepsilon^\beta y)\tau, \mathbb{P}^{-1}_{-v_\mathrm{in}}(\varepsilon^\beta y))\, \mathrm{d}\tau} \cdot$$
$$\phi_x\left(z + R(\zeta(\varepsilon z))\right) \phi_v\left(y\right) \mu_x\left(\tilde{x}\right) \mu_t\left(\tilde{t}\right)\, \mathrm{d}y\, \mathrm{d}z\, \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{t}\,,$$

where the z integration domain $\frac{\mathcal{T}_a^{v_\mathrm{in}}(U)}{\varepsilon}$ was further reduced to its ε independent subset $B^d(0,5)$, for small enough ε, according to Lemma 4.11c.

Step 2: Convergence. In order to establish point-wise convergence and boundedness of the integrand, first notice that continuity of ω, ζ, their explicit formulas in (B.2.1) and boundedness of $z \in B(0,5)$ suggest that $\omega(\varepsilon z) \to v_\mathrm{out}$ and $\zeta(\varepsilon z) \to 0$, with

$$\frac{\frac{t_\mathrm{m}}{2}}{\zeta(\varepsilon z)} = \frac{\varepsilon^\alpha |\langle a - \varepsilon z, -v_\mathrm{in}\rangle|}{\|a - \varepsilon z\|^2} = \frac{|\left\langle \frac{v_\mathrm{out} - v_\mathrm{in}}{2} - \varepsilon^{1-\alpha} z, -v_\mathrm{in}\right\rangle|}{\|\frac{v_\mathrm{out} - v_\mathrm{in}}{2} - \varepsilon^{1-\alpha} z\|^2} \xrightarrow{\varepsilon \to 0} \frac{|\left\langle \frac{v_\mathrm{out} - v_\mathrm{in}}{2}, -v_\mathrm{in}\right\rangle|}{\|\frac{v_\mathrm{out} - v_\mathrm{in}}{2}\|^2} = 1$$

by the definition (4.22) of a. This provides boundedness and point-wise convergence of the fractions in the first line of (4.23) to 1. Point-wise convergence and boundedness of the remaining factors in the integrand follow from continuity and boundendess of K, ϕ_x and $\mathbb{P}^{-1}_{-v_\mathrm{in}}$ with $\mathbb{P}^{-1}_{-v_\mathrm{in}}(0) = v_\mathrm{in}$ by Lemma B.1.1b and the fact that $R(\zeta(\varepsilon z)) \to 0$ by Lemma 4.11a.

This lays the ground for the application of the dominated convergence theorem

$$\begin{aligned}
&\lim_{\varepsilon\to 0} M_{\mu^\varepsilon}(f_1)\\
&= K(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}}) \int_{-1}^{1} \int_{B^d(0,1)} \int_{B^d(0,5)} \int_{B^{d-1}(0,1)} \phi_x(z)\phi_v(y)\mu_x(\tilde{x})\mu_t(\tilde{t})\, \mathrm{d}y\, \mathrm{d}z\, \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{t}\\
&= K(x_{\text{tumb}}, v_{\text{out}}, v_{\text{in}}),
\end{aligned}$$

where $x_{\text{m}} = x_{\text{tumb}} + \frac{t_{\text{m}}}{2} v_{\text{out}} \to x_{\text{tumb}}$ as $t_{\text{m}} = \varepsilon^\alpha \to 0$ was used, as well as the fact that $\phi_x, \phi_v, \mu_x, \mu_t$ integrate to 1. □

The following lemma collects some technical derivations that were used in Step 2 in the above proof of Lemma 4.8 in the change of variables. The notation from the proof is used.

Lemma 4.11. *Let $t_{\text{m}}, x_{\text{in}}, x_{\text{m}}$ be defined as in (4.5)–(4.6) for fixed $x_{\text{tumb}} \in \mathbb{R}^d$, $v_{\text{in}}, v_{\text{out}} \in V$ with $v_{\text{in}} \neq v_{\text{out}}$ and define $a := \frac{t_{\text{m}}}{2}(v_{\text{out}} - v_{\text{in}})$. Furthermore, let $\tilde{t} \in [-1, 1]$, $\tilde{x} \in B^d(0,1)$ and $y \in B^{d-1}(0,1)$.*

a) Let $\varepsilon < 1$. Then the function $R : [0, t_{\text{m}} + \varepsilon^\beta \tilde{t}] \to \mathbb{R}^d$, with

$$R(s) = \frac{(v_{\text{in}} - \mathbb{P}^{-1}_{-v_{\text{in}}}(\varepsilon^\beta y))(t_{\text{m}} - s) + \varepsilon^\beta \tilde{x} - \varepsilon^\beta \tilde{t} \mathbb{P}^{-1}_{-v_{\text{in}}}(\varepsilon^\beta y)}{\varepsilon} \tag{4.24}$$

is bounded uniformly in s by $\|R(s)\| \leq 4\varepsilon^{\beta-1}$, uniformly in $y, \tilde{x}, \tilde{t}$.

b) For sufficiently small $\varepsilon > 0$, the function

$$\psi : [0, t_{\text{m}} + \varepsilon^\beta \tilde{t}] \times V \to \mathbb{R}, \qquad \psi(s, v) = \phi_x\left(\frac{a - s(v - v_{\text{in}})}{\varepsilon} + R(s)\right)$$

is supported in $U := [c_1, t_{\text{m}} + \varepsilon^\beta \tilde{t}] \times \{v \in V \mid \langle v, v_{\text{in}} \rangle \leq 1 - c_2\}$ for $c_1 := \frac{\|a\|}{4}$ and $c_2 := \frac{\|v_{\text{out}} - v_{\text{in}}\|^2}{2^7}$.

c) Consider the function

$$\hat{\psi} : \varepsilon^{-1} \mathcal{T}_a^{v_{\text{in}}}(U) \to \mathbb{R}, \qquad \hat{\psi}(z) = \phi_x\left(z + R(\zeta(\varepsilon z))\right)$$

where ζ is the s-inverse of $\mathcal{T}_a^{v_{\text{in}}}$ defined by (4.21), compare Lemma B.2.1a). For sufficiently small $\varepsilon > 0$, one has $\operatorname{supp} \hat{\psi} \subset B^d(0,5) \subset \varepsilon^{-1} \mathcal{T}_a^{v_{\text{in}}}(U)$.

Proof.

a) The triangle inequality shows

$$\|R(s)\| \leq \frac{|t_{\text{m}} - s|}{\varepsilon} \|v_{\text{in}} - \mathbb{P}^{-1}_{-v_{\text{in}}}(\varepsilon^\beta y)\| + \varepsilon^{\beta-1} \|\tilde{x}\| + \varepsilon^{\beta-1} |\tilde{t}| \|\mathbb{P}^{-1}_{-v_{\text{in}}}(\varepsilon^\beta y)\| \leq 4\varepsilon^{\beta-1}$$

for $\varepsilon \leq 1$, as $|t_{\text{m}} - s| \leq \max(\varepsilon^\beta, \varepsilon^\alpha) \leq 1$, because $s \in [0, t_{\text{m}} + \varepsilon^\beta \tilde{t}]$ and $\alpha < \beta$, and $\|v_{\text{in}} - \mathbb{P}^{-1}_{-v_{\text{in}}}(\varepsilon^\beta y)\| \leq 2\varepsilon^\beta \|y\| \leq 2\varepsilon^\beta$ according to Lemma B.1.1b as well as $|\tilde{t}|, \|\tilde{x}\| \leq 1 = \|\mathbb{P}^{-1}_{-v_{\text{in}}}(\varepsilon^\beta y)\|$.

b) Let $s < c_1 = \frac{\|a\|}{4}$ and $v \in V$. Then the reverse triangle inequality, together with Lemma 4.11a and $\|v - v_{\text{in}}\| \leq 2$, shows that

$$\left\| \frac{a - s(v - v_{\text{in}})}{\varepsilon} + R(s) \right\| \geq \frac{\|a\| - 2s}{\varepsilon} - 4\varepsilon^{\beta-1}$$
$$> \frac{\|a\|}{2\varepsilon} - 4\varepsilon^{\beta-1} = \varepsilon^{\alpha-1} \frac{\|v_{\text{out}} - v_{\text{in}}\|}{4} - 4\varepsilon^{\beta-1} \qquad > 1\,,$$

for sufficiently small $\varepsilon > 0$ by choice of $\alpha < 1 < \beta$ and $v_{\text{in}} \neq v_{\text{out}}$ and the definition (4.22) of a.

Similarly, one has for $v \in V$ with $\langle v, v_{\text{in}} \rangle > 1 - c_2 = 1 - \frac{\|v_{\text{out}} - v_{\text{in}}\|^2}{2^7}$ and $s \in [0, t_{\text{m}} + \varepsilon^\beta \tilde{t}]$ that

$$\left\| \frac{a - s(v - v_{\text{in}})}{\varepsilon} + R(s) \right\| > \frac{\|a\| - 2\varepsilon^\alpha \sqrt{2 \frac{\|v_{\text{out}} - v_{\text{in}}\|^2}{2^7}}}{\varepsilon} - 4\varepsilon^{\beta-1}$$
$$\geq \varepsilon^{\alpha-1} \frac{\|v_{\text{out}} - v_{\text{in}}\|}{4} - 4\varepsilon^{\beta-1} \qquad > 1$$

for small enough ε, where $\|v - v_{\text{in}}\| = \sqrt{2(1 - \langle v_{\text{in}}, v \rangle)}$, the estimate $s \leq t_{\text{m}} + \varepsilon^\beta \tilde{t} \leq \varepsilon^\alpha + \varepsilon^\beta \leq 2\varepsilon^\alpha$ and the definition (4.22) of a were applied.

In both cases, the argument $\frac{a - s(v - v_{\text{in}})}{\varepsilon} + R(s)$ is not contained in the support $\operatorname{supp} \phi_x \subset B^d(0, 1)$ and thus $\psi(s, v) = 0$.

c) Because ϕ_x is supported in $B^d(0, 1)$ and $\|R(\zeta(\varepsilon \tilde{z}))\| \leq 4\varepsilon^{\beta-1} \leq 4$ by a), the triangle inequality shows that all z in the support of $\phi_x(z + R(\zeta(\varepsilon z)))$ satisfy $\|z\| \leq 5$.

The second inclusion follows from an application of Lemma B.2.1b) to U. This shows that there exists a $\mu > 0$ independent of ε such that a small ball $B^d(0, \mu\varepsilon^\alpha)$ with radius of order $t_{\text{m}} = \varepsilon^\alpha$ is contained in $\mathcal{T}_a^{v_{\text{in}}}(U)$ for small enough ε, see Fig B.2.1. Therefore, $\varepsilon^{-1}\mathcal{T}_a^{v_{\text{in}}}(U)$ contains the ball $B(0, \mu\varepsilon^{\alpha-1}) \supset B(0, 5)$ for small enough ε by $\alpha < 1$.

□

4.2.3. The Multiple Tumble Part $f_{\geq 2}$

As $f_{\geq 2}$ has vanishing initial condition, the only mass it contains stems from from the non-negative source in ($T_{\geq 2}$). The source injects mass over time, and the size of $f_{\geq 2}$ corresponds to a time integrated version of this source term (A.1.5). This additional order of measurement time t_{m} will eventually drive the measurement $M(f_{\geq 2}) \to 0$ as $t_{\text{m}} \to 0$. Intuitively, there is no time for the source to act, and bacteria do not have enough time to tumble twice or more frequently.

Proof of Lemma 4.9. Lemma 4.5 provides boundedness of $\|f_{\geq 2}(t)\|_\infty$ by its source in

(4.14), which can be used to bound the measurement $M(f_{\geq 2})$ by

$$\begin{aligned} M(f_{\geq 2}) &= \int_0^T \int_{\mathbb{R}^d} \int_V f_{\geq 2}(t,x,v)\,\mathrm{d}v\ \mu^\varepsilon(t,x)\,\mathrm{d}x\,\mathrm{d}t \\ &\leq \int_0^T |V| \|f_{\geq 2}(t)\|_{L^\infty(\mathbb{R}^d\times V)} \int_{\mathbb{R}^d} \mu^\varepsilon(t,x)\,\mathrm{d}x\,\mathrm{d}t \\ &\leq C_K |V|^2 e^{C_K|V|T} \int_0^T \int_0^t \|f_1(s)\|_{L^\infty(\mathbb{R}^d\times V)}\,\mathrm{d}s \int_{\mathbb{R}^d} \mu^\varepsilon(t,x)\,\mathrm{d}x\,\mathrm{d}t. \end{aligned} \tag{4.25}$$

Step 1: Boundedness of the source term f_1. The explicit form of f_1 in (4.13), and the initial data in (4.7) provide

$$\begin{aligned} &\|f_1(s)\|_{L^\infty(\mathbb{R}^d\times V)} \\ &\leq \int_0^s \frac{C_K}{\varepsilon^d} \left\| \phi_x\left(\tfrac{x-v\tau-v'(s-\tau)-x_{\mathrm{in}}}{\varepsilon}\right)\right\|_{L^\infty(\mathbb{R}^d\times V)} \int_V \frac{1}{\varepsilon^{\beta(d-1)}} \phi_v\left(\tfrac{\mathbb{P}_{-v_{\mathrm{in}}}(v')}{\varepsilon^\beta}\right) j_{\mathbb{P}_{-v_{\mathrm{in}}}}(v')\,\mathrm{d}v'\ \mathrm{d}\tau \\ &\leq \frac{C_K}{\varepsilon^d} s \end{aligned}$$

by boundedness of $K \in \mathcal{A}_K^{\mathrm{cont}}$ and $\phi_x \leq 1$, non-negativity of σ, and the fact that the ϕ_v integral in the second line attains value 1.

Step 2: Boundedness of $f_{\geq 2}$. Inserting the above bound for f_1 in (4.25) and application of the change of variables according to Lemma 4.4b shows

$$\begin{aligned} M(f_{\geq 2}) &\leq \frac{1}{\varepsilon^d} C_K^2 |V|^2 e^{C_K|V|T} \int_0^T \int_0^t s\,\mathrm{d}s \int_{\mathbb{R}^d} \mu^\varepsilon(t,x)\,\mathrm{d}t\,\mathrm{d}x \\ &\leq \frac{C_{v_{\mathrm{in}},v_{\mathrm{out}}} C_K^2 |V|^2 e^{C_K|V|T}}{2^d} \frac{t_{\mathrm{m}}^{d-1}}{\varepsilon^d} \int_{-1}^1 (t_{\mathrm{m}} + \varepsilon^\beta \tilde{t})^2 \mu_t(\tilde{t}) \int_{-1}^1 \mu_x(\tilde{x})\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{t} \\ &\leq C\varepsilon^{\alpha(d+1)-d} \end{aligned}$$

for sufficiently small $\varepsilon > 0$, since $|\varepsilon^\beta \tilde{t}| \leq \varepsilon^\alpha = t_{\mathrm{m}}$ for small ε. The constant $C := C_{v_{\mathrm{in}},v_{\mathrm{out}}} C_K^2 |V|^2 e^{C_K|V|T}/2^{d-2}$ collects all constants.

Step 3: Convergence of $M(f_{\geq 2})$. Non-negativity of $M(f_{\geq 2})$ by non-negativity of μ^ε and $f_{\geq 2}$ according to Lemma 4.5 provides the desired result, since $\alpha > \frac{3}{4}$ was chosen in such a way that the exponent $(d+1)\alpha - d > 0$ is positive and therefore

$$0 \leq \lim_{\varepsilon^\beta \to 0} M_{\mu^\varepsilon}(f_{\geq 2}^{\phi^\varepsilon}) \leq C\varepsilon^{(d+1)\alpha - d} \xrightarrow{\varepsilon\to 0} 0.$$

□

4.3. Proof of Theorem 4.1 in Dimension $d = 1$

In the 1-dimensional setting, the reconstruction procedure can be simplified. Given that only two velocities $v \in \{\pm 1\} = V$ exist, the absorption parameter σ in (3.1) only contains information on one evaluation of the tumbling parameter

$$\sigma(x,v) = K(x,-v,v), \tag{4.26}$$

as $K(x,v,v)=0$ for $K\in\mathcal{A}_K^{\text{cont}}$ by definition (3.3). Chemotaxis model (Ch) then reads

$$\begin{aligned}\partial_t f + v\partial_x f &= K(x,v,-v)f' - K(x,-v,v)f,\\ f(t=0) &= \phi(x,v),\end{aligned}$$

where the notation $f'(x,t,v) := f(x,t,-v)$ is used to simplify the presentation. This shows that information on the decay rate σ is actually sufficient for the reconstruction of $K\in\mathcal{A}_K^{\text{cont}}$ in one space dimension. As observed in the higher dimensional case in (4.12), decay can already be observed in the ballistic part f_0. For this reason, the construction of designs for the singular decomposition approach can be simplified.

This is not possible in higher dimensions $d\in\{2,3\}$, because σ contains information of several tumbling kernel values through integration in (3.1), whose influence cannot be separated. The decay behaviour is not sufficient for the recovery of K in Section 4.2 and the higher order tumbling part f_1 had to be consulted.

The reconstruction strategy presented in the following is based on the reconstruction of σ, as presented in Section 3 of [HKLT24] in the multidimensional case, and results from joint work with Christian Klingenberg, Qin Li and Min Tang. This strategy is adapted to the 1D setting. A detailed comparison is conducted in Remark 4.18

Formal Intuition. The reconstruction process can again be motivated by a formal consideration, as in Section 4.2. Consider (4.3) in the 1D scenario and for initial data $f(t=0,x,v)=\mathbb{1}_{x_{\text{in}}}(x)\mathbb{1}_{v_{\text{in}}}(v)$, where e.g. $\mathbb{1}_{v_{\text{in}}}$ denotes the indicator function[1] over the set $\{v_{\text{in}}\}\subset V$, and assume that the order of the remainder stays unchanged. Then the point wise measurement at location x reads

$$\begin{aligned}\int_V f(t_{\text{m}},x,v)\,\mathrm{d}v \approx \mathbb{1}_{x_{\text{in}}}(x-v_{\text{in}}t_{\text{m}})\left(1-\int_0^{t_{\text{m}}} K(x-v_{\text{in}}s,-v_{\text{in}},v_{\text{in}})\,\mathrm{d}s\right)\\ +\int_0^{t_{\text{m}}} K(x+v_{\text{in}}s,-v_{\text{in}},v_{\text{in}})\mathbb{1}_{x_{\text{in}}}(x+v_{\text{in}}s-v_{\text{in}}(t_{\text{m}}-s))\,\mathrm{d}s+o(t_{\text{m}}).\end{aligned}$$

Choosing measurement location $x_{\text{m}}=x_{\text{in}}+v_{\text{in}}t_{\text{m}}$, the second integral term vanishes, because the indicator function $\mathbb{1}_{x_{\text{in}}}(x_{\text{m}}+v_{\text{in}}s-v_{\text{in}}(t_{\text{m}}-s))=\mathbb{1}_{x_{\text{in}}}(x_{\text{in}}+2v_{\text{in}}s)$ attains a non zero value only on the Lebesgue null set $\{s=0\}$. The measurement simplifies to

$$\int_V f(t_{\text{m}},x_{\text{m}},v)\,\mathrm{d}v \approx 1-\int_0^{t_{\text{m}}} K(x_{\text{in}}+v_{\text{in}}s,-v_{\text{in}},v_{\text{in}})\,\mathrm{d}s+o(t_{\text{m}}).$$

To obtain the evaluation of K at a specific point, the derivative w.r.t. measurement time can be taken

$$\partial_{t_{\text{m}}}\int_V f(t_{\text{m}},x_{\text{m}},v)\,\mathrm{d}v \approx -K(x_{\text{m}},-v_{\text{in}},v_{\text{in}})+\mathcal{O}(t_{\text{m}}).$$

[1] The indicator function $\mathbb{1}_A$ of a set A is defined as $\mathbb{1}_A(a)=1$ if $a\in A$ and zero otherwise. It is convention to write $\mathbb{1}_a:=\mathbb{1}_{\{a\}}$ for singleton sets $\{a\}$ with $a\in A$.

Construction of Initial Data and Measurement Test Function. The detailed construction is described subsequently.

Geometry. Let $(x_\star, v_{\text{in}}) \in \mathbb{R} \times V$ be the point at which the value of the tumbling kernel $K(x_\star, -v_{\text{in}}, v_{\text{in}})$ is supposed to be reconstructed. Furthermore, let $\varepsilon > 0$ be a small scaling parameter. For small measurement times, compatible initial and measurement locations shall be given by

$$t_{\text{m}} \in (\varepsilon, 3\varepsilon) \quad \text{with} \quad t_{\text{m}} < T, \quad \text{and} \tag{4.27}$$

$$x_{\text{in}} = x_{\text{in}}(t_{\text{m}}) = x_\star - v_{\text{in}} t_{\text{m}} \qquad \text{and} \qquad x_{\text{m}} = x_\star. \tag{4.28}$$

Singularity. Given $\phi_x, \mu_x, \mu_t \in \Xi^1$ as defined in (4.9), initial data and measurement test function with concentration points $x_{\text{in}}, v_{\text{in}}$ and $x_{\text{m}}, t_{\text{m}}$, respectively, are constructed as

$$\phi^\varepsilon_{t_{\text{m}}}(x, v) = \phi_x\left(\frac{x - x_{\text{in}}(t_{\text{m}})}{\varepsilon}\right) \mathbb{1}_{v_{\text{in}}}(v)\,, \tag{4.29}$$

$$\mu^\varepsilon_{t_{\text{m}}}(t, x) = \frac{1}{\varepsilon^2}\mu_x\left(\frac{x - x_{\text{m}}}{\varepsilon}\right)\mu_t\left(\frac{t - t_m}{\varepsilon}\right). \tag{4.30}$$

Given that $V = \{\pm 1\}$ is discrete, $\mathbb{1}_{v_{\text{in}}}$ is singular in v. The singularities in initial location and measurement location and time are coupled through the usage of the same scaling parameter ε, that also describes smallness of measurement time $t_{\text{m}} \in (\varepsilon, 3\varepsilon)$. The following lemma summarizes frequently used properties of $\phi^\varepsilon_{t_{\text{m}}}, \mu^\varepsilon_{t_{\text{m}}}$.

Lemma 4.12. *Let $\varepsilon > 0$ be sufficiently small.*

a) Then $\phi^\varepsilon_{t_{\text{m}}}, \mu^\varepsilon_{t_{\text{m}}}$ as constructed above in (4.29)–(4.30) are non-negative, smooth and compactly supported, to be precise

- *$\phi^\varepsilon_{t_{\text{m}}}(\cdot, v) \in C_c^\infty(\mathbb{R})$ for all $v \in V$ with $\operatorname{supp} \phi^\varepsilon_{t_{\text{m}}} \subset B(x_{\text{in}}, \varepsilon) \times \{v_{\text{in}}\}$, and*
- *$\mu^\varepsilon_{t_{\text{m}}} \in C_c^\infty([0, T] \times \mathbb{R})$ with $\operatorname{supp} \mu^\varepsilon_{t_{\text{m}}} \subset B(t_{\text{m}}, \varepsilon) \times B(x_{\text{m}}, \varepsilon)$.*

b) For continuous functions $h : [0, T] \times \mathbb{R}^d \to \mathbb{R}$, the following change of variables formula holds

$$\int_0^T \int_{\mathbb{R}} h(t, x)\mu^\varepsilon_{t_{\text{m}}}(t, x)\,\mathrm{d}x\,\mathrm{d}t = \int_{-1}^1 \int_{-1}^1 h(t_{\text{m}} + \varepsilon\tilde{t}, x_{\text{m}} + \varepsilon\tilde{x})\mu_t(\tilde{t})\mu_x(\tilde{x})\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{t}.$$

Proof. Assertion a) clear from the definition of $\phi^\varepsilon_{t_{\text{m}}}, \mu^\varepsilon_{t_{\text{m}}}$. Application of the transformations $\tilde{t} = \frac{t - t_{\text{m}}}{\varepsilon}$ and $\tilde{x} = \frac{x - x_{\text{m}}}{\varepsilon}$ provides the given form of the integrand in assertion b). The new integration domain is reduced with regards to the supports $\operatorname{supp} \mu_x \subset [-1, 1] \subset \mathbb{R}$ and $\operatorname{supp} \mu_t \subset [-1, 1] \subset [-\frac{t_{\text{m}}}{\varepsilon}, \frac{T - t_{\text{m}}}{\varepsilon}]$. They are contained in the transformed domains, which are spelled out at the end of the inclusion chains, for sufficiently small ε by choice of $t_{\text{m}} \in (\varepsilon, 3\varepsilon)$ and fixed $T > 0$. □

Remark 4.13 (Comparison to the setting in Section 4.2). In this section, a different geometry is constructed with the aim of measuring at the ballistic location $x_b = x_{\text{in}} + t_{\text{m}} v_{\text{in}}$, instead of the scattered location $x_{\text{in}} + \frac{t_{\text{m}}}{2} v_{\text{in}} + \frac{t_{\text{m}}}{2} v_{\text{out}}$. Singularity in the initial location and measurement time and location is generated at the same rate, as smallness of measurement

time is. The type of singularity that is induced by the initial data is qualitatively different, as the former formal intuition showed: in Section 4.2, the initial condition converged to a Dirac delta in space, whereas it is a characteristic function in this setting. This strong singularity in initial location was necessary to fix the tumbling location, and required a balancing by a small multiplicative factors of order t_{m}^{d-1}, to keep the measurement finite. In the 1D setting, it is not necessary, since the geometry explicitly prescribes all considered locations for the ballistic part f_0. The weaker introduced singularity in the initial data will, however, lead to a more complex postprocessing of the data for reconstruction that involves differentiation of the measurement w.r.t. measurement time, as suggested by the above formal intuition.

Decomposition. Given that the ballistic term is already subject to decay with the absorption rate $\sigma(x,v) = K(x,-v,v)$, this term is expected to be sufficient for the reconstruction of K in the one dimensional case. Thus, the solution of f is decomposed into two parts

$$f = f_0 + f_{\geq 1}, \tag{4.31}$$

where the ballistic part f_0 solves

$$\partial_t f_0^\phi + v\partial_x f_0^\phi = -K(x,-v,v) f_0^\phi, \tag{4.32}$$

$$f_0^\phi(t=0) = \phi, \tag{4.33}$$

and the remainder $f_{\geq 1}$ solves

$$\begin{aligned} \partial_t f_{\geq 1}^\phi + v\partial_x f_{\geq 1}^\phi &= K(x,v,-v)((f_0^\phi)' + (f_{\geq 1}^\phi)') - K(x,-v,v) f_{\geq 1}^\phi, \\ f_{\geq 1}^\phi(t=0) &= 0. \end{aligned}$$

In analogy to Lemma 4.5, existence of non-negative solutions f_0, $f_{\geq 1}$ is provided by Corollary A.1.3 and Lemma A.1.5a for non-negative initial data ϕ.

Behaviour of the Measurement. The following Proposition proves Theorem 4.1 in one space dimension by deploying of the formerly constructed setting.

Proposition 4.14. *Let $K \in \mathcal{A}_K^{\mathrm{cont}}$ and consider a fixed point $(x_\star, v_{\mathrm{in}}) \in \mathbb{R} \times V$. Let $(\phi_{t_{\mathrm{m}}}^\varepsilon, \mu_{t_{\mathrm{m}}}^\varepsilon)$ and $t_{\mathrm{m}}, x_{\mathrm{in}}, x_{\mathrm{m}}$ be constructed as in (4.27)-(4.30) for $\varepsilon > 0$. Then*

$$\lim_{\varepsilon \to 0} \partial_{t_{\mathrm{m}}} M_{\mu_{t_{\mathrm{m}}}^\varepsilon}(f^{\phi_{t_{\mathrm{m}}}^\varepsilon}) \,|_{t_{\mathrm{m}}=2\varepsilon} = -K(x_\star, -v_{\mathrm{in}}, v_{\mathrm{in}}) C_{\phi_x,\mu_x,\mu_t}\,,$$

with strictly positive, known constant $0 < C_{\phi_x,\mu_x,\mu_t} := \int_{-1}^1 \int_{\mathbb{R}} \phi_x\left(x - v_{\mathrm{in}} t\right) \mu_x\left(x\right) \,\mathrm{d}x\, \mu_t(t)\,\mathrm{d}t$.

The strict positivity of C_{ϕ_x,μ_x,μ_t} holds due to non-negativity and continuity of ϕ_x, μ_x, μ_t and the fact they attain value 1 at location 0.

As in the multi-dimensional case, this proposition establishes an explicit relationship between a sequence of measurements and the parameter. As a result of the simplified experimental requirements, in particular the weaker introduced singularity in the initial data, this relationship is more complex compared to the one in Section 4.2.

The proof is divided in two lemmas that study the different limiting behaviour of the two parts of the measurement derivative as suggested by the decomposition (4.31):

$$\partial_{t_\mathrm{m}} M(f) = \partial_{t_\mathrm{m}} M(f_0) + \partial_{t_\mathrm{m}} M(f_{\geq 1}). \tag{4.34}$$

The ballistic part f_0 carries the desired information on K.

Lemma 4.15. *Let $K \in \mathcal{A}_K^{\mathrm{cont}}$. Given a point $(x_\star, v_\mathrm{in}) \in \mathbb{R} \times V$, let $\phi_{t_\mathrm{m}}^\varepsilon, \mu_{t_\mathrm{m}}^\varepsilon$ be defined as in (4.29)-(4.30) with $t_\mathrm{m}, x_\mathrm{in}, x_\mathrm{m}$ as in (4.27)–(4.28) for small $\varepsilon > 0$. Then*

$$\lim_{\varepsilon \to 0} \partial_{t_\mathrm{m}} M_{\mu_{t_\mathrm{m}}^\varepsilon}(f_0^{\phi_{t_\mathrm{m}}^\varepsilon}) \,|_{t_\mathrm{m}=2\varepsilon} = -K(x_\star, -v_\mathrm{in}, v_\mathrm{in}) C_{\phi_x, \mu_x, \mu_t}.$$

The contribution of the higher regularity part $f_{\geq 1}$, instead, vanishes asymptotically. The reason lies in the short measurement time, that only allows for maximum one tumbling. One tumble, however, leads to a deviation from the shortest path from x_in to $x_\mathrm{m} = x_\mathrm{b}$, and thus a delayed arrival, which excludes these particles from the measurement.

Lemma 4.16. *Let $K \in \mathcal{A}_K^{\mathrm{cont}}$, and let $\phi_{t_\mathrm{m}}^\varepsilon, \mu_{t_\mathrm{m}}^\varepsilon$ be defined as in (4.28)-(4.30) with $t_\mathrm{m}, x_\mathrm{in}, x_\mathrm{m}$ as in (4.27)–(4.28) for some $\varepsilon > 0$, and an evaluation point $(x_\star, v_\mathrm{in}) \in \mathbb{R} \times V$. Then*

$$\lim_{\varepsilon \to 0} \partial_{t_\mathrm{m}} M_{\mu_{t_\mathrm{m}}^\varepsilon}(f_{\geq 1}^{\phi_{t_\mathrm{m}}^\varepsilon}) \,|_{t_\mathrm{m}=2\varepsilon} = 0.$$

In summary, this proves Proposition 4.14 in the 1D case.

Proof of Proposition 4.14 in 1D. Inserting the results from Lemma 4.15 and Lemma 4.16 into (4.34) shows

$$\lim_{\varepsilon \to 0} \partial_{t_\mathrm{m}} M_{\mu_{t_\mathrm{m}}^\varepsilon}(f^{\phi_{t_\mathrm{m}}^\varepsilon}) \,|_{t_\mathrm{m}=2\varepsilon} = -K(x_\star, -v_\mathrm{in}, v_\mathrm{in}) C_{\phi_x, \mu_x, \mu_t}.$$

□

Remark 4.17. With the same technique as in Remark 4.10, the thus obtained results can be extended to K with continuous time dependence.

Remark 4.18 (Comparison with [HKLT24]). In contrast to the thus presented setting, the derivations in [HKLT24] made use of independent scaling parameters. Singularity in initial velocity and measurement time was generated independently and the limit was taken before the coupled singularities in initial and measurement location were generated. Post processing of the measurement in the limiting case, in particular taking the derivative after the limit is established, yielded the value of σ at the ballistic location $x_b(t_\mathrm{m}) = x_\star + t_\mathrm{m} v_\mathrm{in}$. In that sense, one initial condition could be used to find all values $\sigma(x_b(t_\mathrm{m}), v_\mathrm{in})$, if the measurement could be taken for all these measurement times $t_\mathrm{m} \in (0, \hat{T})$ and the measurement location $x_b(t_\mathrm{m})$ was moved accordingly.

This work uses a setting where all singularities and smallness of measurement time share the same scaling parameter, and derivation of the measurement w.r.t. measurement time is taken before the limit. The reason is the same as in Remark 4.3: in reality, only an approximation to small time and singular initial and test data will be available and has to be processed. This setting requires small measurement time to bound the effect of the scattering part $f_{\geq 1}$, that could be controlled by singularity of the initial and measurement data in [HKLT24].

4.3.1. The Ballistic Part f_0

Similarly to the proofs of Lemma 4.7 and Lemma 4.8, the explicit formula of f_0 and the construction of $\phi^\varepsilon_{t_\mathrm{m}}, \mu^\varepsilon_{t_\mathrm{m}}$ are exploited in the proof of Lemma 4.15.

Proof of Lemma 4.15. Use the explicit form (4.12) of f_0 in the 1D setting (4.26) together with Lemma 4.12b to see

$$M(f_0) = \int_{-1}^{1}\int_{-1}^{1} e^{-\int_0^{t_\mathrm{m}+\varepsilon\tilde{t}} K(x_\mathrm{m}+\varepsilon\tilde{x}-v_\mathrm{in}s,-v_\mathrm{in},v_\mathrm{in})\,\mathrm{d}s} \phi_x\left(\tilde{x} - v_\mathrm{in}\tilde{t} + \tfrac{x_\mathrm{m}-v_\mathrm{in}t_\mathrm{m}-x_\mathrm{in}}{\varepsilon}\right) \mu_x\left(\tilde{x}\right) \mu_t\left(\tilde{t}\right)\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{t},$$

for small enough $\varepsilon > 0$. Because the fraction in the argument of ϕ_x vanishes $\frac{x_\mathrm{m}-v_\mathrm{in}t_\mathrm{m}-x_\mathrm{in}}{\varepsilon} = 0$ independently of t_m by construction of $x_\mathrm{m}, x_\mathrm{in}$ in (4.28), derivation w.r.t. t_m shows that

$$\begin{aligned}\partial_{t_\mathrm{m}} M(f_0) = \int_{-1}^{1}\int_{-1}^{1} &- K(x_\mathrm{m}+\varepsilon\tilde{x}-v_\mathrm{in}(t_\mathrm{m}+\varepsilon\tilde{t}),-v_\mathrm{in},v_\mathrm{in}) e^{-\int_0^{t_\mathrm{m}+\varepsilon\tilde{t}} K(x_\mathrm{m}+\varepsilon\tilde{x}-v_\mathrm{in}s,-v_\mathrm{in},v_\mathrm{in})\,\mathrm{d}s}\cdot \\ &\phi_x\left(\tilde{x} - v_\mathrm{in}\tilde{t}\right) \mu_x\left(\tilde{x}\right) \mu_t\left(\tilde{t}\right)\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{t}.\end{aligned}$$

In $t_\mathrm{m} = 2\varepsilon$, the limit as $\varepsilon \to 0$ is determined by the dominated convergence theorem, since continuity and boundedness of K provide boundedness and point wise convergence of the integrand. In summary, one obtains the assertion

$$\begin{aligned}\lim_{\varepsilon\to 0} \partial_{t_\mathrm{m}} M(f_0)\,|_{t_\mathrm{m}=2\varepsilon} &= -K(x_\star,-v_\mathrm{in},v_\mathrm{in}) \int_{-1}^{1}\int_{-1}^{1} \phi_x\left(\tilde{x} - v_\mathrm{in}\tilde{t}\right) \mu_x\left(\tilde{x}\right) \mu_t\left(\tilde{t}\right)\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{t} \\ &= -K(x_\star,-v_\mathrm{in},v_\mathrm{in}) C_{\phi_x,\mu_x,\mu_t}.\end{aligned}$$

□

4.3.2. The Tumble Part $f_{\geq 1}$

The remainder is further decomposed into parts f_1 that have undergone exactly one tumbling event, and the remainder $f_{\geq 1} = f_1 + f_{\geq 2}$, where f_1 satisfies (T_1) and $f_{\geq 2}$ satisfies ($\mathrm{T}_{\geq 2}$) in the 1D setting i.e. with absorption σ given by (4.26). The decomposition translates to the measurement by linearity

$$M(f_{\geq 1}) = M(f_1) + M(f_{\geq 2})\,. \tag{4.35}$$

This is necessary, because smallness of the contributions of both parts is generated by different mechanisms.

The measurement of f_1 vanishes in the scaling limit, because these particles cannot reach the measurement location x_m, since tumbling means deviation from the direct route.

Lemma 4.19. *Let $K \in \mathcal{A}_K^{\mathrm{cont}}$ and $\phi^\varepsilon_{t_\mathrm{m}}, \mu^\varepsilon_{t_\mathrm{m}}$ be constricted as in (4.27)–(4.30).*
Then $\lim_{\varepsilon\to 0} \partial_{t_\mathrm{m}} M_{\mu^\varepsilon_{t_\mathrm{m}}}(f_1^{\phi^\varepsilon_{t_\mathrm{m}}})\,|_{t_\mathrm{m}=2\varepsilon} = 0$.

Smallness of $M(f_{\geq 2})$ follows by choice of a small measurement time which leaves no time for further tumbling, i.e. no time for the source term f_1 to introduce mass into $f_{\geq 2}$.

Lemma 4.20. *Let $K \in \mathcal{A}_K^{\mathrm{cont}}$ and $\phi_{t_\mathrm{m}}^\varepsilon, \mu_{t_\mathrm{m}}^\varepsilon$ be given as in (4.27)–(4.30).*
Then $\lim_{\varepsilon\to 0} \partial_{t_\mathrm{m}} M_{\mu_{t_\mathrm{m}}^\varepsilon}(f_{\geq 2}^{\phi_{t_\mathrm{m}}^\varepsilon}) \,|_{t_\mathrm{m}=2\varepsilon} = 0$.

This illustrates that the constructed experiment also aims to allow one tumbling event, but prohibit multiple tumblings, which is again achieved by a small experimental time horizon. Lemma 4.16 is a direct consequence of the above lemmas and it remains to prove Lemma 4.19 and Lemma 4.20.

Proof of Lemma 4.19. Similar as in the proof of Lemma 4.15, an explicit form of $\partial_{t_\mathrm{m}} M(f_1)$ provides convergence.

Step 1: Explicit form of $\partial_{t_\mathrm{m}} M(f_1)$. The explicit form of f_1 in (4.13) in the 1D setting and a change of variables according to Lemma 4.12b yield the following form of $M(f_1)$ for sufficiently small ε:

$$M(f_1) = \int_{-1}^{1}\int_{-1}^{1} \mu_t(\tilde{t})\mu_x(\tilde{x}) \int_0^{t_\mathrm{m}+\varepsilon\tilde{t}} e^{-\int_0^s K(x_\mathrm{m}+\varepsilon\tilde{x}+v_\mathrm{in}\tau, v_\mathrm{in}, -v_\mathrm{in})\,\mathrm{d}\tau} K(x_\mathrm{m}+\varepsilon\tilde{x}+v_\mathrm{in}s, -v_\mathrm{in}, v_\mathrm{in}) \cdot e^{-\int_0^{t_\mathrm{m}+\varepsilon\tilde{t}-s} K(x_\mathrm{m}+\varepsilon\tilde{x}+v_\mathrm{in}(s-\tau), -v_\mathrm{in}, v_\mathrm{in})\,\mathrm{d}\tau} \phi_x\left(\tfrac{\varepsilon\tilde{x}+v_\mathrm{in}s-v_\mathrm{in}(\varepsilon\tilde{t}-s)}{\varepsilon}\right) \mathrm{d}s \;\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{t},$$

where the numerator inside ϕ_x was simplified by relation (4.28). The derivative w.r.t. measurement time consists of two parts

$$\partial_{t_\mathrm{m}} M(f_1) = \int_{-1}^{1}\int_{-1}^{1} \mu_t(\tilde{t})\mu_x(\tilde{x})\big((i)+(ii)\big)\,\mathrm{d}\tilde{x}\,\mathrm{d}\tilde{t},$$

that originate from the differentiation of the boundary of the integral in s and differentiation of the integrand, respectively, and are spelled out below.

Step 2: Convergence of (i). Term (i) reads

$$(i) = e^{-\int_0^{t_\mathrm{m}+\varepsilon\tilde{t}} K(x_\mathrm{m}+\varepsilon\tilde{x}+v_\mathrm{in}\tau, v_\mathrm{in}, -v_\mathrm{in})\,\mathrm{d}\tau} K(x_\mathrm{m}+\varepsilon\tilde{x}+v_\mathrm{in}(t_\mathrm{m}+\varepsilon\tilde{t}), -v_\mathrm{in}, v_\mathrm{in})\phi_x\left(\tfrac{\varepsilon\tilde{x}+v_\mathrm{in}(2t_\mathrm{m}+\varepsilon\tilde{t})}{\varepsilon}\right).$$

Taking $t_\mathrm{m} = 2\varepsilon$, this term vanishes for all $\tilde{t}, \tilde{x} \in [-1,1]$, because the argument of ϕ_x is not contained in its support $\operatorname{supp}\phi_x \subset [-1,1]$, as an application of the reverse triangle inequality suggests

$$\left|\tfrac{\varepsilon\tilde{x}+v_\mathrm{in}(2t_\mathrm{m}+\varepsilon\tilde{t})}{\varepsilon}\right| \geq 2\frac{t_\mathrm{m}}{\varepsilon}|v_\mathrm{in}| - |v_\mathrm{in}||\tilde{t}| - |\tilde{x}| \geq 4 - 1 - 1 = 2.$$

Step 3: Convergence of (ii). On the other hand, the term (ii) can be bounded uniformly in $\tilde{x}, \tilde{t}$ by

$$\begin{aligned}(ii) =& \int_0^{t_\mathrm{m}+\varepsilon\tilde{t}} e^{-\int_0^s K(x_\mathrm{m}+\varepsilon\tilde{x}+v_\mathrm{in}\tau, v_\mathrm{in}, -v_\mathrm{in})\,\mathrm{d}\tau} K(x_\mathrm{m}+\varepsilon\tilde{x}+v_\mathrm{in}s, -v_\mathrm{in}, v_\mathrm{in})\phi_x\left(\tfrac{\varepsilon\tilde{x}+v_\mathrm{in}s-v_\mathrm{in}(\varepsilon\tilde{t}-s)}{\varepsilon}\right) \cdot \\ & e^{-\int_0^{t_\mathrm{m}+\varepsilon\tilde{t}-s} K(x_\mathrm{m}+\varepsilon\tilde{x}+v_\mathrm{in}(s-\tau), -v_\mathrm{in}, v_\mathrm{in})\,\mathrm{d}\tau} \big(-K(x_\mathrm{m}+\varepsilon\tilde{x}+v_\mathrm{in}(2s-t_\mathrm{m}-\varepsilon\tilde{t}), -v_\mathrm{in}, v_\mathrm{in})\big)\,\mathrm{d}s \\ =& \mathcal{O}(t_\mathrm{m}+\varepsilon\tilde{t}) \leq \mathcal{O}(t_\mathrm{m}+\varepsilon),\end{aligned}$$

by uniform boundedness of the integrand by C_K^2. Choosing $t_\mathrm{m} = 2\varepsilon$, this means that $\lim_{\varepsilon\to 0}(ii) = 0$ uniformly in $\tilde{t}, \tilde{x}$. In summarty, this proves the lemma.

□

Proof of Lemma 4.20. The formal derivation of this part of the measurement yields

$$|\partial_{t_\mathrm{m}} M(f_{\geq 2})| = \left| \int_0^T \int_{\mathbb{R}} \int_V \left((\partial_{t_\mathrm{m}} f_{\geq 2})\ \mu^\varepsilon + f_{\geq 2} \partial_{t_\mathrm{m}} \mu^\varepsilon \right) \mathrm{d}v\, \mathrm{d}x\, \mathrm{d}t \right| \tag{4.36}$$
$$\leq \|\partial_{t_\mathrm{m}} f_{\geq 2}\|_{L^\infty([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R}\times V)} \|\mu^\varepsilon\|_{L^1} + \|f_{\geq 2}\|_{L^\infty([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R}\times V)} \|\partial_{t_\mathrm{m}} \mu^\varepsilon\|_{L^1}$$

by Hölder's inequality. Note that the time domain in the norm of $f_{\geq 2}, \partial_{t_\mathrm{m}} f_{\geq 2}$ was reduced to $[0, t_\mathrm{m} + \varepsilon]$ by consideration of the temporal domain of μ^ε and $\partial_{t_\mathrm{m}} \mu^\varepsilon$ in Lemma 4.12a. Dependency on t_m is introduced in $f_{\geq 2}$ by its source term $\mathcal{L}(f_1)$.

Step 1: L^∞ bound for $f_{\geq 2}$. A direct application of Corollary A.1.3 shows

$$|f_{\geq 2}(t,x,v)| \leq e^{2C_K t} \int_0^t \sup_{x,v'} |K(x,v',-v') f_1(s,x,-v')|\, \mathrm{d}s \leq e^{2C_K t} C_K \int_0^t \sup_{x,v'} |f_1(s,x,v')|\, \mathrm{d}s$$

recalling that $|V| = 2$ in 1D. By (4.13), one can bound $f_1(s,x,v')$ by

$$|f_1(s,x,v)| \leq C_K s \|\phi^\varepsilon_{t_\mathrm{m}}\|_{L^\infty} \leq C_K s.$$

In total, this gives

$$|f_{\geq 2}(t,x,v)| \leq e^{2C_K t} C_K^2 \frac{t^2}{2}.$$

Step 2: L^∞ bound for $\partial_{t_\mathrm{m}} f_{\geq 2}$. First, note that by linearity of the equations $\partial_{t_\mathrm{m}} f_{\geq 2}, \partial_{t_\mathrm{m}} f_1, \partial_{t_\mathrm{m}} f_0$ solve the same systems ($\mathrm{T}_{\geq 2}$), (T_1), (T_0) as $f_{\geq 2}, f_1, f_0$, but with source term $\mathcal{L}(\partial_{t_\mathrm{m}} f_1), \mathcal{L}(\partial_{t_\mathrm{m}} f_0)$ and initial condition $\partial_{t_\mathrm{m}} \phi^\varepsilon_{t_\mathrm{m}}(x,v) = v_\mathrm{in} \varepsilon^{-1} \phi_x' \left(\frac{x - x_\mathrm{in}}{\varepsilon}\right) \mathbb{1}_{v_\mathrm{in}}(v)$, respectively, where ϕ_x' denotes the derivative of ϕ_x. The same derivation as above yields

$$|\partial_{t_\mathrm{m}} f_{\geq 2}(t,x,v)| \leq e^{2C_K t} C_K^2 \frac{t^2}{2} \|\partial_{t_\mathrm{m}} \phi^\varepsilon_{t_\mathrm{m}}\|_{L^\infty} \leq \varepsilon^{-1} e^{2C_K t} C_K^2 \frac{t^2}{2} C_{\phi_x'},$$

where the derivative $\|\phi_x'\|_{L^\infty} \leq C_{\phi_x'} < \infty$ is bounded by construction of $\phi_x \in C_c^\infty(\mathbb{R})$.

Step 3: L^1 bounds for $\mu^\varepsilon, \partial_{t_\mathrm{m}} \mu^\varepsilon$. The standard change of variables from Lemma 4.12b shows that for small enough ε

$$\|\mu^\varepsilon\|_{L^1([0,T]\times\mathbb{R}\times V)} = 2 \int_{-1}^1 \int_{-1}^1 \mu_x(\tilde{x}) \mu_t(\tilde{t})\, \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{t} = 2,$$

where $2 = |V|$ was used. In analogy, one obtains for small $\varepsilon > 0$ that $\partial_{t_\mathrm{m}} \mu^\varepsilon(t,x) = -\varepsilon^{-3} \mu_x \left(\frac{x - x_\mathrm{m}}{\varepsilon}\right) \mu_t' \left(\frac{t - t_\mathrm{m}}{\varepsilon}\right)$ and with $\|\mu_t'\|_{L^\infty} \leq C_{\mu_t'} < \infty$ this is bounded by

$$\|\partial_{t_\mathrm{m}} \mu^\varepsilon\|_{L^1([0,T]\times\mathbb{R}\times V)} = 2\varepsilon^{-1} \int_{-1}^1 \int_{-1}^1 \mu_x(\tilde{x}) \left|\mu_t'\left(\tilde{t}\right)\right| \mathrm{d}\tilde{x}\, \mathrm{d}\tilde{t} \leq \varepsilon^{-1} 4 C_{\mu_t'}.$$

Step 4: Convergence. Inserting everything in (4.36) shows

$$|\partial_{t_\mathrm{m}} M(f_{\geq 2})| \leq \varepsilon^{-1} (t_\mathrm{m} + \varepsilon)^2 e^{2C_K(t_\mathrm{m}+\varepsilon)} \frac{C_K^2}{2} (2 C_{\phi_x'} + 4 C_{\mu_t'}).$$

For $t_\mathrm{m} = 2\varepsilon$, this proves the assertion $|\partial_{t_\mathrm{m}} M(f_{\geq 2})| \to 0$ as $\varepsilon \to 0$. □

5

Sensitivity Based & Cost Function Identifiability: Experimental Design

In the previous chapter, structural identifiability of the chemotaxis kernel reconstruction problem (Ch^{-1}) was shown under access to the full infinite dimensional input to output map $\mathrm{ItO}_K : L^1_+ \cap L^\infty(\mathbb{R}^d \times V) \times L^1([0,T] \times \mathbb{R}^d) \to \mathbb{R}$. In real or numerical experiments, this is typically a too strong assumption: only a finite number of experiments can be conducted, with only a finite number of sensors each. This reduces the data space to finite dimension, and the inverse problem requires a new balancing: the dimension of the admissible set has to be reduced accordingly, to sustain injectivity of the forward map. The tighter controlled form of the parameter adds a-priori knowledge in order to leverage the reduced information available in the data. The precise manner in which the experiment compresses the parameter information in the data, and with it the required form of induced a-priori knowledge on the parameter, is dictated by the specific choice of the experimental design, as anticipated by source (II) of non-identifiability. Vice versa, designs are typically constructed in adaptation to a given admissible set to maximize the information on the parameter in the data. This intricate interplay between admissible set, design and identifiability is, however, far from obvious, and it is not immediately clear which data are required to reconstruct a parameter of a prescribed form.

This chapter is thus devoted to the analytical investigation of this balance between a-priori knowledge in the admissible set and demonstrate the 'Relaxation of Theory' approach to experimental design for the chemotaxis inverse problem (Ch^{-1}), in the framework of sensitivity based identifiability. First, equivalence of strong convexity in the quadratic cost function and a sensitivity based criterion [MXPW11, VRWL89] is established. Then, this criterion is used to characterize experimental designs for their suitability:

(i) Degeneracy of the inverse problem is shown, if the data is not sufficiently diverse, in the sense that two out of a minimal number of experimental setups are asymptotically close. This makes sense intuitively, because the two respective measurements asymptotically coincide and carry the same information on K. By minimality of the data dimension, the inverse problem thus becomes underdetermined.

(ii) In contrast to that, one experimental setting is carefully crafted, for which identifiability is proven analytically in the 1-dimensional case. It represents a relaxation of the construction of the singular decomposition experiments in Section 4.3, which is possible by a suitable choice of the admissible set. This demonstrates existence of a minimal data experimental design that renders the chemotaxis inverse problem least squares identifiable and provides a direct connection between the theoretical infinite dimensional and the practical finite dimensional setting.

The analysis demonstrates capabilities and limitations of the inverse chemotaxis setting: it gives insight, which type of parameter information can be reconstructed from a given number of data, if these data are well-prepared. This presents a valuable information in applications where real experimentation is expensive and a traded off between experimentation cost and reconstruction accuracy is typical. It should be mentioned, that even for higher data dimension, a poorly chosen experimental design can still prohibit a successful reconstruction, for instance if the problem is nonetheless underdetermined.

The presented analysis is set in a noise-free setting where data is generated by a ground truth parameter in the admissible set, and no regularization is applied in the cost function, in order to focus only on the sole effect of the experimental design in the dimension reduction process. Noise can be incorporated in the analysis of cost function analysis as demonstrated in [VRWL89] which brings the analysis closer to the framework of Bayesian optimal experimental design for linear inverse problems, as breifly introduced in Section 2.5.

This chapter is based on a joint work with Christian Klingenberg, Qin Li and Min Tang [HKLT25].

Remark 5.1 (Difference to [HKLT25]). A major difference to [HKLT25] is the explicit explanation of the derivation of the sensitivity based identifiable experimental design from the theoretical results of the previous chapter. More emphasis is put on the connection of the results to the identifiability framework, and the analysis is adapted to the framework of this work, where the measurement operator is described by spatio-temporal test functions in instead of a point wise measurement at a prescribed measurement time in [HKLT25]. Moreover, the non-identifiability results under (i) are studied in a generalized setting where the admissible set prescribes an arbitrary fixed basis element expansion in an arbitrary spatial dimension, and closeness is generalized to simultaneous closeness of initial data and measurement test function. An alternative proof allows us to remove the small time assumption. By proving sufficient regularity of the forward model (Ch)–(iCh) in the parameter, the local boundedness assumption on the measurement Hessian in [HKLT25, Thm3.2] could be removed and equivalence of sensitivity based identifiability and strong convexity of the least squares cost is shown to hold. The suggested design under (ii) is extended to spatial dimension $d = 2$ of the model, with the obvious extension to higher dimension $d = 3$.

Related Literature. Literature on experimental design for chemotaxis mostly belongs to the realm of biology, biophysics and related areas, where it focuses on designing suitable

experimental apparatuses, so called chemotaxis assay chambers, through which chemotactic behaviour can be studied in a controlled environment, and statistics on the microscopic behaviour or density information can be extracted. Designs are for instance compared by experimental data reproducibility [MBJPM85], or the effects of different experimental setups such as the initial data preparation are investigated experimentally [EMJ10]. The latter already represents a step towards comparing experimental designs via sensitivity analysis. A model based approach in this direction was applied in [PMMA02] for an autochemotactic Keller-Segel model, where the influence of different choices of certain constants in a parametric form of the chemotactic drift coefficient on the data was investigated. Sensitivity of the model w.r.t. these constants and thus their identifiability was derived heuristically from comparisons of numerical forward simulations.

A more structured framework of sensitivity analysis was applied to a microscopic receptor model for chemotaxis in [DT11]. The sensitivity matrix was used to detect important, i.e. highly sensitive, parameters, and non-identifiability could of some coefficients could be revealed.

Moreover, regularization based cost function identifiability was shown for an inverse problem of determining the chemotactic sensitivity in the auto-chemotactic Patlak Keller Segel model, as a non-linear function of the chemoatractant concentration, from noisy data, assuming access to the macroscopic bacteria concentration [EPS15], or the bacteria and chemoattractant concentration [FM08] for all space-time.

Novelty of this Work. Studying (optimal) experimental design is novel for inverse problems related to chemotactic behaviour. In comparison to previous literature, designs in this work are characterized by a finite number of localized measurement test functions, which reduces the experimental measurement effort. Sensitivity analysis is used to test suitability of different designs, and a specific design is constructed for which local identifiability can be proven analytically. Similar constructions have not been proposed yet in the realm of chemotaxis.

The construction relies on the relaxation of theoretical findings and constructions in Chapter 4 and emphasizes the importance of the choice of the admissible set for experimental design. The idea of transferring the theoretical identifiability or well-posedness proof to the practical finite data level through a relaxation is not frequently adopted in inverse problems so far, but it has the potential to connect theory and numerics of inverse problems. It should be mentioned, however, that the thus presented construction largely benefits from the constructive nature of the theoretical proof in the singular decomposition approach.

5.1. General Setting

This section introduces the finite dimensional inverse problem under investigation and motivates the choice of the sensitivity based identifiability criterion by its connection to local strong convexity of the cost function which guarantees favourable properties of the

numerical optimization based reconstruction.

Admissible Set. The finite dimensional admissible set is chosen to prescribe a fixed expansion form of the parameter, that might for instance emerge from discretization of a continuous function. It allows more flexibility in comparison to a parametrized form based on physical insight [RHL86, TZL88, FL91, WECW97], which might not always be available. The choice of the discretization will later be specified when a specific experimental design is constructed in Section 5.2.2 to provide well-posedness. It should be mentioned, that for applied contexts, the choice of the discretization is also limited by the experimental capabilities, for instance the pixel size of the camera that takes the photograph of the cell density or the accuracy up to which an initial condition can be generated, that limit the available data and thus its reconstruction potential.

To construct the admissible set, consider a fixed finite set of linearly independent basis elements $b_q : \mathbb{R}^d \times V \times V \to \mathbb{R}$ that are uniformly bounded in L^∞ by a constant $\|b_q\|_\infty \leq C_b$ for all $q = 1, ..., Q$ and satisfy $b_q(\cdot, v, v) \equiv 0$ for all $v \in V$. Then the admissible set consists of all nonnegative K that attain an expansion into these basis elements:

$$\mathcal{A}_K^{\text{fin}} := \left\{ \mathcal{A}_K \ni K(x, v, v') = \sum_{q=1}^{Q} K_q b_q(x, v, v') \quad \text{for some } K_q \in \mathbb{R} \right\} \tag{5.1}$$

Note that a $K \in \mathcal{A}_K^{\text{fin}}$ is uniquely determined by the Q scalar coefficients $K_q \in \mathbb{R}$, and $\mathcal{A}_K^{\text{fin}}$ is of dimension Q.

Remark 5.2. In comparison to $\mathcal{A}_K^{\text{cont}}$, continuity of K is exchanged for its expansion form to guarantee a certain structure of the admissible set. This allows for instance also piecewise constant K.

Experimental Data. A finite number $L \in \mathbb{N}$ of synthetic, noise free data of form (M) is generated by a ground truth parameter $K_\star$ in the interior of $\mathcal{A}_K^{\text{fin}}$, and a predefined set of initial data $\phi_l \in L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V)$ and spatio-temporal test functions $\mu_l \in L^1([0, T] \times \mathbb{R}^d)$, for $l = 1, ..., L$, i.e.

$$y_l = M_l(f_{K_\star}) := M_{\mu_l}(f^{\phi_l}_{K_\star}), \qquad l = 1, ..., L\,. \tag{5.2}$$

Denote by $C_\phi, C_\mu > 0$ constants that uniformly bound all ϕ_l, μ_l, respectively

$$\|\phi_l\|_{L^1 \cap L^\infty(\mathbb{R}^d \times V)} \leq C_\phi, \qquad \|\mu_l\|_{L^1([0,T] \times \mathbb{R}^d)} \leq C_\mu \qquad \text{for all } l = 1, ..., L. \tag{5.3}$$

This ensures existence and uniform boundedness of f^{ϕ_l} by Proposition 3.3

$$\|f^{\phi_l}(t)\|_{L^1 \cap L^\infty(\mathbb{R}^d \times V)} \leq e^{|V| C_K t} C_\phi. \tag{5.4}$$

Inverse Problem. The inverse problem consists in finding the Q values $(K_q)_{q=1,...,Q}$ that determine $K \in \mathcal{A}_K^{\text{fin}}$:

$$\text{find} \quad K \in \mathcal{A}_K^{\text{fin}} \quad \text{such that} \quad (y_l)_l = F^{\text{fin}}(K) := (M_l(f_K))_l \in \mathbb{R}^L. \tag{5.5}$$

Using the quadratic cost function, its minimization form reads

$$\min_{K\in\mathcal{A}_K^{\text{fin}}} \quad \mathcal{C}(K) = \min \frac{1}{2L}\sum_{l=1}^{L}\left(M_l(f_K) - y_l\right)^2 \qquad (5.6)$$
$$\text{subject to} \quad \text{(Ch) and (iCh)}.$$

Strong Convexity of the Quadratic Cost. The convexity of $\mathcal{C}$ describes how strong the measurement changes in comparison to the data, when the parameter is varied. It thus resembles the expressivity of the data in terms of the parameter and is dictated by the experimental design $(\phi_l, \mu_l)_{l=1,\dots,L}$, given that structural identifiability of the inverse problem (Ch^{-1}) has already been established in Chapter 4. It will be used as a measure to compare suitability for different designs, and existence of both, identifiable and non-identifiable designs, will be established in the subsequent Section 5.2.

Strong convexity of the cost function, at least locally around the ground truth parameter $K_\star$, is particularly favourable: because $K_\star$ is by definition of the noise free data (5.2) a minimum of $\mathcal{C}$, strong convexity would guarantee its uniqueness locally around $K_\star$ and thus provide local square loss identifiability of K from the respective forward map. Furthermore, optimization of a convex problem (5.6) is significantly easier than in the non convex setting and reconstruction, even by simple gradient-based methods as described for instance in Section 2.2.1, is guaranteed to converge to the ground truth [Pol63] - given that the starting value is chosen in the convexity neighbourhood of $K_\star$. Given sufficient regularity, strong convexity of $\mathcal{C}$ expresses in the fact that the eigenvalues of the Hessian are bounded away from 0.

This regularity is established in the following lemma, and allows application of gradient based optimization methods. For the chemotaxis inverse problem (5.5), it is a direct consequence of the regularity of the forward model solution f_K w.r.t. K as given in Lemma 3.4, and linearity of M_l.

Lemma 5.3. *The quadratic cost function $\mathcal{C}$ from (5.6) and the measurements $M_l(f_K)$ are twice continuously differentiable in K. The Hessian of the cost function, given by (2.16), is Lipschitz continuous in K w.r.t. the L^∞ norm.*

Note that this in particular implies local C_H-smoothness of $\mathcal{C}$.

Connection to Sensitivity based Identifiability. The previously established regularity of the cost function establishes equivalence of local strong convexity of the quadratic cost and sensitivity based identifiability for inverse problem (Ch^{-1}) through Proposition 2.11. This facilitates the testing and comparing of different experimental designs, as the Hessian needs to be evaluated only at one point.

Proposition 5.4. *For inverse problem (5.5), sensitivity based identifiability is equivalent to local strong convexity of the quadratic cost function around the ground truth. It implies local square loss identifiability.*

Computing the Sensitivity Matrix. To compute the entries the sensitivity matrix (2.15), different strategies can be employed, each one offering its own advantages and disadvantages:

- Difference approximations of the Q partial derivatives offer an easy reconstruction formula, but might be inaccurate if the parameter difference ε is chosen too large. Furthermore, they are computationally costly, as they require at least $Q+1$ forward computations, if one-sided approximations

$$\partial_{K_q} M_l(f_K) \approx \frac{M_l(f_{K+\varepsilon e_q}) - M_l(f_K)}{\varepsilon} \tag{5.7}$$

  with unit vectors e_q in the direction of basis element b_q were used.

- By independence of the measurement operator from K, alternatively, the derivative can be pulled inside the measurement which leads to the still fairly simple formula $\partial_{K_q} M_l(f_K) = M_l(\partial_{K_q} f_K)$. The Q partial derivatives of f_K can be computed by solving their corresponding evolution equations (A.1.11), and one additional computation of f_K by (Ch)–(iCh) to compute the source term. In total, this amounts again in $Q+1$ PDE solves.

- A very efficient way of computing the gradients is given by the adjoint gradient method that requires solving only two PDEs, namely for the chemotaxis inverse problem the forward model (Ch)–(iCh) and the adjoint equation corresponding the to l-th measurement

$$-\partial_t g_{\mu_l} - v \cdot \nabla g_{\mu_l} = \mathcal{K}^\star(g_{\mu_l}) - \mu_l \tag{5.8}$$
$$g_{\mu_l}(t = T, x, v) = 0\,, \tag{5.9}$$

  with adjoint tumbling operator given by

$$\mathcal{K}^\star(g_{\mu_l}) := \int_V K(x, v', v)(g_{\mu_l}(t, x, v') - g_{\mu_l}(t, x, v))\,\mathrm{d}v'. \tag{5.10}$$

  Existence of mild solutions $g_{\mu_l} \in C^0([0,T]; L^1(\mathbb{R}^d; L^1 \cap L^\infty(V)))$ to this equation is established in Lemma A.1.9 and uniform boundedness follows from uniform boundedness of the μ_l in (5.3) as

$$\|g_{\mu_l}(t)\|_{L^1(\mathbb{R}^d \times V)} \leq |V| e^{C_K |V| (T-t)} C_\mu. \tag{5.11}$$

  The partial derivatives $\partial_{K_q} M_l(f_K)$ are then computed by means of a more involved and model dependent formula, as given for the thus considered inverse problem in the following Lemma.

The representations of the gradient entries, i.e. the partial derivatives, for the latter two cases are summarized in the following lemma.

Lemma 5.5. *Denote by f_K, $\partial_{K_q} f_K$ and g_{μ_l} the mild solution to the forward model (Ch)–(iCh), the partial derivative evolution equation (A.1.11) for variation in the q-th unit vector $\eta = e_q$, and the adjoint model (5.8)–(5.9), respectively. Then, the partial derivatives can be represented as*

$$\frac{\partial M_l(f_K)}{\partial K_q} = \iint_{[0,T]\times\mathbb{R}^d} \int_V \partial_{K_q} f_K \,\mathrm{d}v \;\mu_l \,\mathrm{d}(t,x) \tag{5.12}$$

$$= \iiiint_{[0,T]\times\mathbb{R}^d\times V\times V} b_q(x,v,v') f'(g'_{\mu_l} - g_{\mu_l}) \,\mathrm{d}(t,x,v,v')\,, \tag{5.13}$$

where the short notation from (3.2) was applied in the last line.

Proof. The representation in the first line is a direct consequence of the Leibnitz integral rule [Fol99, Thm. 2.27], whose application is justified by uniform boundedness of the partial derivative $\partial_{K_q} f_K \in C([0,T]; L^1 \cap L^\infty(\mathbb{R}^d \times V))$ by Lemma A.1.7 and finiteness of the measure $\mathrm{d}t \;\mu_l \,\mathrm{d}(t,x)$.

The proof of the second line representation follows from a calculation-of-variation argumentation and can be found in Section 5.3.1. The low regularity of the mild solutions f and g_{μ_l} does not allow a standard application of integration by parts and specific attention has to be paid when deriving the adjoint equation.

□

5.2. Qualitative Experimental Design

Ideally, an experimental design leads to data that contains information on all aspects of K and for which changes in the parameter are reflected in the data. This requires a sufficiently high dimension of the data, initial data that triggers all different aspects of K and measurement specifications μ_l that capture the respective triggered aspect of the parameter. Suitability of designs can be tested by the previously studied sensitivity identifiability criterion, under which Proposition 2.13 suggests to measure at least Q data points.

5.2.1. Non Identifiability for Close Measurement Specifications

Even if a sufficient number of data is collected, the inverse problem (5.5) can still be non-identifiable in the sensitivity sense, if this data does not represent the kernel K well.

An example is provided by the following theorem. It studies the minimal data setting and shows that identifiability degenerates when two experimental setups become asymptotically close, leading to a decay of diversity in the data. This is not surprising: Intuitively, it makes sense that if two experimental setups almost coincide, they offer almost the same information and the data lacks one degree of freedom to specify K. Asymptotically, one data point is redundant and the inverse problem becomes underdetermined.

Theorem 5.6. *Consider a minimal design consisting of $L = Q$ experimental setups $((\phi_l, \mu_l))_{l=1,\dots,L}$. Then sensitivity based identifiability of the corresponding inverse problem (5.5) and with it local strong convexity of the cost function (5.6) around $K_\star$ decay as the first experimental setup approaches the second one, in the sense that $\phi_1 \to \phi_2$ strongly in $L^1 \cap L^\infty(\mathbb{R}^d \times V)$ and $\mu_1 \rightharpoonup \mu_2$ weakly in $L^1([0,T] \times \mathbb{R}^d)$.*

Convergence in the measurement test function holds in a distributional sense:

Definition 5.7. A sequence $\mu^{(m)}$ of $L^1([0,T] \times \mathbb{R}^d)$ functions is said to converge weakly to $\mu \in L^1([0,T] \times \mathbb{R}^d)$, if for any test function $h \in L^\infty([0,T] \times \mathbb{R}^d)$ the duality product converges $\int_0^T \int_{\mathbb{R}^d} \mu^{(m)}(t,x)h(t,x)\,\mathrm{d}x\,\mathrm{d}t \to \int_0^T \int_{\mathbb{R}^d} \mu(t,x)h(t,x)\,\mathrm{d}x\,\mathrm{d}t$.

Remark 5.8. Similar weak notions of closeness are required when working with lower regularity measurement test functions such as measures, e.g. represented by Dirac delta function corresponding to point wise measurements. It is expected that the result generalizes to this pointwise measurement case, if additional regularity of the forward solution f is established to provide sufficient measurability.

In the limiting case, convergence of gradient descend methods to reconstruct $K_\star$ can then no longer be guaranteed. In fact, a significant decay of their reconstruction can already be observed in the case where two measurements are close but distinct, as numerical experiments in Section 6.2.2 demonstrate.

It should be mentioned that the result does not prove non-convexity of the quadratic cost. In fact, local strict convexity and thus cost function identifiability could still hold in certain cases. In this case, gradient descend methods would still converge to the true parameter, however, convergence would be slowed down by the vanishing eigenvalue of the Hessian at the ground truth parameter. To be precise, convergence would hold at linear instead of an exponential rate [WR22, Thm.3.3]. Numerical experiments in Section 6.2.2, however, demonstrate actual degeneracy in the sense that the cost function becomes flat in one direction.

The proof of Theorem 5.6 is laid out in Section 5.3.2. It amounts in proving the increasing alignment of the gradients $\nabla_K M_1(f_{K_\star})$ with $\nabla_K M_2(f_{K_\star})$ with the help of formula (5.12), which results in the Hessian at ground truth

$$H_K\mathcal{C}(K_\star) = L^{-1}J^TJ = L^{-1}\sum_{l=1}^{L} \nabla_K M_l(f_{K_\star}) \otimes \nabla_K M_l(f_{K_\star}),$$

by (2.17), lacking at least one rank and thus a collapse of the objective function landscape.

The results of the previous two subsections are not surprising and apply to many other inverse problems. Often they are understood as common knowledge and not rigorously justified.

5.2.2. An Identifiable Design in Dimension $d = 1$

In contrast to the previous section where non identifiability even under a sufficient number of data was considered, in this section one minimal data ($L = Q$) design is constructed

that generates an identifiable inverse problem for a compatibly chosen admissible set. This proves existence of identifiable minimal designs.Together with the previous Theorem 5.6 on non identifiability, this demonstrates that the preparation of the experiment, in combination with the choice of the admissible set, is crucial to the reconstruction success.

The construction of the design is based on a relaxation of the singular decomposition construction in the theoretical identifiability proof from Section 4.3. In this section, the $d = 1$ dimensional case is considered, that allows for a simplified design based on the ballistic part of f, which facilitates the analytical identifiability analysis.

Relaxing the Design from Section 4.3.

Recall that in the proof of structural identifiability in Section 4.3, point values of K were reconstructed by a sequence of increasingly singular initial data and spatio-temporal measurement test functions centered around the ballistic location $x_b = x_{\text{in}} + v_{\text{in}} t_{\text{m}}$, with asymptotically vanishing measurement time t_{m}. Asymptotically, the the experimental geometry, i.e. initial and measurement location, concentrate at the reconstruction location, such that the data only experienced the value of K at that location.

This is not feasible experimentally, where a finite number of experimental setups $(\phi_l, \mu_l)_{l=1}^L$ with positive measurement times and bounded bacterial density needs to be considered. Under such designs, only a finite dimensional version of K will be sensitivity based identifiable according to Proposition 2.13.

The choice of the parametric form of K will be informed by the theoretical results in Section 4.3: To allow a relaxation of the short time singular design, the idea is to extend the domain in which K attains a certain value by choosing a piecewise constant form. This way particles can travel for a short time while still only observing one value of the tumbling kernel. Note that this form of K emerges naturally when approximating a continuous function by a step function.

Admissible set. As the velocity space $V = \{\pm 1\}$ is already discrete, it remains to discretize K in space. To do so, prescribe a partition of $\mathbb{R} = \bigcup_{r=0}^{R-1} I_r$ into $R \in \mathbb{N}$ intervals

$$\mathbb{R} = \dot{\bigcup_{r=0}^{R-1}} I_r \quad \text{with} \quad I_0 = (-\infty, a_1), I_r = [a_r, a_{r+1}) \quad \text{for fixed } a_1 < \ldots < a_{R-1} < a_R = \infty. \tag{5.14}$$

We then restrict ourselves to tumbling kernels $K(x, v, v')$ that are piecewise constant in space in these interval and denote their constant value on I_r by $K_r^{v'} = K(x, -v', v')$ for all $x \in I_r$, $v' \in V$. The admissible set collects all such K that are non-negative and uniformly bounded by a constant C_K:

$$\mathcal{A}_K^{\text{pwc}} = \Big\{ K \in \mathcal{A}_K \;\Big|\; K(x, v, v') = \sum_{r=0}^{R-1} \mathbb{1}_{I_r}(x) \mathbb{1}_{-v'}(v) (K_r^{+1} \mathbb{1}_{+1}(v') + K_r^{-1} \mathbb{1}_{-1}(v')) \tag{5.15}$$
$$\text{for some } K_r^{\pm 1} \in [0, C_K] \Big\}$$

where $\mathbb{1}_A$ again denotes the characteristic function of a set A. Note that K has $Q = 2R$ free parameters $K_r^{\pm 1} \in [0, C_K]$. This choice of $\mathcal{A}_K^{\text{pwc}}$ represents a special case of $\mathcal{A}_K^{\text{fin}}$ for

the choice of basis elements

$$b_q(x,v,v') = \mathbb{1}_{I_r}(x)\mathbb{1}_{-v'}(v)\mathbb{1}_{\pm 1}(v') \quad \text{with } q = 1 + 2r + (1 \pm 1)/2. \tag{5.16}$$

Experimental Design. This piecewise constant form of K allows sufficiently singular initial data to travel for a small positive time within the domains I_r still only experiences the local value of K. This is the basic idea behind the following design:

Design (1D). *Consider the partition of* $\mathbb{R}$ *from* (5.14) *and define the middle points* $a_{r+\frac{1}{2}} := \frac{a_r + a_{r+1}}{2}$ *for* $r = 1, ..., R-2$ *and the grid size* $\alpha_{\min} := \min_{r=1,...,R-2}(a_{r+1} - a_r)$ *and set* $a_{\frac{1}{2}} := a_1 - \frac{\alpha_{\min}}{2}$ *and* $a_{R-1/2} := a_{R-1} + \frac{\alpha_{\min}}{2}$. *Furthermore, let* $\psi_x, \psi_t \in L^1 \cap L^\infty(\mathbb{R})$ *be nonnegative, symmetric functions that are monotonously decreasing in* $|x|$ *and compactly supported in the domain* $[-\varepsilon, \varepsilon]$ *for some* $\varepsilon > 0$. *Then consider an experiment consisting of* $2R$ *sub-experiments with*

- *initial conditions* $\phi_r(x,v) = \psi_x(x - a_{r+1/2}) \in L^1_+ \cap L^\infty(\mathbb{R})$ *supported in* $a_{r+1/2} + [-\varepsilon, \varepsilon]$, *and*
- 2 *measurement test functions* $\mu_r^{\pm}(t,x) = \psi_t(t - t_{\mathrm{m}})\psi_x(x - a_{r+1/2} \mp t_{\mathrm{m}})$ *that are centered in time around a measurement time* t_{m} *and in space around the ballistic locations* $x^{\pm}_{b,r} = a_{r+1/2} \pm t_{\mathrm{m}}$, *each,*

for $r = 0, ..., R-1$, *where* ε *is balanced against* t_{m} *through the relation*

$$0 < 4\varepsilon < t_{\mathrm{m}} < T - \varepsilon \quad \text{and} \quad t_{\mathrm{m}} + 2\varepsilon < \frac{\alpha_{\min}}{2}. \tag{5.17}$$

This represents a minimal design, as the dimension of the data $2R$ coincides with the dimension of the parameter $K \in \mathcal{A}_K^{\mathrm{pwc}}$. Figures 5.1 and 6.1 visualize the geometries of Design (1D).

Typical choices of ψ_x, ψ_t include characteristic functions as pixel readings, hat functions or cut-off Gaussians to mimic a detector profile.

The connection to the design in Section 4.3 is rather obvious:

- Localized initial data and measurement test functions approximate the singular data in the singular decomposition construction.
- A small time $0 < t_{\mathrm{m}} < \frac{\alpha_{\min}}{2} - 2\varepsilon$ is fixed around which the temporal component of the measurement test function centers. By $t_{\mathrm{m}} < T - \varepsilon$, its domain is fully contained in $[0, T]$.
- Prescribing the smallness of the support of the initial data and measurement test function in comparison to the measurement time $4\varepsilon < t_{\mathrm{m}}$ by (5.17) ensures that the ballistic data has traveled at least a distance of 2ε away from the spatial support of the initial data, before it is measured. It has the same effect as the slower speed at which measurement time vanishes in comparison to the generation of the singularity in Section 4.3. Transport and thus tumbling is required to collect information on K.

- The measurement location is set to the ballistic location $x_b = x_{\text{in}} + v_{\text{in}} t_{\text{m}}$. This indicates that the ballistic part f_0 of f is measured, which contains enough information to recover K in the $1D$ scenario according to Section 4.3.
- Finally, by putting the initial location in the middle of I_r and choosing t_{m} small according to (5.17), the constant speed of propagation in (Ch) provides

$$\operatorname{supp} f^{\phi_r}(t,\cdot,v) \subset a_{r+1/2} + [-\varepsilon - t, \varepsilon + t] \subset I_r, \qquad \text{for all } t \leq t_{\text{m}} + \varepsilon, v \in V, \quad (5.18)$$

  as explained in detail in Lemma A.1.5b, i.e. the data measured by $\mu_r^{\pm}$ has in fact not interacted with values of $K_s^{\pm 1}$ from any other interval I_s with $s \neq r$, and measurement is taken at the location where we want to recover K. This also justifies why boundary conditions can be neglected. Moreover, it shows that the experiment lives on a finer spatial scale than the parameter discretization.

Remark 5.9. Obvious generalizations of Design (1D) include the choice of unequal lengths of the supports of measurement test functions and initial data, as well as individual designs on every interval I_r, for instance through different choices of initial data, measurement test functions or measurement time, as long as the these are reasonably balanced through an analogon of (5.17).

Inverse Problem. The inverse problem then reads

$$\text{find} \quad K \in \mathcal{A}_K^{\text{pwc}} \quad \text{such that} \quad (y_r^i)_{\substack{r=1,\dots,R \\ i\in\{+,-\}}} = F^{1D}(K) := (M_{\mu_r^i}(f_K^{\phi_r}))_{\substack{r=0,\dots,R-1 \\ i\in\{+,-\}}}. \quad (5.19)$$

Properties of Design (1D)

Because each datum $y_r^{\pm} = M_{\mu_r^{\pm}}(f_{K_\star}^{\phi_r})$ is only affected by the values of K in I_r, the reconstruction of $K_r^{\pm 1}$ only relies on these two data points. The experiments and the reconstruction of the $2R$ values of K is thus decoupled into R smaller cell problems of reconstructing 2 the local value $K_r^{\pm 1}$ in the intervals I_r. This opens the door for parallelization in experimentation as well as in the computation. The simplification in the inversion comes with the cost of prescribing very detailed experiments, involving tightly controlled initial data and very fine measurements, in dependence on the discretization scales of K through the size of the intervals I_r. Proposition 5.10 describes this property mathematically.

Proposition 5.10. *Let $v', v'' \in V = \{\pm 1\}$ and $r, s \in \{0, \dots, R-1\}$ be distinct. Under Design (1D), the reconstruction of $K_r^{v'}$ is decoupled from that of $K_s^{v''}$, in the sense that measurements taken in interval I_r are not sensitive to parameter values $K_s^{\pm 1}$ in other intervals I_s. All entries of the gradient $\nabla_K M_{\mu_r^{v'}}(f_K^{\phi_r})$ vanish, except for those two correspond to the values $K_r^{\pm 1}$.*

Proof. By choice of the basis b_q in (5.16), the gradient formula (5.13) reduces to

$$\frac{\partial M_{\mu_r^{\pm}}(f_K^{\phi_r})}{\partial K_s^{v'}} = \iint_{[0,T]\times I_s} f^{\phi_r}(t,x,v')(g_{\mu_r^{\pm}}(t,x,v') - g_{\mu_r^{\pm}}(t,x,-v'))\,\mathrm{d}(t,x)\,. \quad (5.20)$$

For $s \neq r$, then (5.18) shows that f^{ϕ_r} is not supported in I_s for $t \leq t_m + \varepsilon$, and $g_{\mu_r^\pm}(t)$ vanishes for $t > t_m + \varepsilon$ by Corollary A.1.10, i.e. the respective gradient entry vanishes. □

This provokes a block diagonal structure in the sensitivity matrix, whose rows are exactly the gradients, and in the Gauss-Newton $H_K\mathcal{C}(K_\star)$ according to (2.17).

Corollary 5.11. *Under Design (1D), the sensitivity matrix J and the Gauss-Newton Hessian $H_K\mathcal{C}(K_\star)$ attain a block diagonal structure. The 2×2 blocks of the sensitivity matrix are given by the sub-sensitivity matrices*

$$J_{K_r} = (\nabla_{K_r} M_{\mu_r^+}, \nabla_{K_r} M_{\mu_r^-})^T. \tag{5.21}$$

Most importantly, Design (1D) ensures local identifiability of inverse problem (5.19), if the measurement times t_m are chosed sufficiently small.

Theorem 5.12. *Consider the inverse parameter identification problem (5.19) of finding $K \in \mathcal{A}_K^{\text{pwc}}$ from measurements generated by Design (1D), for which*

$$\psi_x = \psi_t = \mathbb{1}_{[-\varepsilon,\varepsilon]}, \tag{5.22}$$

$$C_K(t_m + \varepsilon) < 1, \qquad \textit{and} \tag{5.23}$$

$$3e^{-(t_m+\varepsilon)C_K} t_m - 12\varepsilon - 32\max(1, 2\varepsilon)(t_m + \varepsilon)\tfrac{C_K(t_m+\varepsilon)}{1-C_K(t_m+\varepsilon)}\left(\tfrac{1}{1-C_K(t_m+\varepsilon)} + 1\right) > 0. \tag{5.24}$$

Then inverse problem (5.19) is sensitivity based and least squares identifiable and the quadratic cost function is locally strongly convex around the ground truth parameter $K_\star$.

Because the proof is based on the sensitivity matrix, only local square loss identifiability can be guaranteed. However, the proof shows that local identifiability holds uniformly over the admissible set, which means that the design is uniformly persistently exciting. Numerical experiments in Section 6.2 even suggest global identifiability.

Remark 5.13 (About assumptions (5.22)–(5.24)). Characteristic functions for ψ_x, ψ_t are considered purely for readability of the proof. The result generalizes to other forms of test functions ψ_x, ψ_t, when the assumptions (5.22)–(5.24) are adapted accordingly, as the derivations in [HKLT25] show.

Assumption (5.23) explicitly expresses smallness of t_m in comparison to C_K. This poses the model in the rare tumbling regime and serves to control the high regularity parts of the forward and adjoint solution in the gradient according to (5.13), in analogy to the theoretical proofs in Chapter 4.

The complicated form of assumption (5.24) originates from estimation in the proof, where it eventually guarantees positivity of the minimal eigenvalue of J^TJ and thus identifiability. It relates t_m, ε and C_K and postulates smallness of the measurement time t_m, as a limiting consideration shows: consider ε that is a fixed fraction $\varepsilon = \lambda t_m$ of the measurement time for some $\lambda \in (0, \frac{1}{4})$ as suggested by (5.17). Then the left hand side of (5.24) is of order

$$\begin{aligned} &t_m \left(3e^{-(1+\lambda)t_m C_K} - 12\lambda - 32\max(1, 2\lambda t_m)(1+\lambda)\tfrac{(1+\lambda)t_m C_K}{1-(1+\lambda)t_m C_K}\left(\tfrac{1}{1-(1+\lambda)t_m C_K} + 1\right)\right) \\ &= t_m(3 - 12\lambda + \mathcal{O}(t_m)) \end{aligned}$$

as $t_\mathrm{m} \to 0$. With $\lambda < \frac{1}{4}$, this shows that (5.24) holds for small enough t_m. One suitable choice is for instance given by $t_\mathrm{m} \leq 0.02 C_K^{-1}$ and $\varepsilon \leq \min(\frac{1}{2}, 0.1 t_\mathrm{m})$.

The rough estimation that we used in the the proof lets us expect that the bound (5.24) can be improved. This path is not followed in this work in order to concentrate on the main arguments and in favour of readability of the proof.

By Proposition 5.4, proving Theorem 5.12 amounts in proving sensitivity based identifiability $J^T J > 0$, which is equivalent to the fact that J attains full rank Q. The block diagonal structure of J according to Corollary 5.11 suggests that this is equivalent to the fact that all 2×2-blocks J_{K_r}, $r = 0, ..., R-1$, in (5.21) are of full rank. By symmetry in the design, the remaining proof concentrates without loss of generality (w.l.o.g.) on J_{K_1}.

Proposition 5.14. *Under assumptions (5.22)–(5.24), Design (1D) generates a full rank sub-sensitivity matrix J_{K_1}.*

It is easy to see that J_{K_1} is full rank if its entries satisfy the following inequalities

$$\left| \frac{\partial M_{\mu_1^+}(f_{K_\star}^{\phi_1})}{\partial K_1^{+1}} \right| > \left| \frac{\partial M_{\mu_1^+}(f_{K_\star}^{\phi_1})}{\partial K_1^{-1}} \right| \qquad \text{and} \qquad \left| \frac{\partial M_{\mu_1^-}(f_{K_\star}^{\phi_1})}{\partial K_1^{+1}} \right| < \left| \frac{\partial M_{\mu_1^-}(f_{K_\star}^{\phi_1})}{\partial K_1^{-1}} \right|. \tag{5.25}$$

It is thus the goal of the proof of Proposition 5.14 to verify these inequalities. The symmetry in Design (1D) suggests that it is sufficient to study the first inequality.

Figure 5.1 helps building an intuition why this inequality should be true: in formula (5.20) for the partial derivative, consider for a moment only the ballistic parts f_0 of f^{ϕ_1} and g_0 of $g_{\mu_1^+}$ that solve, respectively, (4.32) with initial condition $f_0(t=0) = \phi_1$, and

$$-\partial_t g_0 - v \partial_x g_0 = -K(x, v', v) g_0 - \mu_1^+, \qquad g_0(t = T) = 0. \tag{5.26}$$

Existence of f_0 and g_0 is established in Lemmas 4.5 and 5.18. Then the ballistic part of the first inequality in (5.25) reads $B > 0$ with

$$B := \left| \int_0^T \int_{I_1} f_0(v')(g_0(v') - g_0(v)) \,\mathrm{d}x \,\mathrm{d}t \right| - \left| \int_0^T \int_{I_1} f_0(v)(g_0(v) - g_0(v')) \,\mathrm{d}x \,\mathrm{d}t \right| \tag{5.27}$$

for $(v, v') = (-1, +1)$. Positivity of B can be anticipated by a comparison of the supports of f_0 and g_0, as depicted in Figure 5.1: the supports of $f_0(v')$ and $g_0(v')$ overlap initially and due to the same direction and speed of transport, they overlap for the whole travel time $t \in [0, t_\mathrm{m} + \varepsilon]$ until $g_0(v')$ vanishes. In contrast to that, $f_0(v')$ overlaps with $g_0(v)$ only for a very short period of time when $t \approx t_\mathrm{m}$, whose length is determined by ε. Due to positivity of f_0, g_0, the value of the first summand is thus supposed large. In contrast to that, considering the second summand, the spatial supports of $f_0(v)$ and $g_0(v)$ never overlap, and only a small contribution of the product of $f_0(v)$ and $g_0(v')$ is expected, given that their supports overlap only for a short time at the beginning when $t \approx 0$, determined again by ε. The size of the second summand is thus expected to be small, which shows positivity of B in total, if measurement time and the support parameter ε are sufficiently balanced to control the contribution of oppositely facing f_0 and g_0. It remains to show that the remaining non ballistic parts do not destroy this positivity.

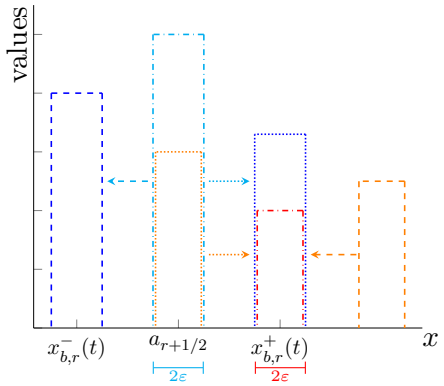


Figure 5.1.: Motion of the ballistic parts $f_0(t=0,v)$ (cyan, dashdotted) centered around $a_{r+1/2}$ to $f_0(t=t_\mathrm{m},v=+1)$ (blue, dotted) and $f_0(t=t_\mathrm{m},v=-1)$ (blue, dashed) centered around the ballistic locations $x^\pm_{b,r}=a_{r+1/2}\pm t_\mathrm{m}$ and $g_0(t=0,v=+1)$ (orange, dotted) and $g_0(t=0,v=-1)$ (orange, dashed) to $g_0(t=t_\mathrm{m},v)$ (red, dashdotted).

This intuition is made rigorous in Section 5.3.3, first proving that the difference of the ballistic parts of the partial derivatives B is bounded away from zero, before the contribution of the remaining high regularity part is bounded smaller than this distance.

5.2.3. Extension to $d=2$

In higher dimensions, for demonstration $d=2$, the same relaxation strategy can be applied to construct a design in analogy to the theory presented in Section 4.2. The strategy can readily be extended to $d=3$.

Admissible set. The reconstruction of K in Section 4.2 is point wise in space and in the velocities, therefore, we adopt a piecewise constant form of K in space x, and in the incoming and outgoing velocities v',v to relax the singular design.

Hence, prescribe partitions

- of $\mathbb{S}^1=\bigcup_{j=0}^{N_v-1}S_j$ into uniform segments $S_j=\{(\cos\zeta,\sin\zeta)\mid\zeta\in 2\pi\frac{j}{N_v}+[-\frac{2\pi}{2N_v},\frac{2\pi}{2N_v})\}$ of the unit circle, as well as
- of $\mathbb{R}^2=\bigcup_{r_1,r_2=0}^{R-1}A_{(r_1,r_2)}$ into, for simplicity, two dimensional intervals $A_{(r_1,r_2)}=I_{r_1}\times I_{r_2}$ for intervals I_r described by a fixed partition $-\infty<a_1<\ldots<a_{R-1}<a_R=\infty$ of $\mathbb{R}$ as in (5.14).

The admissible set is then constructed as

$$\mathcal{A}_K^{\mathrm{pwc}}=\left\{\mathcal{A}_K\ni K(x,v,v')=\sum_{i,j=0}^{N_v-1}\sum_{\substack{r_1,r_2=0\\ r=(r_1,r_2)}}^{R-1}K_r^{j,i}\mathbb{1}_{S_j}(v)\mathbb{1}_{S_i}(v')\mathbb{1}_{A_r}(x)\;\middle|\;0\le K_r^{j,i}\le C_K\right\}. \tag{5.28}$$

Experimental Design. The following design is inspired by the singularity inducing construction in Section 4.2.

Design (2D). *The middle points $a_{(r_1,r_2)+1/2} = (a_{r_1+1/2}, a_{r_2+1/2})^T$ of the intervals A_{r_1,r_2} shall be defined through the one dimensional interval middle points $a_{r_i+1/2}$ as defined in Design (1D), for $0 \leq r_1, r_2 \leq R-1$, and let again $\alpha_{\min} := \min_{r=1,\dots,R-2}(a_{r+1} - a_r)$ be the grid size. Furthermore, let v_j be the middle point of segment S_j. Consider non negative, radially symmetric functions $\Psi_x \in L^1 \cap L^\infty(\mathbb{R}^2)$ and $\psi_t \in L^1 \cap L^\infty(\mathbb{R})$ that are monotonously decreasing in $\|x\|$ and supported in the balls $B^{(2)}(0,\varepsilon)$ and $B^{(1)}(0,\varepsilon)$ of dimension 2 and 1, respectively, for some $\varepsilon > 0$.*

Then define for $r = (r_1, r_2) \in \{0, R-1\}^2$ and $i \in \{0, N_v - 1\}$ the (r,i)-th sub-experiments through

- *an initial condition $\phi_r^i(x,v) = \Psi_x(x - a_{r+1/2})\mathbb{1}_{S_i \cap B^{(2)}(v_i,\delta)}(v)$ for some $\delta > 0$, that is centered around the middle points a_r of the 2D intervals A_r in space and in velocity around the middle points of the velocity range S_i, and*
- *measurement test functions*

$$\mu_r^{j,i}(t,x) = \psi_t(t - t_{\mathrm{m}})\Psi_x\left(x - \left(a_{r+1/2} + \frac{t_{\mathrm{m}}}{2}v_i + \frac{t_{\mathrm{m}}}{2}v_j\right)\right), \qquad \text{for } j \in \{0, N_v - 1\}\backslash\{i\}$$

centered around the measurement time t_{m} and in space around the scattered location $a_{(r_1,r_2)+1/2} + \frac{t_{\mathrm{m}}}{2}v_i + \frac{t_{\mathrm{m}}}{2}v_j$,

where t_{m}, ε and δ satisfy

$$0 < 5\varepsilon < t_{\mathrm{m}} + \varepsilon < T \qquad \text{and} \qquad t_{\mathrm{m}} + 2\varepsilon < \frac{\alpha_{\min}}{2}, \qquad \text{and} \tag{5.29}$$

$$\frac{t_{\mathrm{m}}}{2}\left(1 - \cos\left(\frac{2\pi}{N_v}\right)\right)\left(\frac{t_{\mathrm{m}}}{2}\left(1 - \cos\left(\frac{2\pi}{N_v}\right)\right) - 2(t_{\mathrm{m}} + \varepsilon)\delta\right) > 4\varepsilon^2. \tag{5.30}$$

Typical choices for Ψ_x, ψ_t are again characteristic functions, hat functions, or cut off Gaussians of the corresponding dimension. Unlike Design (1D), Design (2D) prescribes an initial direction of travel for the particles by the velocity dependency of the initial data. This becomes necessary in the two dimensional setting, where the velocity space $V = \mathbb{S}^1$ is continuous.

Assumption (5.29) shares the same purpose as (5.17) in the one dimensional case: Note in particular, that smallness of t_{m} w.r.t. the grid size $\alpha_{\min}$ in the second inequality again ensures that the measured data in the r-th experiment does not leave the domain A_r of constant value of K, due to the finite speed of propagation in Lemma A.1.5b

$$\operatorname{supp} f^{\phi_r^i}(t,\cdot,v) \subset B^{(2)}(a_{r+1/2}, \varepsilon + t) \subset A_r, \qquad \text{for all } t \leq t_{\mathrm{m}} + \varepsilon, v \in V, i \in \{0, \dots, N_v - 1\}. \tag{5.31}$$

This entails the same local decoupling of the inverse problem, as observed already for Design (1D) in Proposition 5.10 and Corollary 5.11.

Proposition 5.15. *Under Design (2D), the reconstruction of $K_r^{j,i}$ is decoupled from that of $K_s^{k,l}$ for all $r \neq s$ and $i, j, k, l \in \{0, \dots, N_v - 1\}$, in the sense that measurements taken*

inside A_r are not sensitive to parameter values $K_s^{k,l}$ in other areas A_s. All entries of the gradient $\nabla_K M_{\mu_r^{j,i}}(f_K^{\phi_r^i})$ vanish, except for those correspond to the values $K_r^{k,l}$ for distinct $k,l \in \{0,...,N_v-1\}$.

Hence, the sensitivity matrix J and the Gauss-Newton Hessian $H_K\mathcal{C}(K_\star)$ attain block diagonal structures, where the $N_v(N_v-1) \times N_v(N_v-1)$ blocks of the sensitivity matrix are given by the sub-sensitivity matrices $J_{K_r} = (\nabla_{K_r} M_{\mu_r^{j,i}})_{j,i}^T$.

In contrast to Design (1D), the measurement location of Design (2D) centers around the scattered instead of the ballistic location. Together with the additional assumption (5.30), this excludes the ballistic part f_0 of the solution $f^{\phi_r^i}$ from the measurement, as shown in the subsequent Lemma 5.16. This is necessary in the higher dimensional case, where f_0 does not contain sufficient information for the recovery of the tumbling kernel, as explained in Section 4.2, but could distort - especially numerical - reconstructions, as it exceeds the size of the desired one tumbling part f_1 by one order of magnitude in $TC_K|V|$.

Lemma 5.16. *Consider Design (2D) and choose $r \in \{0,...,R-1\}^2$ and $i \in \{0,...,N_v-1\}$ arbitrary, but fixed. Then the ballistic part f_0 of $f^{\phi_r^i}$ does not contribute to any measurement in the (r,i)-th sub-experiment, i.e. $M_{\mu_r^{j,i}}(f_0) = 0$ for all $j \neq i$.*

The argument is based on a lack of overlap of the supports of the ballistic part f_0 and the measurement test function, just as the proof in Section 4.2.1.

Proof. Recall that by construction of Design (2D),

$$\begin{aligned}
&\operatorname{supp} \mu_r^{j,i} = [t_\mathrm{m}-\varepsilon, t_\mathrm{m}+\varepsilon] \times B^{(2)}\left(a_{r+1/2} + \frac{t_\mathrm{m}}{2}v_i + \frac{t_\mathrm{m}}{2}v_j, \varepsilon\right), \\
&\operatorname{supp} f_0(t,\cdot,v) = B^{(2)}(a_{r+1/2}+tv,\varepsilon), && \text{for } t \geq 0, v \in S_i \cap B^{(2)}(v_i,\delta), \text{ and} \\
&f_0(\cdot,\cdot,v) \equiv 0, && \text{for } v \notin S_i \cap B^{(2)}(v_i,\delta),
\end{aligned}$$

by the explicit formula (4.12) for f_0. The spatial supports of $\mu_r^{j,i}$ and f_0 do not overlap for any $t \in [t_\mathrm{m}-\varepsilon, t_\mathrm{m}+\varepsilon]$, given that for all $v = v_i + w \in S_i \cap B^{(2)}(v_i,\delta)$ with $\|w\| \leq \delta$, the Cauchy Schwarz inequality shows

$$\begin{aligned}
&\left\| a_{r+1/2} + \frac{t_\mathrm{m}}{2}v_i + \frac{t_\mathrm{m}}{2}v_j - (a_{r+1/2}+tv) \right\|^2 = \left\| \frac{t_\mathrm{m}}{2}v_i + \frac{t_\mathrm{m}}{2}v_j - t(v_i+w) \right\|^2 \\
&\geq \left\| \left(t - \frac{t_\mathrm{m}}{2}\right) v_i - \frac{t_\mathrm{m}}{2}v_j \right\| \left(\left\| \left(t - \frac{t_\mathrm{m}}{2}\right) v_i - \frac{t_\mathrm{m}}{2}v_j \right\| - 2t\delta \right) \\
&\geq \frac{t_\mathrm{m}}{2}\left(1 - \cos\left(\frac{2\pi}{N_v}\right)\right)\left(\frac{t_\mathrm{m}}{2}\left(1-\cos\left(\frac{2\pi}{N_v}\right)\right) - 2(t_\mathrm{m}+\varepsilon)\delta\right) && > 4\varepsilon^2
\end{aligned}$$

by assumption (5.30). In the last line, it was used that $\langle v_i, v_j \rangle \leq \cos(2\pi/N_v)$ for $i \neq j$, and thus by a simple minimization

$$\left\|\left(t-\tfrac{t_\mathrm{m}}{2}\right)v_i - \tfrac{t_\mathrm{m}}{2}v_j\right\| \geq \sqrt{\left(t-\tfrac{t_\mathrm{m}}{2}\right)^2 + \left(\tfrac{t_\mathrm{m}}{2}\right)^2 - t_\mathrm{m}\left(t-\tfrac{t_\mathrm{m}}{2}\right)\cos\left(\tfrac{2\pi}{N_v}\right)} \geq \tfrac{t_\mathrm{m}}{2}\left(1-\cos\left(\tfrac{2\pi}{N_v}\right)\right).$$

□

Tuples $t_\mathrm{m}, \varepsilon, \delta$ that satisfy the assumptions (5.29)–(5.30) exist: approximating small $\delta \approx 0$, (5.30) becomes $t_\mathrm{m}(1-\cos(2\pi/N_v)) > 4\varepsilon$. Therefore, after fixing t_m and ε accordingly, δ can be chosen sufficiently small such that (5.30) is fulfilled.

The similarities to Design (1D) and the construction in Section 4.2, as well as numerical experiments in Section 6.3, let us conjecture that Design (2D) might also yield a locally or even globally identifiable inverse problem, potentially under complementing the above assumptions by further smallness requirements on t_m w.r.t. C_K in order to place the experiment in the rare tumbling regime.

Analytical proofs of identifiability in analogy to Theorem 5.12 are left for future work, as derivations are expected to become even more technical, given that the design is based on measuring the one-tumbling part f_1 instead of the ballistic part f_0. The proof can again take advantage of the block diagonal structure of J, in the sense that the full rank property has to be established only for one block. Then a decomposition of the forward and adjoint solutions according to regularity may provide a similar size structure in the partial derivatives $\partial_{K_1^{j,i}} M_{\mu_1^{k,l}}$, where the construction of Design (2D) in analogy to the construction in Section 4.2 suggests that $\partial_{K_1^{j,i}} M_{\mu_1^{j,i}}$ will be the dominant entries.

5.3. Proofs of this Chapter

This section collects the proofs from the previous chapter that were postponed.

5.3.1. Proof of Lemma 5.5

Because the first representation (5.12) of the partial derivative of the measurement w.r.t. the parameter is a direct consequence of the Leibnitz integral rule, only the adjoint gradient representation (5.13) is proven in the following.

Proof of representation (5.13). Consider the l-th data as a measurement of a generic function f, under the constraint that $f = f_K$ is the mild solution to (Ch)–(iCh), i.e. that (A.1.8) holds. The corresponding Lagrangian function is defined as

$$\begin{aligned}\mathbb{L}: \quad & \mathcal{A}_K^{\mathrm{fin}} \times C^0([0,T]; L^1_{+,c} \cap L^\infty(\mathbb{R}^d \times V)) \times L^1([0,T] \times \mathbb{R}^d \times V) \to \mathbb{R}, \\ & \mathbb{L}(K,f,h) = M_{\mu_l}(f) + \left\langle h, f(t,x,v) - \mathsf{T}(t)\phi - \int_0^t \mathsf{T}(t-s)\mathcal{K}_K(f)(s)\,\mathrm{d}s \right\rangle_{t,x,v},\end{aligned}$$

where the notation $\mathcal{K}_K := \mathcal{L} - \sigma$ explicitly expresses the K dependence, $(\mathsf{T}(t))_t$ is the transport semigroup from Lemma A.1.2 such that the right hand side of the duality bracket $\langle \cdot, \cdot \rangle_{t,x,v}$ w.r.t. integration in variable t, x, v describes exactly the mild solution form for f. Choosing $f = f_K$ as the mild solution to (Ch)–(iCh), one obtains $\mathbb{L}(K, f_K, h) = M_l(f_K)$. The gradient at a point $\hat{K}$ can thus be expressed as

$$\left.\frac{\mathrm{d}M_l(f_K)}{\mathrm{d}K}\right|_{K=\hat{K}} = \left.\frac{\partial \mathbb{L}}{\partial K}\right|_{\substack{K=\hat{K} \\ f=f_{\hat{K}}}} + \left.\frac{\partial \mathbb{L}}{\partial f}\right|_{\substack{K=\hat{K} \\ f=f_{\hat{K}}}} \left.\frac{\mathrm{d}f_K}{\mathrm{d}K}\right|_{K=\hat{K}}.$$

The multiplier $h = \hat{h}$ will be chosen such that second summand vanishes, and one is left with

$$\left.\frac{\partial M_l(f_K)}{\partial K_q}\right|_{K=\hat{K}} = \left.\frac{\partial \mathbb{L}}{\partial K_q}\right|_{\substack{K=\hat{K}, f=f_{\hat{K}},\\ h=\hat{h}}} = \left.\frac{\partial \left\langle h, -\int_0^t \mathsf{T}(t-s)\mathcal{K}_K(f)(s)\,\mathrm{d}s\right\rangle_{t,x,v}}{\partial K_q}\right|_{\substack{K=\hat{K}, f=f_{\hat{K}},\\ h=\hat{h}}}.$$

By the definition of the transport semigroup $(\mathsf{T}(t))_t$ from Lemma A.1.2, a change of variables in x and in the order of integration in s and t shows

$$\begin{aligned}
&\left\langle h, -\int_0^t \mathsf{T}(t-s)\mathcal{K}_K(f)(s)\,\mathrm{d}s\right\rangle_{t,x,v}\\
&= -\iiint_{\mathbb{R}^d\times V\times V} \int_0^T K(x,v,v') f(s,x,v') \int_s^T h(t, x+v(t-s), v) - h(t, x+v'(t-s), v')\,\mathrm{d}t\,\mathrm{d}s\,\mathrm{d}(x,v,v')\\
&= -\left\langle f, \mathcal{K}_K^\star\left(\int_t^T \mathsf{T}^\star(s-t)h(s)\,\mathrm{d}s\right)\right\rangle_{t,x,v},
\end{aligned}$$

where $(\mathsf{T}^\star(t))_{t\geq 1}$ denotes the backward transport semigroup from Lemma A.1.9, and K dependence of the adjoint tumbling operator $\mathcal{K}^\star$ from (5.10) is explicitly expressed. Formula (5.13) then follows from the second line by using the form of K prescribed by the admissible set in (5.1) and denoting $g(t,x,v) := \int_t^T h(s, x+v(s-t), v)\,\mathrm{d}s$.

It remains to compute such $\hat{g}$ corresponding to $\hat{h}$, for which $\partial\mathbb{L}/\partial f\,|_{f=f_{\hat{K}}, K=\hat{K}, h=\hat{h}} = 0$. From the last manipulation, one sees that

$$\frac{\partial \mathbb{L}}{\partial f} = \frac{\partial}{\partial f}\left\langle f, \mu_l + h - \mathcal{K}_K^\star\left(\int_t^T \mathsf{T}^\star(s-t)h(s)\,\mathrm{d}s\right)\right\rangle_{t,x,v}.$$

This always vanishes, if $h = \hat{h}$ satisfies

$$\hat{h}(t) = \mathcal{K}_K^\star\left(\int_t^T \mathsf{T}^\star(s-t)\hat{h}(s)\,\mathrm{d}s\right) - \mu_l.$$

Existence of such $\hat{h} \in L^1([0,T]\times\mathbb{R}^d\times V)$ can be established by a similar argument as in the proof of Lemma A.1.4, by proving that the source iteration is a contraction mapping w.r.t. the weighted norm $\|h\| := \|e^{4C_K|V|t}h\|_{L^1}$ on $L^1([0,T]\times\mathbb{R}^d\times V)$. Then $\hat{g}$ is continuous in time by regularity of $h \in L^1$ because $\|\hat{g}(t+\zeta) - \hat{g}(t)\|_{L^1(\mathbb{R}^d\times V)} = \|\int_t^{t+\zeta}\hat{h}(s, x+v(s-t), v)\,\mathrm{d}s\|_{L^1(\mathbb{R}^d\times V)} \to 0$ for $\zeta\to 0$. Furthermore, $\hat{g}$ satisfies

$$\begin{aligned}
\hat{g}(t,x,v) &= \int_t^T \mathsf{T}^\star(s-t)\hat{h}(s)\,\mathrm{d}s = \int_t^T \mathsf{T}^\star(s-t)\left(\mathcal{K}_K^\star\left(\int_s^T \mathsf{T}^\star(\tau-s)\hat{h}(\tau)\,\mathrm{d}\tau\right) - \mu_l\right)\mathrm{d}s\\
&= \int_t^T \mathsf{T}^\star(s-t)(\mathcal{K}_K^\star(\hat{g}(s)) - \mu_l(s))\,\mathrm{d}s.
\end{aligned}$$

Lemma A.1.9 thus indicates that $\hat{g}$ is in fact the mild solution to the adjoint problem (5.8)–(5.9). □

5.3.2. Proof of Theorem 5.6

From (2.17) it is clear that $H_K\mathcal{C}(K_\star)$ shares the same rank as J, which is determined by the number of its linearly independent rows $\nabla_K M_l(f^{\phi_l})$. In a minimal setting where $L = Q$, an increasing alignment of two gradients $\nabla_K M_1(f^{\phi_1}) \to \nabla_K M_2(f^{\phi_2})$ hence results in a loss of strong convexity.

Proposition 5.17. *As the first experimental setup converges to the second such that $\mu_1 \rightharpoonup \mu_2$ weakly in $L^1([0,T]\times\mathbb{R}^d)$ and $\phi_1 \to \phi_2$ strongly in $L^1\cap L^\infty(\mathbb{R}^d\times V)$, the corresponding measurement gradients converge in the Euclidean norm $\nabla_K M_1(f_K) \to \nabla_K M_2(f_K)$ for all $K \in \mathcal{A}_K^{\mathrm{fin}}$.*

Then Theorem 5.6 follows from the continuity of eigenvalues w.r.t. matrix perturbations.

Proof of Theorem 5.6. Let $\nabla_K M_1(f_{K_\star}) \to \nabla_K M_2(f_{K_\star})$ in the Euclidean norm. To show that the Hessian at $K_\star$ degenerates asymptotically, rewrite (2.17) as

$$J^T J = \underbrace{\left(2\nabla_K M_2\otimes\nabla_K M_2 + \sum_{l=3}^{Q}\nabla_K M_l\otimes\nabla_K M_l\right)}_{=:A} + \underbrace{(\nabla_K M_1\otimes\nabla_K M_1 - \nabla_K M_2\otimes\nabla_K M_2)}_{=:B},$$

in the short notation $M_l := M_{\mu_l}(f^{\phi_l})$. The matrix A has a rank of at most $Q-1$ and thus it has at least one vanishing eigenvalue: $\lambda_j(A) = 0$ for some j. Furthermore, $\|B\|_F \to 0$ due to $\|\nabla_K M_1(f_{K_\star}) - \nabla_K M_2(f_{K_\star})\|_2 \to 0$. Continuity of the j-th eigenvalue w.r.t. perturbation of matrix A by B [HJ85, Corr.6.3.8] then provides

$$|\lambda_j(H_K\mathcal{C}(K_\star))| = |\lambda_j(H_K\mathcal{C}(K_\star)) - \lambda_j(A)| \le \|B\|_F \to 0, \qquad \text{as } \nabla_K M_1(f_{K_\star}) \to \nabla_K M_2(f_{K_\star}).$$

This shows that asymptotically the Hessian looses one rank, it thus lacks positive definiteness, and the strong convexity of the cost function around $K_\star$ decays. □

The increasing alignment of the gradients can be anticipated, as it is well known that the duality bracket of elements of a weakly convergent sequence μ_1 with elements of a strongly convergent sequence ϕ_1 in its dual space yields a convergent sequence in $\mathbb{R}$ [Bre11, Prop. 3.5], and the strong convergence of ϕ_1 translates to f^{ϕ_1} by (A.1.12).

Proof of Proposition 5.17. Using representation (5.12), one obtains

$$\left|\frac{\partial M_1(f^{\phi_1})}{\partial K_q} - \frac{\partial M_2(f^{\phi_2})}{\partial K_q}\right| = \left|\iiint_{[0,T]\times\mathbb{R}^d\times V} (\partial_{K_q} f^{\phi_1-\phi_2})\mu_1 + (\partial_{K_q} f^{\phi_2})(\mu_1-\mu_2)\,\mathrm{d}(t,x,v)\right|$$
$$\le |V|\|\partial_{K_q} f^{\phi_1-\phi_2}\|_\infty\|\mu_1\|_1 + \left|\iint_{[0,T]\times\mathbb{R}^d}\int_V (\partial_{K_q} f^{\phi_2})\,\mathrm{d}v\;(\mu_1-\mu_2)\,\mathrm{d}(t,x)\right|$$

by Hölder's inequality. As $\phi_1 \to \phi_2$ strongly in $L^1\cap L^\infty(\mathbb{R}^d\times V)$, then (A.1.12) shows

$$\|\partial_{K_q} f^{\phi_1-\phi_2}\|_\infty \le 2|V|e^{2|V|C_K T}T\|\phi_1-\phi_2\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)} \to 0.$$

Moreover, $\|\mu_1\|_1$ as a weakly convergent sequence is bounded by the uniform boundedness principle, compare for instance [Bre11, Prop. 3.5, Cor. 2.4]. In total, this shows that the first summand vanishes asymptotically. The same holds true by Definition 5.7 for the second summand, given that $\int_V \partial_{K_q} f^{\phi_2}\,\mathrm{d}v \in L^\infty([0,T]\times\mathbb{R}^d)$ by regularity of $\partial_{K_q} f^{\phi_2} \in C([0,T]; L^1\cap L^\infty(\mathbb{R}^d\times V))$ from Lemma A.1.7. □

5.3.3. Proof of Proposition 5.14

Existence of a mild solution g_0 to the ballistic adjoint equation (5.26) is ensured by Lemma A.1.9 and an explicit formula can be obtained from representation (A.1.16).

Lemma 5.18. *The ballistic adjoint equation (5.26) attains a unique mild solution $g_0 \in C^0([0,T]; L^1(\mathbb{R}^d\times V))$ that is explicitly given by*

$$g_0(t,x,v) = \int_t^T e^{-\int_0^{s-t}\sigma(x+v\tau,v)\,\mathrm{d}\tau}\mu(s,x+v(s-t))\,\mathrm{d}s. \tag{5.32}$$

The approximation of the measurement gradient by its ballistic part then allows estimation of the gradient difference by

$$\left|\frac{\partial M_{\mu_1^+}(f^{\phi_1}_{K_\star})}{\partial K_1^{+1}}\right| - \left|\frac{\partial M_{\mu_1^+}(f^{\phi_1}_{K_\star})}{\partial K_1^{-1}}\right| \geq B - |R^{+1}| - |R^{-1}|,$$

where B as in (5.27) collects the ballistic part of the gradient difference and

$$R^{v'} = \frac{\partial M_{\mu_1^+}(f_{K_\star})}{\partial K_1^{v'}} - \int_0^T\int_{I_1} f_0(v')(g_0(v') - g_0(-v'))\,\mathrm{d}x\,\mathrm{d}t, \qquad v'\in V$$

are the remainders. The proof of the first inequality in (5.25) then consists in showing that B is sufficiently large such that its positivity is not destroyed by the remainders.

Lemma 5.19. *Denote $(v,v') = (-1,+1)$ and consider an experiment under the Design (1D) with $\psi = \mathbb{1}_{[-\varepsilon,\varepsilon]}$ as in (5.22). Then B from (5.27) satisfies*

$$B \geq e^{-(t_\mathrm{m}+\varepsilon)C_K} 3\varepsilon^2 t_\mathrm{m} - 12\varepsilon^3.$$

Lemma 5.20. *In the framework of Design (1D) and under (5.22)–(5.24), the remainder $R^{v'}$ is uniformly bounded in $v'\in V$ by*

$$|R^{v'}| \leq 4(2\varepsilon)^2\max(1,2\varepsilon)(t_\mathrm{m}+\varepsilon)\tfrac{C_K(t_\mathrm{m}+\varepsilon)}{1-C_K(t_\mathrm{m}+\varepsilon)}\left(\tfrac{1}{1-C_K(t_\mathrm{m}+\varepsilon)}+1\right).$$

Proposition 5.14 is a direct consequence of these bounds and inequality (5.25).

Proof of Proposition 5.14. By symmetry, it is sufficient to verify the first inequality of (5.25). Using the bounds from Lemma 5.19 and 5.20, one obtains

$$\begin{aligned}&\left|\frac{\partial M_{\mu_1^+}(f_{K_\star})}{\partial K_1^{+1}}\right| - \left|\frac{\partial M_{\mu_1^+}(f_{K_\star})}{\partial K_1^{-1}}\right| \geq B - 2\max_{v'}|R^{v'}|\\ &\geq \varepsilon^2\left(3e^{-(t_\mathrm{m}+\varepsilon)C_K}t_\mathrm{m} - 12\varepsilon - 32\max(1,2\varepsilon)(t_\mathrm{m}+\varepsilon)\tfrac{C_K(t_\mathrm{m}+\varepsilon)}{1-C_K(t_\mathrm{m}+\varepsilon)}\left(\tfrac{1}{1-C_K(t_\mathrm{m}+\varepsilon)}+1\right)\right) > 0,\end{aligned}$$

by choice of t_m and ε that satisfy (5.24). □

It remains to prove Lemmas 5.19 and 5.20. The proof of the former is based on a consideration of the overlap in the supports of the ballistic forward and adjoint solution, as described after Proposition 5.14.

Proof of Lemma 5.19. Explicit representations of f_0 and g_0 are available in (4.12) and (5.32). Setting $(v, v') = (-1, +1)$, the first summand in B can be estimated as

$$\begin{aligned}
i_1 &:= \int_0^T \int_{I_1} f_0(v') g_0(v') \,\mathrm{d}x \,\mathrm{d}t \\
&= \int_0^T \int_{I_1} \int_t^T e^{-\int_0^t \sigma(x-\tau,+1)\,\mathrm{d}\tau} \mathbb{1}_{a_{3/2}+[-\varepsilon,\varepsilon]}(x-t) \mathbb{1}_{t_\mathrm{m}+[-\varepsilon,\varepsilon]}(s) e^{-\int_0^{s-t} \sigma(x+\tau,+1)\,\mathrm{d}\tau} \cdot \\
&\qquad \mathbb{1}_{a_{3/2}+t_\mathrm{m}+[-\varepsilon,\varepsilon]}(x+(s-t)) \,\mathrm{d}s \,\mathrm{d}x \,\mathrm{d}t \\
&\geq e^{-(t_\mathrm{m}+\varepsilon)C_K} \int_0^{t_\mathrm{m}+\varepsilon} \int_{I_1} \int_{\max(t,t_\mathrm{m}-\varepsilon)}^{t_\mathrm{m}+\varepsilon} \mathbb{1}_{a_{3/2}+t+([-\varepsilon,\varepsilon]\cap[t_\mathrm{m}-s-\varepsilon,t_\mathrm{m}-s+\varepsilon])}(x) \,\mathrm{d}s \,\mathrm{d}x \,\mathrm{d}t \\
&= e^{-(t_\mathrm{m}+\varepsilon)C_K} 3\varepsilon^2 t_\mathrm{m}.
\end{aligned}$$

In the second to last row, the fact that $g_0(t) = 0$ for $t > t_\mathrm{m} + \varepsilon$ was used, together with the form (4.26) of $\sigma = K(x, v', v)$ in $1D$, to bound the exponential functions. Furthermore, the indicator function in s was evaluated. In the last row the integral was evaluated by a case-by-case consideration, keeping in mind that $a_{3/2}+[-\varepsilon,\varepsilon]+t \subset I_1$ for all $t \in [0, t_\mathrm{m}+d_\mu]$ by choice of t_m and ε in (5.17). Similarly, one obtains

$$\begin{aligned}
i_2 &:= \int_0^T \int_{I_1} f_0(v') g_0(v) \,\mathrm{d}x \,\mathrm{d}t \\
&= \int_0^T \int_{I_1} \int_t^T e^{-\int_0^t \sigma(x-\tau,+1)\,\mathrm{d}\tau} \mathbb{1}_{a_{3/2}+[-\varepsilon,\varepsilon]}(x-t) \mathbb{1}_{t_\mathrm{m}+[-\varepsilon,\varepsilon]}(s) e^{-\int_0^{s-t} \sigma(x-\tau,-1)\,\mathrm{d}\tau} \cdot \\
&\qquad \mathbb{1}_{a_{3/2}+t_\mathrm{m}+[-\varepsilon,\varepsilon]}(x-(s-t)) \,\mathrm{d}s \,\mathrm{d}x \,\mathrm{d}t \\
&\leq \int_0^{t_\mathrm{m}+\varepsilon} \int_{I_1} \int_{\max(t,t_\mathrm{m}-\varepsilon)}^{t_\mathrm{m}+\varepsilon} \mathbb{1}_{a_{3/2}+([t-\varepsilon,t+\varepsilon]\cap[t_\mathrm{m}+s-t-\varepsilon,t_\mathrm{m}+s-t+\varepsilon])}(x) \,\mathrm{d}s \,\mathrm{d}x \,\mathrm{d}t \\
&\leq \int_{t_\mathrm{m}-1.5\varepsilon}^{t_\mathrm{m}+\varepsilon} \int_{I_1} \int_{\max(t,t_\mathrm{m}-\varepsilon)}^{t_\mathrm{m}+\varepsilon} \mathbb{1}_{a_{3/2}+t+[-\varepsilon,\varepsilon]}(x) \,\mathrm{d}s \,\mathrm{d}x \,\mathrm{d}t \qquad = 6\varepsilon^3
\end{aligned}$$

In the last row, $t \geq t_\mathrm{m} - 1.5\varepsilon$ was set since the integrand vanishes for all $t < t_\mathrm{m} - 1.5\varepsilon$ and $s \geq t_\mathrm{m} - \varepsilon$ and x, by an empty support of the characteristic function, as

$$t + \varepsilon < t_\mathrm{m} - 0.5\varepsilon \leq t_\mathrm{m} + s - t - \varepsilon.$$

Furthermore, the third summand vanishes

$$\begin{aligned}
0 \leq i_3 &:= \int_0^T \int_{I_1} f_0(v) g_0(v) \,\mathrm{d}x \,\mathrm{d}t \\
&= \int_0^T \int_{I_1} \int_t^T e^{-\int_0^t \sigma(x+\tau,-1)\,\mathrm{d}\tau} \mathbb{1}_{a_{3/2}+[-\varepsilon,\varepsilon]}(x+t) \mathbb{1}_{t_\mathrm{m}+[-\varepsilon,\varepsilon]}(s) e^{-\int_0^{s-t} \sigma(x-\tau,-1)\,\mathrm{d}\tau} \cdot \\
&\qquad \mathbb{1}_{a_{3/2}+t_\mathrm{m}+[-\varepsilon,\varepsilon]}(x-(s-t)) \,\mathrm{d}s \,\mathrm{d}x \,\mathrm{d}t \\
&\leq \int_0^{t_\mathrm{m}+\varepsilon} \int_{I_1} \int_{\max(t,t_\mathrm{m}-\varepsilon)}^{t_\mathrm{m}+\varepsilon} \mathbb{1}_{a_{3/2}-t+([-\varepsilon,\varepsilon]\cap[s+t_\mathrm{m}-\varepsilon,s+t_\mathrm{m}+\varepsilon])}(x) \,\mathrm{d}s \,\mathrm{d}x \,\mathrm{d}t \qquad = 0,
\end{aligned}$$

given that the integrand vanishes for all $s \geq t_\mathrm{m} - \varepsilon$, because the domain of the characteristic function is empty, since

$$s + t_\mathrm{m} - \varepsilon \geq 2(t_\mathrm{m} - \varepsilon) \geq 6\varepsilon$$

by (5.17). Finally,

$$\begin{aligned} i_4 &:= \int_0^T \int_{I_1} f_0(v) g_0(v') \,\mathrm{d}x \,\mathrm{d}t \\ &= \int_0^T \int_{I_1} \int_t^T e^{-\int_0^t \sigma(x+\tau,-1)\,\mathrm{d}\tau} \mathbb{1}_{a_{3/2}+[-\varepsilon,\varepsilon]}(x+t) \mathbb{1}_{t_\mathrm{m}+[-\varepsilon,\varepsilon]}(s) e^{-\int_0^{s-t} \sigma(x+\tau,+1)\,\mathrm{d}\tau} \cdot \\ &\qquad \mathbb{1}_{a_{3/2}+t_\mathrm{m}+[-\varepsilon,\varepsilon]}(x+(s-t)) \,\mathrm{d}s \,\mathrm{d}x \,\mathrm{d}t \\ &\leq \int_0^{t_\mathrm{m}+\varepsilon} \int_{I_1} \int_{\max(t,t_\mathrm{m}-\varepsilon)}^{t_\mathrm{m}+\varepsilon} \mathbb{1}_{a_{3/2}+([-t-\varepsilon,-t+\varepsilon]\cap[t-s+t_\mathrm{m}-\varepsilon,t-s+t_\mathrm{m}+\varepsilon])}(x) \,\mathrm{d}s \,\mathrm{d}x \,\mathrm{d}t \\ &\leq 6\varepsilon^3 \end{aligned}$$

because the integrand vanishes for $t > 1.5\varepsilon$ for all $s \leq t_\mathrm{m} + \varepsilon$ and x, since

$$t - s + t_\mathrm{m} - \varepsilon \geq 1.5\varepsilon - \varepsilon - \varepsilon = -0.5\varepsilon > -t + \varepsilon.$$

In summary, this shows

$$B \geq i_1 - i_2 - i_3 - i_4 \geq e^{-(t_\mathrm{m}+\varepsilon)C_K} 3\varepsilon^2 t_\mathrm{m} - 12\varepsilon^3.$$

□

To prove Lemma 5.20, $R^{v'}$ is further decomposed by regularity as

$$\begin{aligned} R^{v'} = &\sum_{\substack{n,k=0 \\ n+k\geq 1}}^{N} \int_0^T \int_{I_1} f_k(v')(g_n(v') - g_n(v)) \,\mathrm{d}x \,\mathrm{d}t + \int_0^T \int_{I_1} f(v')(g_{\geq N+1}(v') - g_{\geq N+1}(v)) \,\mathrm{d}x \,\mathrm{d}t \\ &+ \sum_{n=0}^{N} \int_0^T \int_{I_1} f_{\geq N+1}(v')(g_n(v') - g_n(v)) \,\mathrm{d}x \,\mathrm{d}t\,, \end{aligned} \tag{5.33}$$

for some $N \in \mathbb{N}$ where $f_k, f_{\geq N+1}, g_n$ and $g_{\geq N+1}$ are defined as follows:

- In extension to (4.11)–($\mathrm{T}_{\geq 2}$), the f_k collect all particles that have tumbles exactly $k \in \mathbb{N}$ times until time t. They satisfy

  $$\partial_t f_k + v \cdot \nabla_x f_k = \mathcal{L}(f_{k-1}) - \sigma f_k, \qquad f_k(t=0) = 0. \tag{T_k}$$

  The remainder $f_{\geq N+1}$ collects all particles that have tumbled more than N times and evolves according to

  $$\partial_t f_{\geq N+1} + v \cdot \nabla_x f_{\geq N+1} = \mathcal{L}(f_N + f_{\geq N+1}) - \sigma f_{\geq N+1}, \qquad f_{\geq N+1}(t=0) = 0. \tag{$\mathrm{T}_{\geq N+1}$}$$

  Again, the diffusive effect of tumbling leads to an increasing regularity of f_k in k.

- Similarly, the g_n solve

$$-\partial_t g_n - v\partial_x g_n = -\sigma g_n + \mathcal{L}^\star(g_{n-1}), \qquad g_n(t=T) = 0, \tag{5.34}$$

and represent that part of the adjoint solution that has been affected by exactly $n \in \mathbb{N}$ tumblings after time t, on their way of being transported backwards in time from the final condition. The remaining part is collected in $g_{\geq N+1}$ that satisfies

$$-\partial_t g_{\geq N+1} - v\partial_x g_{\geq N+1} = -\sigma g_{\geq N+1} + \mathcal{L}^\star(g_N) + \mathcal{L}^\star(g_{\geq N+1}), \qquad g_{\geq N+1}(t=T) = 0, \tag{5.35}$$

where one has $\sigma = K(x,-v,v)$ by (4.26) and $\mathcal{L}^\star(g)(t,x,v) := K(x,-v,v)g(t,x,-v)$ in 1D for the operator $\mathcal{L}^\star = \mathcal{K}^\star + \sigma$. The regularity of the g_n also increases in n.

Existence of f_k, $f_{\geq N+1}$ and their smallness follows in analogy to Lemma 4.5 and Lemma A.1.9.

Lemma 5.21. *Let $K \in \mathcal{A}_K^{\mathrm{pwc}}$ and consider Design (1D) with $\psi = \mathbb{1}_{[-\varepsilon,\varepsilon]}$ as in (5.22). Then (Ch)–(iCh), (T_k) and ($\mathrm{T}_{\geq N+1}$) attain unique mild solutions $f, f_k, f_{\geq N+1} \in C^0([0,T]; L^1_+ \cap L^\infty(\mathbb{R}\times V))$ for all $N \in \mathbb{N}$ and $k \in \{0,...,N\}$, that are bounded by*

$$\begin{aligned} &\|f(t)\|_{L^\infty(\mathbb{R}\times V)} \leq e^{C_K t}\max(1,2\varepsilon), \\ &\|f_k\|_{L^\infty([0,t]\times\mathbb{R}\times V)} \leq (C_K t)^k \max(1,2\varepsilon), \qquad \textit{and} \\ &\|f_{\geq N+1}\|_{L^\infty([0,t]\times\mathbb{R}\times V)} \leq e^{C_K t}(C_K t)^{N+1}\max(1,2\varepsilon). \end{aligned}$$

Proof. An iterative application of Corollary A.1.3 provides existence of $f_k, f_{\geq N+1}$ and the recursive bounds

$$\begin{aligned} &\|f_k\|_{L^\infty([0,t]\times\mathbb{R}\times V)} \leq \|f_k\|_{L^\infty([0,t];L^1\cap L^\infty(\mathbb{R}\times V))} \leq C_K t\|f_{k-1}\|_{L^\infty([0,t];L^1\cap L^\infty(\mathbb{R}\times V))}, \qquad \text{and} \\ &\|f_{\geq N+1}\|_{L^\infty([0,t]\times\mathbb{R}\times V)} \leq e^{C_K t}C_K t\|f_N\|_{L^\infty([0,t];L^1\cap L^\infty(\mathbb{R}\times V))}, \end{aligned}$$

when using the tighter bound $\|\mathcal{L}\| \leq C_K$ in 1D. Similarly, the bound (3.4) for f can be improved to $\|f(t)\|_\infty \leq e^{C_K t}\|\phi\|_{L^1\cap L^\infty(\mathbb{R}\times V)}$. The bounds in the lemma then follow by iteration and the fact that $\|f_0\|_{L^\infty([0,t];L^1\cap L^\infty(\mathbb{R}\times V))} \leq \|\phi\|_{L^1\cap L^\infty(\mathbb{R}\times V)} = \max(1,2\varepsilon)$ by (4.12) for Design (1D). □

Existence and smallness of solutions $g_n, g_{\geq N+1}$ to (5.34) and (5.35) is established in Lemma A.1.9.

Lemma 5.22. *Let $K \in \mathcal{A}_K^{\mathrm{pwc}}$ and consider Design (1D) with $\psi = \mathbb{1}_{[-\varepsilon,\varepsilon]}$ as in (5.22). Then for all $N \in \mathbb{N}$, unique mild solutions g_n and g_{N+1} to (5.34) and (5.35), respectively, exist in $C^0([0,T]; L^1(\mathbb{R}^d; L^\infty(V)))$ for all $n \in \{1,...,N\}$. They vanish for $t > t_{\mathrm{m}} + \varepsilon$ and satisfy*

$$\begin{aligned} &\|g_n\|_{L^1([0,t_{\mathrm{m}}+\varepsilon]\times\mathbb{R};L^\infty(V))} \leq 2(C_K(t_{\mathrm{m}}+\varepsilon))^n(t_{\mathrm{m}}+\varepsilon)(2\varepsilon)^2, \qquad \textit{and} \\ &\|g_{\geq N+1}(t)\|_{L^1(\mathbb{R};L^\infty(V))} \leq 2e^{C_K(t_{\mathrm{m}}+\varepsilon-t)}(C_K(t_{\mathrm{m}}+\varepsilon))^{N+1}(2\varepsilon)^2. \end{aligned}$$

Proof. Lemma A.1.9 and iteratively obtained regularity of $g_{n-1} \in C([0,T]; L^1(\mathbb{R}^d; L^1 \cap L^\infty(V)))$ yield existence, uniqueness and the recursive bounds

$$\begin{aligned}\|g_n\|_{L^1([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R};L^\infty(V))} &\le \|g_n\|_{L^1([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R};L^1\cap L^\infty(V))} \\ &\le C_K(t_\mathrm{m}+\varepsilon)\|g_{n-1}\|_{L^1([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R};L^1\cap L^\infty(V))}, \qquad \text{and} \\ \|g_{\ge N+1}(t)\|_{L^1(\mathbb{R};L^\infty(V))} &\le e^{C_K(t_\mathrm{m}+\varepsilon-t)} C_K \|g_N\|_{L^1([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R};L^1\cap L^\infty(V))},\end{aligned}$$

by the improved bound $\|\mathcal{L}^\star\| \le C_K$, because g_n and $g_{\ge N+1}$ vanish for $t \ge t_\mathrm{m}+\varepsilon$ by Corollary A.1.10. The absolute bounds follow from (A.1.14) by

$$\|g_0\|_{L^1([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R};L^1\cap L^\infty(V))} \le (t_\mathrm{m}+\varepsilon)\|\mu_1^+\|_{L^1([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R};L^1\cap L^\infty(V))} \le 2(t_\mathrm{m}+\varepsilon)(2\varepsilon)^2.$$

□

Note that $g_{\ge N+1}$ vanishes strongly in L^1 as $N \to \infty$, due to $C_K(t_\mathrm{m}+\varepsilon) < 1$ by assumption (5.22). Lemma 5.20 then follows from the previously derived bounds on $f_k, f_{\ge N+1}, g_n$ and $g_{\ge N+1}$.

Proof of Lemma 5.20. The parts in $R^{v'}$ are estimated separately by Hölder's inequality. Noting that $g_n(t) = 0 = g_{\ge N+1}(t)$ for $t \ge t_\mathrm{m}+\varepsilon$ by Corollary A.1.10, then representation (5.33) in connection with the previous two Lemmas yields

$$\begin{aligned}|R^{v'}| &\le 2\Bigg(\sum_{\substack{n,k=0\\ n+k\ge 1}}^{N} \|f_k\|_{L^\infty([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R}\times V)} \|g_n\|_{L^1([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R};L^\infty(V))} \\ &\qquad + \int_0^{t_\mathrm{m}+\varepsilon} \|f(t)\|_{L^\infty(\mathbb{R}\times V)} \|g_{\ge N+1}(t)\|_{L^1(\mathbb{R};L^\infty(V))}\,\mathrm{d}t \\ &\qquad + \sum_{n=0}^{N} \|f_{\ge N+1}\|_{L^\infty([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R}\times V)} \|g_n\|_{L^1([0,t_\mathrm{m}+\varepsilon]\times\mathbb{R};L^\infty(V))} \Bigg) \\ &\le 4(2\varepsilon)^2 \max(1,2\varepsilon)(t_\mathrm{m}+\varepsilon)\Bigg(\sum_{\substack{n,k=0\\ n+k\ge 1}}^{N} (C_K(t_\mathrm{m}+\varepsilon))^{n+k} + e^{C_K(t_\mathrm{m}+\varepsilon)}(C_K(t_\mathrm{m}+\varepsilon))^{N+1} \\ &\qquad + \sum_{n=0}^{N} e^{C_K(t_\mathrm{m}+\varepsilon)}(C_K(t_\mathrm{m}+\varepsilon))^{N+1+n} \Bigg) \\ &\le 4(2\varepsilon)^2 \max(1,2\varepsilon)(t_\mathrm{m}+\varepsilon)\Bigg(\tfrac{C_K(t_\mathrm{m}+\varepsilon)}{1-C_K(t_\mathrm{m}+\varepsilon)} \left(\tfrac{1}{1-C_K(t_\mathrm{m}+\varepsilon)} + 1 \right) \\ &\qquad + e^{C_K(t_\mathrm{m}+\varepsilon)}(C_K(t_\mathrm{m}+\varepsilon))^{N+1} \left(1 + \tfrac{1}{1-C_K(t_\mathrm{m}+\varepsilon)} \right) \Bigg)\end{aligned}$$

where we estimated $\sum_{n=0}^{N}(C_K(t_\mathrm{m}+\varepsilon))^n \le \sum_{n=0}^{\infty}(C_K(t_\mathrm{m}+\varepsilon))^n = \frac{1}{1-(C_K(t_\mathrm{m}+\varepsilon))}$ by the geometric series formula, justified by $C_K(t_\mathrm{m}+\varepsilon) < 1$ by (5.23). Because this estimate holds independently of $N \in \mathbb{N}$, letting $N \to \infty$ shows the desired bound. □

6

Numerical experiments

In this chapter, we present numerical evidence to support the theoretical results from the previous chapter on identifiability of the kinetic tumbling kernel $K \in \mathcal{A}_K^{\mathrm{pwc}}$ from macroscopic data. To study and compare the identifiability behaviour of inverse problems corresponding to different experimental designs, the cost minimization inversion framework is implemented numerically, and the convexity of the cost function (5.6) is studied.

Inverse Problem. The inverse problem under investigation in this chapter coincides with that of the previous chapter, with admissible set chosen as $\mathcal{A}_K^{\mathrm{pwc}}$ from (5.15) or (5.28), according to the considered spatial dimension d. For a prescribed experimental design $(\phi_l, \mu_l)_{l=1,\dots,L}$ the noise free synthetic data is generated by forward computations as in (5.2), and the inverse problem is posed as the cost minimization (5.6).

While several literature exists on computing the forward kinetic autochemotactic chemotaxis model [FY13, CKLMNO19, Yas17, Vau10], this work thus represents the first numerical treatment of an inverse problem related to the kinetic chemotaxis model.

The following numerical results represent adaptations of the numerical experiments in 1D in [HKLT25] to the setting where measurement is distributed in space and time through a space-time test function. The framework is extended to 2D, where identifiability of the inverse problem related to Design (1D) is shown.

Outline. After a short introduction of the numerical framework in Section 6.1, then identifiability of Design (1D) and the decay of identifiability as two experimental setups become close are discussed in spatial dimension $d = 1$ in Section 6.2, and a design is investigated that provides local but not global convexity of the cost function to demonstrate the local nature of Proposition 5.4. In Section 6.3, identifiability of Design (2D) is numerical demonstrated. The chapter is concluded in **??**.

6.1. Numerical Setting

In order to keep the focus of this chapter on demonstrating identifiability, questions of computational efficiency and high accuracy will not be addressed. Instead, results are based on a simple fixed step size gradient descend method for the inversion through cost minimization (5.6) and an easy first order explicit finite difference solver for the forward and adjoint equations in the gradient computation. Details of the numerical techniques are placed in Appendix C.

It is worth mentioning that more sophisticated methods can readily be integrated in the inversion framework. Possible adaptations are discussed in Section 6.1.3.

6.1.1. Optimization by Gradient Descent

If local strong convexity of the cost function $\mathcal{C}$ around $K_\star$ is given, then a simple gradient descend method with a suitable step size $\eta_n \in \mathbb{R}^+$ is guaranteed to converge to $K_\star$, if the initial guess $K^{(0)}$ is sufficiently close to $K_\star$. A brief introduction into the gradient descend method can be found in Section 2.2.1. The iterative reconstruction step reads

$$K^{(n+1)} = K^{(n)} - \eta_n \nabla_K \mathcal{C}(K^{(n)})\,. \tag{6.1}$$

Choice of Step Size. A constant step size $\eta_n \equiv \eta = \frac{1}{\lambda_{\max}}$ is used in the numerical examples, where $\lambda_{\max}$ is the maximal eigenvalue of $H_K\mathcal{C}(K_\star)$. This step size guarantees convergence of the gradient descend method locally under strong convexity by Proposition 2.3 and [Pol63], because Lipschitz continuity of of $H_K\mathcal{C}(K)$ in K according to Lemma 5.3 ensures that this step size falls below the upper bound $\eta \leq \frac{2}{M}$, where M denotes an upper bound on the eigenvalues of $H_K\mathcal{C}(K)$ in a sufficiently small neighbourhood of $K_\star$.

6.1.2. Calculation of the Gradient

To solve (5.6), a first optimize, then discretize approach is adopted. In the thus considered framework, this means that the formula for the gradient is derived for the continuous variable model. The numerical gradient update is then computed as the numerical analogon of this formula, using the numerical solutions to the forward and adjoint equations.

Gradient Formula. In analogy to Lemma 5.5, the gradient $\nabla_K\mathcal{C}(K_\star)$ can be computed by a calculus of variation method via the augmented Lagrangian. This yields the formulas

$$\frac{\partial \mathcal{C}}{\partial K_r^{\pm}} = \iint_{[0,T]\times I_r} f(t,x,\pm 1)(g(t,x,\pm 1) - g(t,x,\mp 1))\,\mathrm{d}(t,x), \qquad \text{for } d=1, \text{ and}$$

$$\frac{\partial \mathcal{C}}{\partial K_r^{j,i}} = \iiiint_{[0,T]\times A_r\times S_j\times S_i} f(t,x,v')(g(t,x,v') - g(t,x,v))\,\mathrm{d}(t,x,v,v'), \qquad \text{for } d=2,$$

where g denotes the solution to the adjoint cost equation

$$\begin{aligned} -\partial_t g - v \cdot \nabla_x g &= \mathcal{K}^\star(g) - \tfrac{1}{L}\textstyle\sum_{l=0}^{L}(M_l(f_K) - y_l)\mu_l, \\ g(t=T,x,v) &= 0. \end{aligned} \tag{6.2}$$

Note that g is a linear combination of the adjoint measurement functions g_{μ_l} solving (5.8)–(5.9) by linearity of the PDE, and existence of a mild solution $g \in C^0([0,T]; L^1(\mathbb{R}^d; L^1 \cap L^\infty(V)))$ is established in Lemma A.1.9.

PDE Discretization. In dimension $d = 1$, the velocity space is already discrete $V = \{\pm 1\}$. For dimension $d = 2$, a discrete ordinate (S_N) method with equidistant quadrature points is employed to discretize the velocity domain $V = \mathbb{S}^1$, which aligns well with the piecewise constant in velocity tumbling kernels in $\mathcal{A}_K^{\text{pwc}}$ in the 2D setting in (5.28), as further explained in Appendix C.1.3. The resulting system of coupled equations for the forward and adjoint model are then solved by the first order explicit finite difference schemes (C.1.7) and (C.1.8) that are based on a Lax Wendroff discretization for the transport part and treat the tumbling operators and source terms explicitly. A detailed derivation can be found in Appendix C.1.3.

6.1.3. Integration of Advanced Methods.

More sophisticated methods can readily be integrated in the presented inversion framework. Some possible adaptations include:

Velocity quadrature. The choice of more advanced quadrature rules, for instance Gauss quadrature, for the discrete ordinates method could improve accuracy of the velocity integral approximation in the numerical measurement procedure as well as the numerical evaluation of the tumbling operator. Thus would entail adaptations in the structure of the admissible set, by adjusting the domains of constant parameter value to the new velocity quadrature points.

Improved Inversion Mechanisms. Convergence of the parameter reconstruction can be improved by the choice of more sophisticated optimization techniques. Starting with integrating a line search for the gradient descend step size [WR22], there exists a vast range of possibilities [Ren10, AS09]. In the realm of gradient based inversion mechanisms, for instance, the Landweber Kaczmarz method [PD16] or stochastic gradient descend methods [CLL18], as well as the Gauss-Newton method [ES13], the method of augmented Lagrangian and (quasi-)Newton methods [Ren10] were successfully implemented for kinetic inverse problems. Sequential quadratic programming [STLH21, HAO00, BM02] or the conjugate gradient method [AS09] present further possible choices.

Numerical forward and adjoint solvers. The development of numerical methods to solve kinetic PDEs constitutes an own area of research, where the main goals are improving

the approximation accuracy and computational efficiency. In the realm of chemotaxis, two second order finite difference schemes were developed for the autochemotaxis system [CKLMNO19, FY13], and a semi-Lagrangian method [Vau10] as well as a Monte Carlo individual path based simulation [Yas17] were proposed.

Such approaches can readily be incorporated in the inversion framework, as long as forward and adjoint schemes are chosen in a compatible fashion w.r.t. the "first optimize, then discretize" gradient computation [HPUU08, AF12].

First discretize, then optimize. On the other hand, the order can be changed and a "first discretize, then optimize" approach brings advantages such as automatic compatibility of forward and adjoint solvers [AF12, Gun02], at the cost of a reduced flexibility in the choice of discretization of the adjoint [HPUU08] and the potential introduction of additional difficulties, such as spurious oscillations in the optimized parameter [LW19].

6.1.4. Computational Setting.

The following settings are adopted for all subsequent computations, in both dimension $d = 1$ and $d = 2$.

Admissible set. The piecewise constant form of the parameter $K \in \mathcal{A}_K^{\mathrm{pwc}}$ is described by the choice of $C_K = 1$ and a partition of the spatial domain $\mathbb{R}$, which is chosen as $a_q = \frac{q}{Q}$ and yields the intervals $I_0 = (-\infty, \frac{1}{Q})$, $I_q = [\frac{q}{Q}, \frac{q+1}{Q})$ and $I_{Q-1} = [\frac{Q-1}{Q}, \infty)$. In case $d = 2$, the two dimensional intervals A_r are the Cartesian products of these intervals.

Experimental Designs. The spatial test functions ψ_x and Ψ_x are chosen as cut off 1- and 2-dimensional Gaussians to minimize spurious oscillations of the numerical scheme, whereas the temporal part ψ_t is chosen as a characteristic function for simplicity. They are the building blocks of all considered experimental designs. Furthermore, all measurements share the same temporal component $\psi_t(t - t_{\mathrm{m}})$, with t_{m} and ε that satisfy (5.17). Assumptions (5.23)–(5.24) will not be enforced for Design (1D).

Experimental Data. The entries of the ground truth parameter $K_\star$ to generate synthetic data $y_l = M_l(f_{K_\star})$ are chosen randomly, following independent uniform distributions in $[0, C_K]$, unless stated otherwise.

Temporal computational domain. Designs will be constructed such that the temporal support of all considered test functions is bounded by $T = t_{\mathrm{m}} + \varepsilon$, and the temporal domain is set to $[0, T]$.

Spatial computational domain. The discretization of the spatial domain is chosen a lot finer than the spatial discretization of the parameter K, to resolve fine structures as those generated by Design (1D).

The spatial computational domains for f and g are set to the d-dimensional intervals $[0,1]^d$ and $[-(t_\mathrm{m}+\varepsilon), 1+(t_\mathrm{m}+\varepsilon)]^d$, respectively: For any initial datum ϕ prescribed by Design (1D) or Design (2D), (5.18) or (5.31) and the choice of $t_\mathrm{m}+\varepsilon$ show that $[0,1]^d$ contains the support of all considered forward solutions $f^\phi(t)$ for $t \leq t_\mathrm{m}+\varepsilon$ of (Ch)–(iCh). Similarly, by Corollary A.1.10 the adjoint solution g is supported in the extended domain $\operatorname{supp} g(t,\cdot,v) \subset [-(t_\mathrm{m}+\varepsilon), 1+(t_\mathrm{m}+\varepsilon)]^d$ for all $t \in [0,T], v \in V$. Other considered designs will share these properties and computational domains.

Zero boundary conditions are imposed, as the boundary will not be reached by the finite speed of propagation.

Space-time discretization. The spatial computational domains are discretized by equidistant points of grid size $\frac{1}{N_x} = \frac{1}{300R}$ in the 1D experiments and $\frac{1}{N_x} = \frac{1}{200R}$ for the 2D experiments, to reduce the computational cost. To obey the stability constraint of the numerical schemes (C.1.7) and (C.1.4), a time step $\tau = T/\lceil T d^{3/2} N_x \rceil$ is chosen.

6.2. 1D Numerical examples.

Numerical experiments provide insight into the identifiability of the inverse problems corresponding to different experimental designs.

6.2.1. Well-Posedness of Design (1D).

Theorem 5.12 analytically proves local identifiability of the inverse problem related to Design (1D) and local strong convexity of the cost function. In this section, numerical evidence is presented that even suggests global strong convexity and identifiability. This in particular shows that the reconstruction of $K \in \mathcal{A}_K^{\mathrm{pwc}}$ from velocity averaged data is possible, if a suitable experimental design was chosen, i.e. structural identifiability holds in the parametric setting. Uniform strong convexity even indicates possible stability of the parameter reconstruction: in the sense of practical identifiability with noise, the confidence intervals related to the parameter reconstruction are finite. A more detailed investigation, including noise in the data, is left for future investigation.

The geometry of Design (1D) in the first interval I_0 is summarized in the two panels of Figure 6.1. The experiments in the remaining intervals are constructed symmetrically. The left panel shows the propagation of the bacteria density $\int_V f \,\mathrm{d}v$ in time from $t_\mathrm{m}-\varepsilon$ to $t_\mathrm{m}+\varepsilon$ in comparison to the stationary value of $\mu_1^\pm$ in this time interval. The ballistic part is the most dominant, and it decays over time due to tumbling. In order to observe this effect, $R=2$ was chosen in this computation to allow for larger measurement times t_m by (5.17). As the right panel suggests, the ballistic trajectory of the data passes through the support of the measurement test function and they perfectly overlap at $t=t_\mathrm{m}$, indicating that the ballistic part is measured well in this design.

In what follows, $R=20$ is chosen to demonstrate the flexibility of the reconstruction

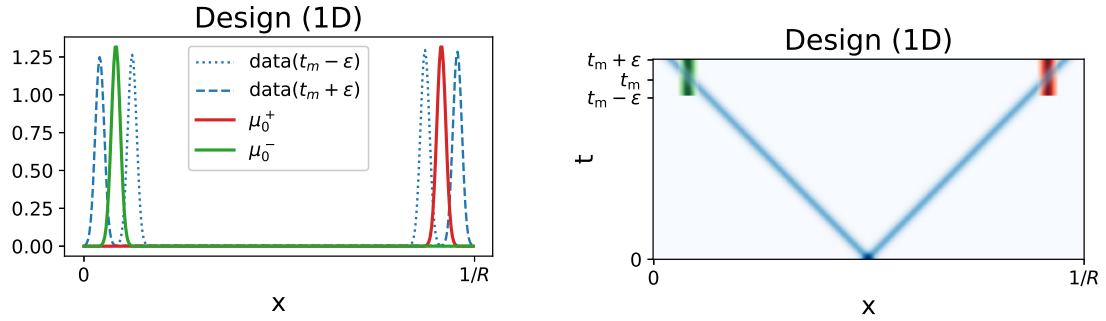


Figure 6.1.: Propagation of the bacteria density data $\int_V f \, dv$ (blue) and the test functions μ_0^+ (red) and μ_0^- (green) in the first interval I_0 under Design (1D), for $R = 2$. Left panel: Function values at times $t_m - \varepsilon$ and $t_m + \varepsilon$. Right panel: space time propagation.

setting to cope with high dimensional ($Q = 40$) parameters. By the modular construction of Design (1D), the considered properties are supposed to hold independently of the precise choice of R.

Marginal Strong Convexity. Local strong convexity of the quadratic cost $\mathcal{C}(K)$ with noise free data around a ground truth parameter $K_\star$, as suggested by Theorem 5.12, can be observed in numerical experiments. These experiments not only show positivity of the minimum eigenvalue, but even exhibit a good conditioning of the Gauss-Newton Hessian, with an inverse condition number $c^{-1} = \frac{\lambda_{\min}(H_K\mathcal{C}(K_\star))}{\lambda_{\max}(H_K\mathcal{C}(K_\star))} \approx 0.791$ close to 1, which guarantees a weak eigenvalue decay of the sensitivity matrix J and thus well posedness of the linearized inverse problem.

Moreover, marginal (global) strong convexity of the cost function is observed: Figure 6.2 displays the marginal cost functions that fix all but two parameters of K to the ground truth values of $K_\star$ and vary only the two remaining parameters, i.e. for prescribed $r, s \in \{0, ..., R-1\}$ and $i, j \in \{\pm 1\}$ one considers the cost function

$$\mathcal{C}(\hat{K}_{k_1,k_2}) \quad \text{with} \quad \hat{K}_r^i = k_1, \hat{K}_s^j = k_2 \text{ and } \hat{K}_q^{\pm 1} = K_{\star,q}^{\pm 1} \text{ for the remaining values.} \tag{6.3}$$

Global strong convex throughout the full parameter domain $k_1, k_2 \in [0, 1]$, suggests marginal convexity.

To obtain further insight into the shape of the full $2R = 40$-dimensional cost landscape, its values $\mathcal{C}(K_\lambda)$ on lines

$$\{K_\lambda = K_\star + \lambda\omega, \lambda \in \mathbb{R}\} \tag{6.4}$$

through $K_\star$ in a randomly chosen direction $\omega \in \mathbb{S}^{39}$ are plotted in the subsequent Figure 6.3 against the distance λ of K_λ to $K_\star$. Because only those line segments inside the parameter domain $[0, 1]^{40}$ are considered, the lines start and end at different λ values by randomness of the direction ω. It can be observed that the cost function is almost uniformly convex in each direction, starting at $K_\star$. This is also indicated (locally) by the very weak eigenvalue decay of the Hessian $H_K\mathcal{C}(K_\star)$, expressed in the high inverse condition number of 0.791.

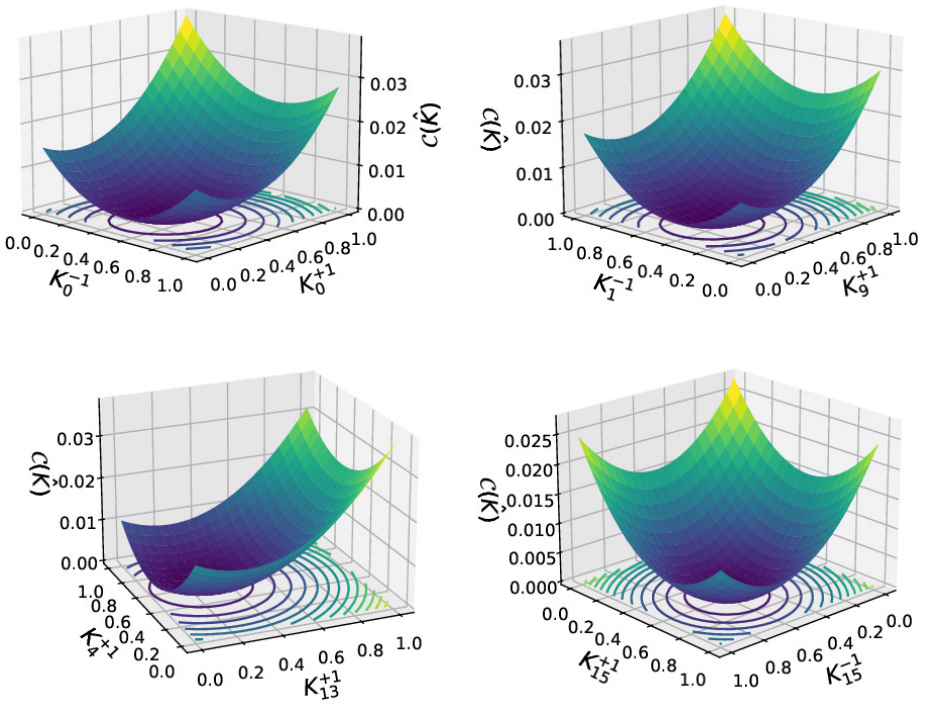


Figure 6.2.: Strong convexity of the marginal cost function landscapes under Design (1D).

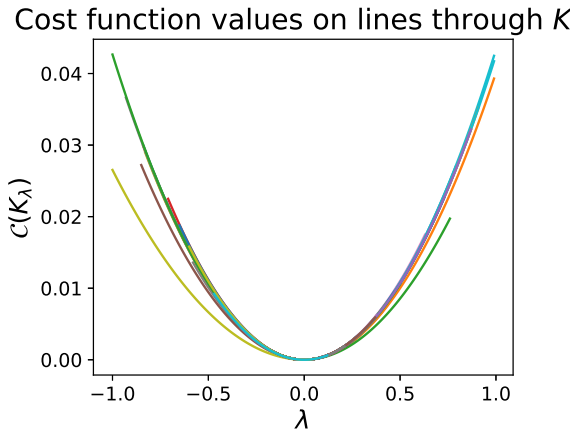


Figure 6.3.: Values of $\mathcal{C}$ on lines through $K_\star$ in 100 randomly chosen directions under Design (1D).

Moreover, marginal strong convexity can be observed as strict positivity in the minimum eigenvalue landscape of the Hessian. To speed up the computation, a smaller setting with $R = 2$ is considered and the ground truth parameter is set to $K_{\star,r}^{\pm 1} = 0.5$ for all r for an simpler visualization. In this setting, the minimal eigenvalue $\lambda_{\min}(H_K \mathcal{C}(\hat{K}))$ is plotted against two varying parameters, which are denoted on the axes in Figure 6.4. If the remaining parameters of K are set to the values of the ground truth (left panel), then

this describes the minimal eigenvalue landscape of the marginal cost as in (6.3). In the right panel, these remaining parameters are chosen at random. Positivity even in this case strongly suggests global strong convexity, as to be further investigated in Section 6.2.4.

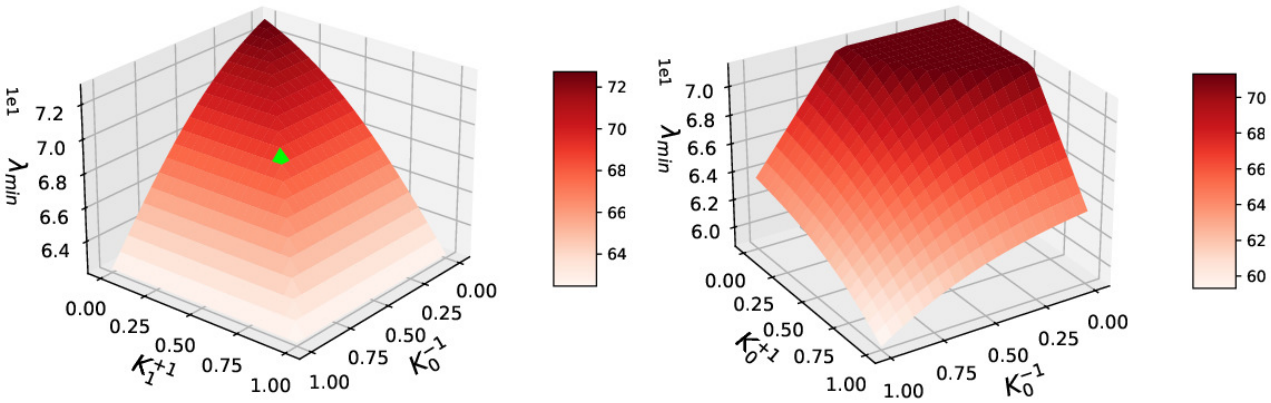


Figure 6.4.: Minimum eigenvalues of the Hessian $H_K\mathcal{C}(\hat{K})$ for Design (1D) at varying values of $\hat{K}$, where all entries apart from those on the x and y axis are fixed to either the ground truth, whose value is marked by the green dot (left panel), or randomly selected values (right panel).

Parameter Reconstruction. Iterative reconstructions $K^{(n)}$ by the simple gradient descent scheme (6.1) attain exponential convergence towards the ground truth parameter $K_\star$, independent of the randomly chosen the initial value $K^{(0)}$, as predicted by Proposition 2.3. Figure 6.5 shows this convergence for all 40 entries $K_r^{\pm 1}$ of K when $R = 20$ in linear and logarithmic scale, as well as the cost function.

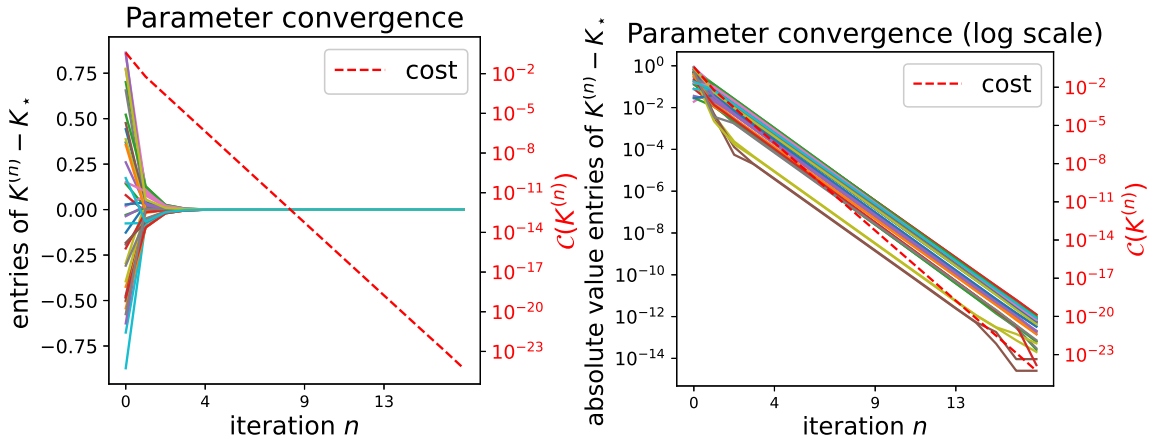


Figure 6.5.: Convergence of the values $K_r^{(n),\pm 1}$ of $K^{(n)}$ (solid lines) to those $K_{\star,r}^{\pm 1}$ of $K_\star$ in linear (left panel) and logarithmic scale (right panel) under Design (1D). The red dashed line depicts the cost of the chosen parameter $K^{(n)}$ in the n-th iteration.

6.2.2. Ill-conditioning for Close Experimental Setups

In this section we study the behaviour of the reconstruction as two measurement test functions that share the same initial data become asymptotically close. The following family of designs is constructed, in dependence on a closeness parameter $s \in [-1, 1]$:

Design (Close^s). *The designs coincide with Design (1D), except for the first measurement function: μ_0^- is exchanged for $\mu_0^s(t,x) = \psi(t - t_\mathrm{m})\psi(x + t_\mathrm{m} \cdot s - a_{1/2})$ for a fixed value $s \in [-1, 1]$.*

For $s = -1$, this yields the original Design (1D), whereas $\mu_0^s \to \mu_0^+$ in $L^1([0,T] \times \mathbb{R})$ as $s \to 1$. Figure 6.6 shows Design (Close^s) for two values of $s \in \{0.9, 0.97\}$. The measurement test function $\mu_0^{0.9}$ is close to μ_0^+ but still distinguishable, whereas $\mu_0^{0.97}$ almost overlaps μ_0^+. The experiments in the remaining intervals are constructed in analogy to the first interval of Design (1D), as displayed in Figure 6.1.

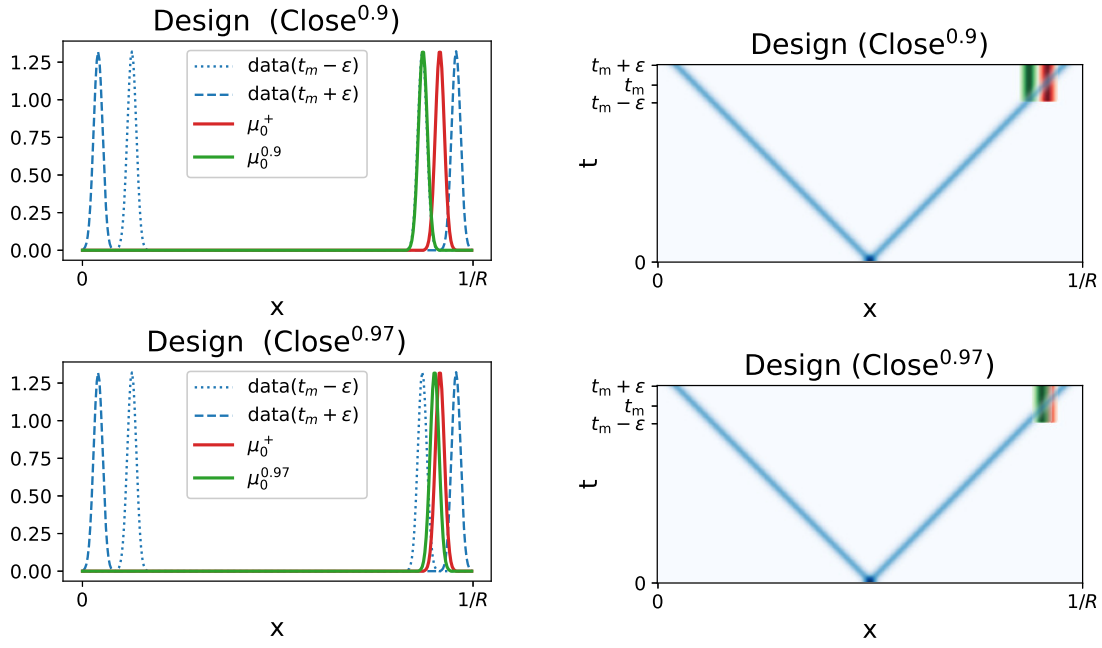


Figure 6.6.: Propagation of the ballistic parts of the bacteria density $\int_V f \, \mathrm{d}v$ (blue) and the test functions μ_0^+ (red) and μ_0^s (green) in the first interval I_0 from $t_\mathrm{m} - \varepsilon$ to $t_\mathrm{m} + \varepsilon$ (left panel) and in the full space-time domain (right panel) for Design (Close^s) with $s = 0.9$ (upper row) and $s = 0.97$ (lower row).

Loss of Convexity. Theorem 5.6 suggests that the rank structure of the Hessian deteriorates as $\mu_0^s \to \mu_0^+$. This is observed numerically in Figure 6.7 by plotting the minimum eigenvalue $\lambda_\mathrm{min}(H_K\mathcal{C}(K_\star))$ as well as the inverse condition number $c^{-1}(H_K\mathcal{C}(K_\star)) = \frac{\lambda\mathrm{min}(H_K\mathcal{C}(K_\star))}{\lambda\mathrm{max}(H_K\mathcal{C}(K_\star))}$ of the Hessian of the cost function at the ground truth parameter $K_\star$ against s.

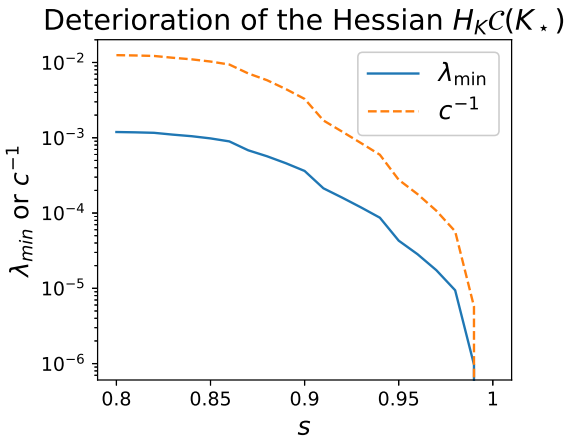


Figure 6.7.: Deterioration of the minimum eigenvalue $\lambda_{\min}(H_K\mathcal{C}(K_\star))$ of the Hessian of the cost function at ground truth (blue), and of its inverse condition number $c^{-1}(H_K\mathcal{C}(K_\star)) = \frac{\lambda\min(H_K\mathcal{C}(K_\star))}{\lambda\max(H_K\mathcal{C}(K_\star))}$ (orange, dotted), as $s \to 0$ in Design (Close^s).

This is a consequence of a loss in sensitivity of the cost function w.r.t. parameter K_0^-. Figure 6.6 shows that the negative velocity part of f^{ϕ_0} in Design (1D), previously measured by μ_0^-, is not captured well by any sensor. This part was mostly affected by the decay due to tumbling with K_0^-. The lack of sensitivity expresses through flatness of the marginal cost functions in Figure 6.8, for fixed $s = 0.9$, as well as the small minimal eigenvalue of $\lambda_{\min}(H_K\mathcal{C}(K_\star)) \approx 0.00036$ and inverse conditioning $c^{-1} \approx 0.0033$.

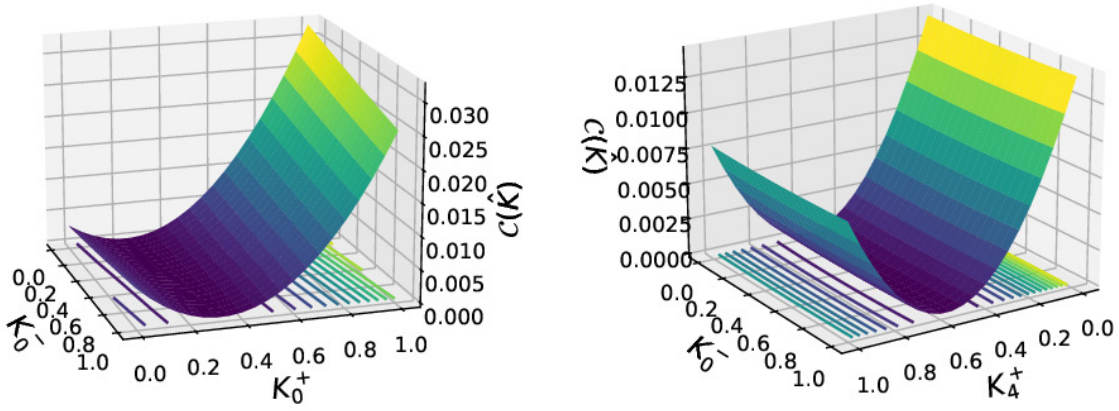


Figure 6.8.: Flat marginal cost functions w.r.t. change in K_0^- and one further parameter under Design (1D) with $s = 0.9$.

The sensitivity w.r.t. other parameters of K is not affected, as Figure 6.9 shows. This makes sense, given that the relevant sensors were not moved by Design (Close^s) in comparison to Design (1D).

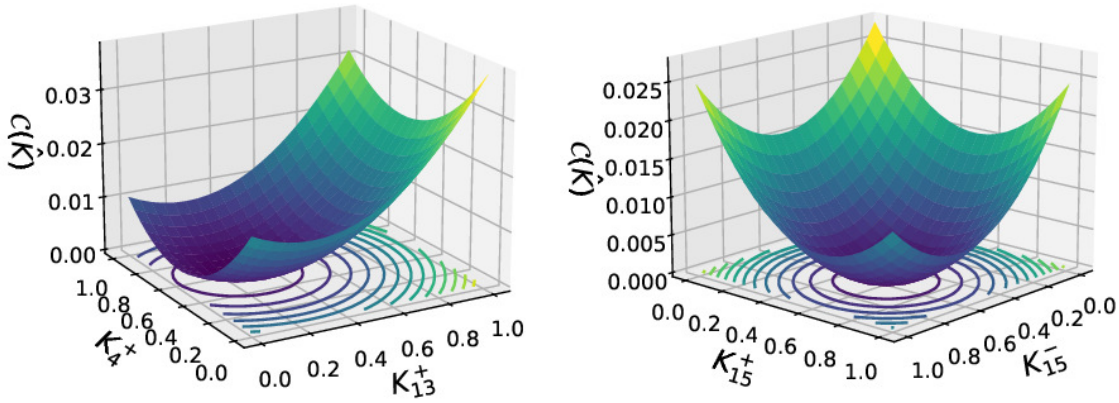


Figure 6.9.: Examples of marginal cost functions that do not vary the parameter K_0^- under Design (Close^s) with $s = 0.9$.

The existence of very flat directions in the cost landscape can also be observed in a plot of $\mathcal{C}(K_\lambda)$ along random lines through $K_\star$, as defined in (6.4), in Figure 6.10.

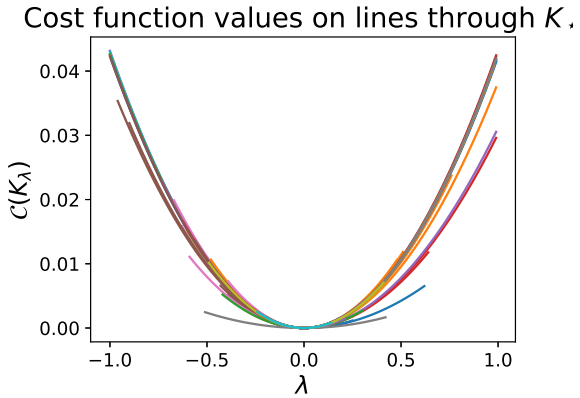


Figure 6.10.: Values of $\mathcal{C}$ on lines through $K_\star$ in 100 randomly chosen directions, under Design (Close^s) with $s = 0.9$.

Failure in Parameter Reconstruction. The decay of convexity and flatness of the cost in certain directions provokes a significant slow down in the parameter convergence of gradient based minimization methods, such as the thus applied gradient descend scheme (6.1), as can be observed in Figure 6.11. The slow convergence expresses most prominently in the parameter convergence of K_0^{-1}, the direction in which the cost function is flat. For $s = -1$, the design equals Design (1D) and reconstructions converge very fast, and after only 5 iteration steps a parameter accuracy of 10^{-3} is reached, compare Figure 6.5. For $s = 0.9$, instead, Figure 6.12 shows that convergence is so slow that after 1000 iteration steps of the gradient descend scheme, though reaching accuracy 10^{-3}, the parameter error is still larger compared to the 5-step accuracy with Design (1D). This behaviour is even worse for $s = 0.97$, where the value of $K_0^{(n),-1}$ is improved only very mildly towards $K_{\star,0}^{-1}$

within the first 1000 iterations, and convergence is not clear. In the underdetermined case $s = 1$ where $\mu_0^s = \mu_0^+$ and one measurement is redundant, the parameter K_0^{-1} is not updated by the gradient scheme after step 4 and remains far away from the true value.

The reconstruction of other parameters $K_r^{\pm 1}$ in intervals $I_r \neq I_0$ is not affected by the deterioration, as they do not rely on the measurements inside I_0 by Proposition 5.10. Neither is the minimization of the cost, which explicitly expresses that the measurement is not sensitive w.r.t. K_0^{-1}.

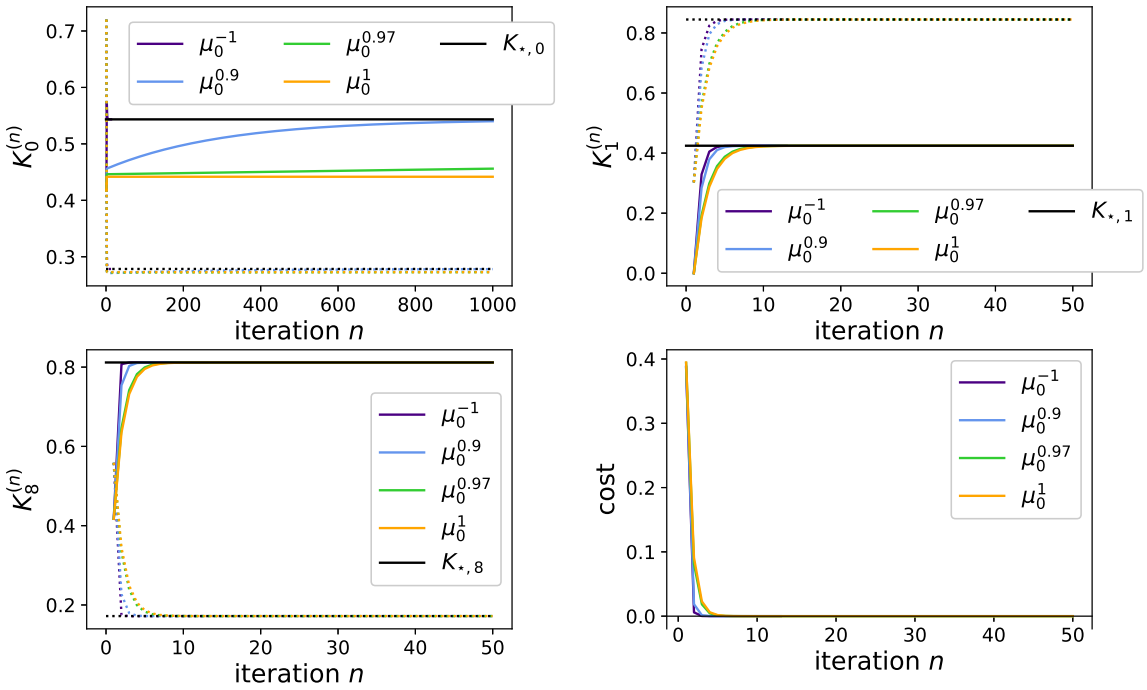


Figure 6.11.: Comparison of the convergence behaviour of reconstruction of the parameters K_r^{-1} (solid lines) and K_r^{+1} (dashed lines) with $r \in \{0, 1, 8\}$ and the cost under Design (Close^s) with different values $s \in \{-1, 0.9, 0.97, 1\}$, in the first 50 gradient descend steps.

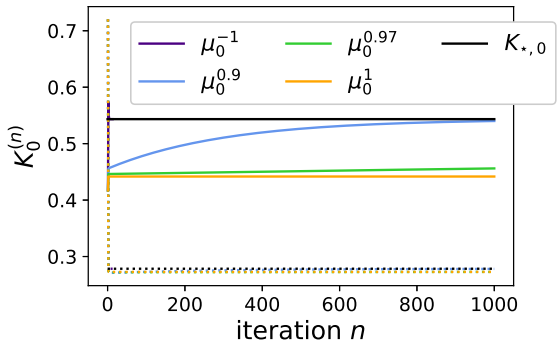


Figure 6.12.: Convergence of reconstruction of K_0 under Design (Close^s) with different values $s \in \{-1, 0.9, 0.97, 1\}$, for 1000 gradient descend steps.

6.2.3. A Locally but Non-Globally Convex Design

Proposition 5.4 ensures local convexity, if sensitivity based identifiability holds. The result cannot be improved to imply global convexity, as the following design demonstrates.

To study global convexity of the cost function, a simplified setting with $R = 2$ different values of K in space is adopted, which reduces the dimension of K to 4. This facilitates the Hessian computation and allows plotting the minimum Hessian eigenvalue $\lambda_{\min}(H_K\mathcal{C}(K))$ at a large variety of K values. Furthermore, the ground truth is placed in the middle of the domain by setting $K^{\pm 1}_{\star,r} = 0.5$ for better observability in the plots.

Even in this reduced setting, a design can be constructed that leads to local convexity of the cost function around the ground truth, but that is not globally convex. Its geometry is depicted in Figure 6.13

Design (Conv_{loc}). *With the same choice of $t_{\text{m}}, \varepsilon, T$ and ψ_x, ψ_t as in the previous designs, choose*

- *one velocity independent initial datum $\phi(x, v) = \psi_x(x - x_0) + \psi_x(x - x_1)$ centered around the positions $x_0 = 0.5 - \frac{T}{3}, x_1 = 0.5 + \frac{T}{3}$, and*
- *measurement test functions $\mu_r^{\pm}(t, x) = \mathbb{1}_{t_{\text{m}}+[-\varepsilon,\varepsilon]}(t)\psi_x(x - (x_r \pm (t_{\text{m}} - \frac{T}{6})))$, for $r = 0, 1$.*

Note that the initial data is not located in the middle of intervals I_0, I_1, but shifted towards their joint boundary such that the data starting in one interval will cross the border at time $\frac{T}{3}$. The 'inner' measurement test functions μ_0^+ and μ_1^- are located in the respective other interval and do not capture the ballistic parts of the data well.

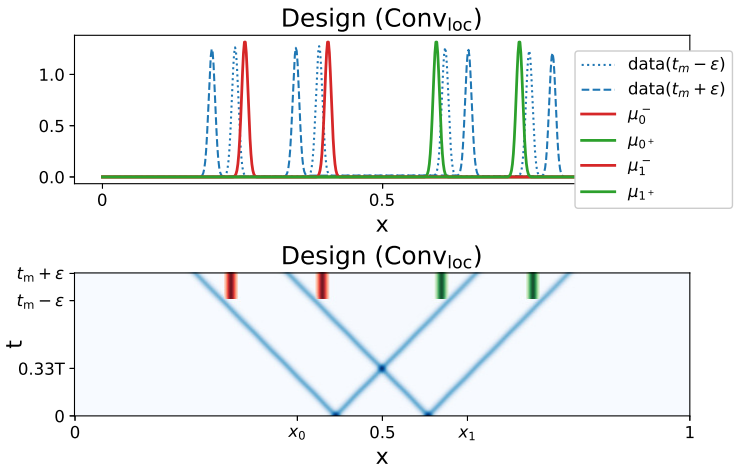


Figure 6.13.: Geometry of Design (Conv_{loc}) and propagation of the ballistic part of $\int_V f \, \mathrm{d}v$ from $t_{\text{m}} - \varepsilon$ to $t_{\text{m}} + \varepsilon$ (upper panel), and in full space time (lower panel).

The cost function related to Design (Conv_{loc}) is not globally convex, as can be observed in Figure 6.14, where the eigenvalues of the cost function Hessian become negative at certain marginal parameter values $\hat{K}$, as defined in (6.3). At the ground truth parameter (green dot), positivity of the minimum eigenvalue suggests local identifiability and

convexity of the cost, and as predicted by Proposition 5.4, one can find a neighbourhood where this remains true.

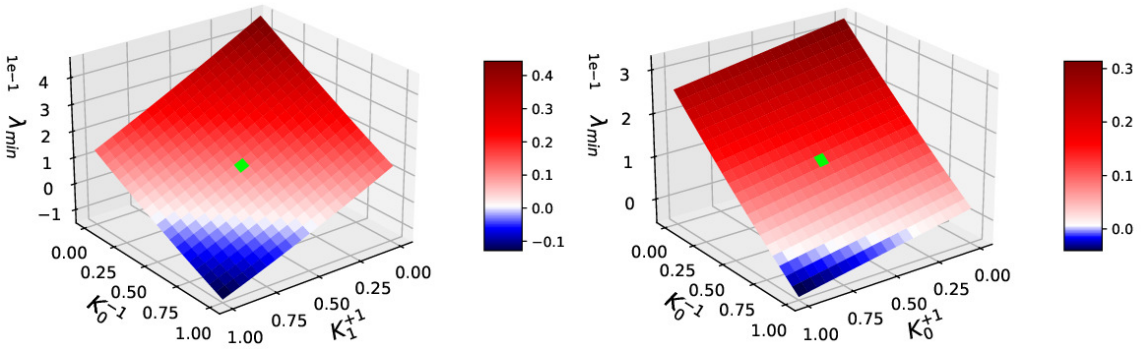


Figure 6.14.: Minimum eigenvalues of the Hessian $H_K\mathcal{C}(\hat{K})$ for Design (Conv$_{\text{loc}}$) at varying values of $\hat{K}$, where all entries apart from those on the x and y axis are fixed to the ground truth. The ground truth parameter is marked by the green dot.

It should be mentioned, however, that the thus observed non convexity was so mild that it could not be observed with bare eye in marginal cost function plots in Figure 6.15. In all numerical examples, it belonged to a saddle point and numerical reconstructions still converged, even if initial values are taken in the corresponding region of non convexity. This is not surprising, given the well-known fact that first order methods, such as gradient descend, tend to avoid saddle points almost surely [LPP$^+$19]. An example is shown in Figure 6.16, where the peaks in the logarithmic parameter convergence plot in the right panel in Figure 6.16 originate from changes in the sign of the parameter difference.

It should be mentioned that although Design (Conv$_{\text{loc}}$) is convex by positivity of its minimum eigenvalue of order 10^{-1}, the quality of convexity is very low, as the minimum eigenvalue, as well as the inverse conditioning of order 10^{-2} are very small. Precise values can be found in Table 6.1. Especially in numerical contexts, where positivity of the Hessian is exchanged for small positive lower bound, this design might not be considered convex anymore.

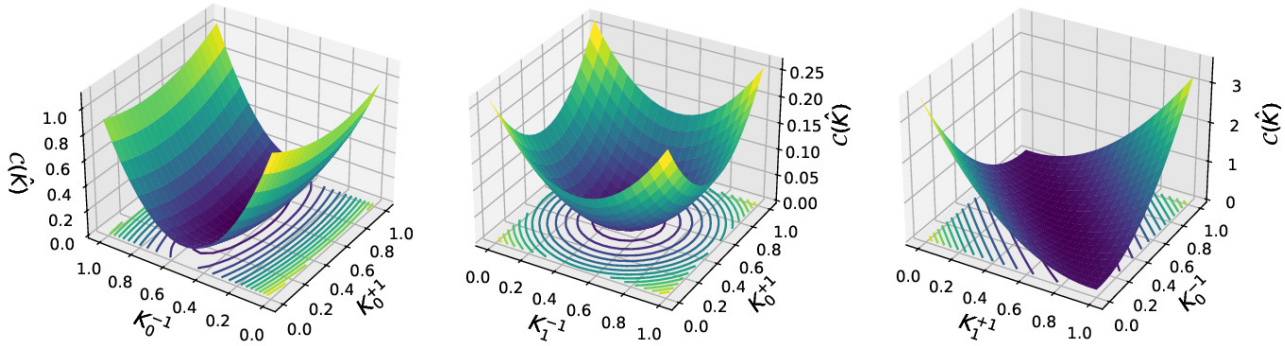


Figure 6.15.: Marginal cost function landscapes under Design (Conv$_{\text{loc}}$).

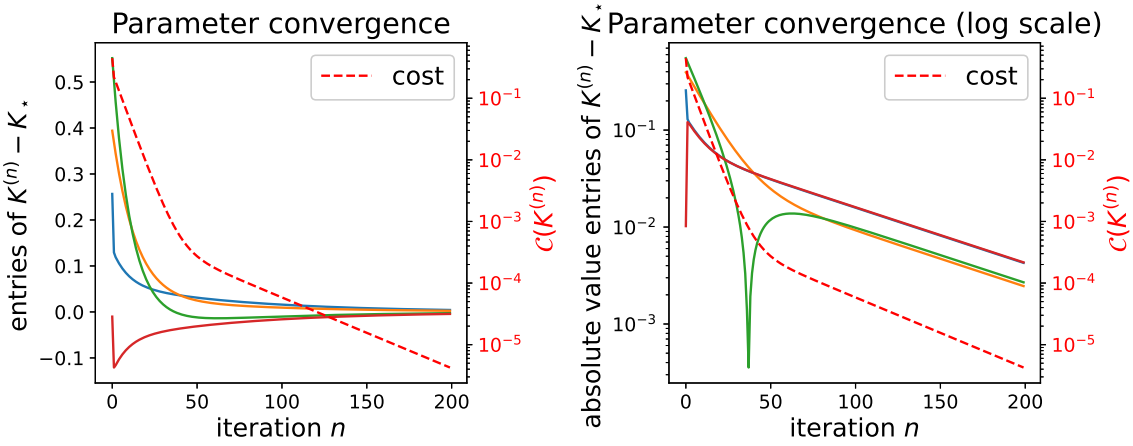


Figure 6.16.: Convergence of the values $K_r^{(n),\pm}$ of $K^{(n)}$ (solid lines) to those $K_{\star,r}^{\pm 1}$ of $K_\star$ in linear (left panel) and logarithmic scale (right panel) under Design (Conv$_{\text{loc}}$). The red dashed line depicts the cost of the chosen parameter $K^{(n)}$ in the n-th iteration. The number of iterations was increased for the logarithmic plot.

6.2.4. Comparison of the Convexity of considered Designs.

To compare the convexity of the cost functions corresponding to the thus considered Designs (1D), (Close^s) and (Conv$_{\text{loc}}$), the setting of the previous subsection is adopted, where $R = 2$ and $K_{\star,r}^{\pm 1} = 0.5$ for all $r = 1, 2$ is fixed.

Local Strong Convexity. Proposition 5.4 suggests to study local strong convexity of the cost function around the ground truth by means of the eigenvalue structure of the Gauss-Newton Hessian $H_K\mathcal{C}(K_\star)$, in particular by consideration of its minimal eigenvalue as a theoretical indicator of convexity. Moreover, the (inverse) conditioning $c^{-1} = \frac{\lambda_{\min} H_K\mathcal{C}(K_\star)}{\lambda_{\max} H_K\mathcal{C}(K_\star)}$ gives insight into the rate at which the experiment compresses information on the parameter and thus well-posedness of the inverse problem, and affects the convergence of numerical minimization schemes, as described in Proposition 2.3.

Comparing the cost function plots Figures 6.2, 6.8 and 6.15, it is obvious that Design (1D) attains the best local strong convexity, underlined by the largest minimum Hessian eigenvalue and a conditioning close to 1. Design (Conv$_{\text{loc}}$) and Design (Close^s) both adhere very flat directions, as indicated by very small inverse condition numbers and minimal eigenvalues in Table 6.1, which anticipates worse identifiability properties, especially with noisy data which might easily perturb the flat cost landscape.

Global Convexity. To study global convexity of the cost functions corresponding to the three considered designs, the minimal eigenvalues and inverse conditionings of their Hessians are plotted for 1000 randomly sampled parameter values in the 4 dimensional parameter space $\mathcal{A}_K^{\text{pwc}}$ in Figure 6.17.

Strong positivity of the minimal eigenvalues and a conditioning close to 1 in all consid-

Design	Design (1D)	Design (Close^s) with $s = 0.9$	Design (Conv$_{\text{loc}}$)
$\lambda_{\min}(H_K\mathcal{C}(K_\star))$	69.093	0.30028	0.13165
c^{-1}	0.79703	0.0029271	0.0098018

Table 6.1.: Local convexity of the cost function related to different designs, expressed through the minimal Gauss-Newton Hessian eigenvalue and its inverse conditioning.

ered points suggests that Design (1D) in fact yields a globally strongly convex quadratic cost function. The conditioning for Design (Close^s) with $s = 0.9$ is much smaller. By positivity of the minimal eigenvalue, minimization algorithms will still converge, but at very much slower rates. This could also be observed in Figure 6.12. Design (Conv$_{\text{loc}}$), instead, also attains negative eigenvalues and thus non-convexity of the cost function in certain regions of the K domain, as observed already in Figure 6.14. It can be noticed that the eigenvalues are nicely aligned on one line, which indicates that the maximum eigenvalue of Design (Conv$_{\text{loc}}$) does not vary (much) throughout the parameter space.

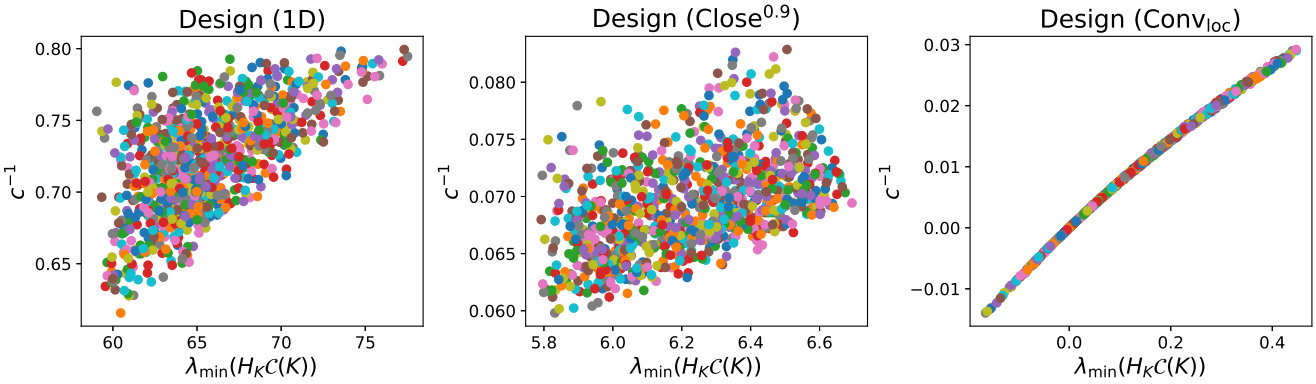


Figure 6.17.: Minimum eigenvalue and inverse conditioning of the cost function Hessian corresponding to Design (1D), Design (Close^s) with $s = 0.9$ and Design (Conv$_{\text{loc}}$) at 1000 randomly sampled values of K in the parameter space.

6.3. Extension to 2D

The effects of close measurements and the local nature of the well-posedness result have been extensively studied in the 1D case, such that this section is restricted to the demonstration of well-posedness of Design (2D).

The 2D setting comes with increased computational complexity in connection with the higher dimensional x and v space, and the following computational setting shall considered: a $N_v = 4$ velocity model is considered under $R = 2$, i.e. the spatial domain is divided into $R^2 = 4$ intervals in which the tumbling kernel is spatially constant. Note that this setting already amounts in a $R^2 N_v(N_v - 1) = 48$-dimensional parameter K. The spatial

discretization is reduced to $\frac{1}{N_x} = \frac{1}{200R}$, to speed up the computations. As before, it is expected that the numerical results of this section generalize to finer discretizations by the modular construction of Design (2D).

The geometry of Design (2D) is summarized in Figure 6.18. In the left panel, the full data $\int_V f^{\phi_r^0} \, \mathrm{d}v$ is shown for all intervals A_r simultaneously. The ballistic part dominates and the first order tumbling part $\int_V f_1 \, \mathrm{d}v$ becomes only visible once the ballistic part is cut out in the right panel. Three rays correspond to the three directions of tumbling in the $N_v = 4$ velocity model. The corresponding densities f_1 are measured by test functions that centre around red dots, marking the scattered locations $a_{r+1/2} + \frac{t_\mathrm{m}}{2} v_0 + \frac{t_\mathrm{m}}{2} v_j$.

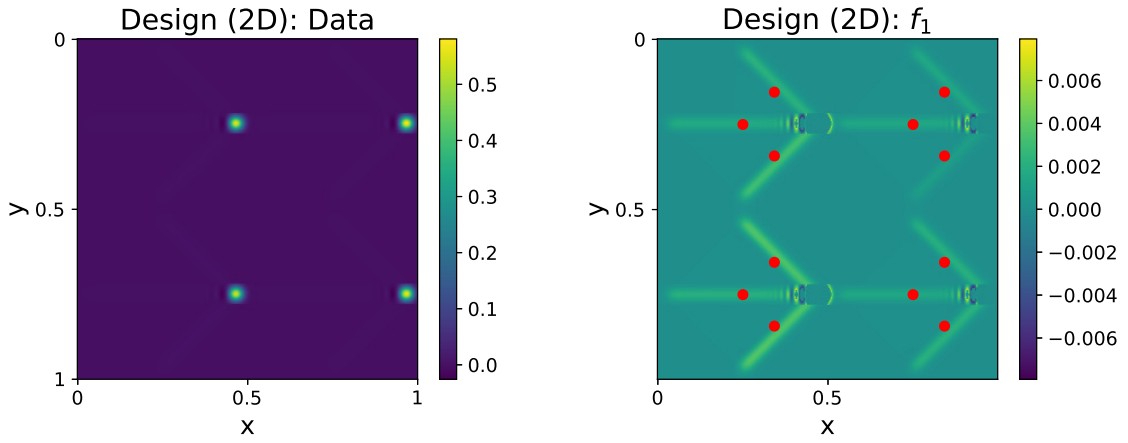


Figure 6.18.: Distribution of the full data (left panel), and the data where the ballistic part is cut out (right panel) for ϕ_r^0 under Design (2D) with $R = 2$ and under simultaneous experimentation in the 4 spatial domains of K. The red dots mark the scattered locations $a_{r+1/2} + \frac{t_\mathrm{m}}{2} v_0 + \frac{t_\mathrm{m}}{2} v_j$ at which the measurements are taken.

Convexity of the Cost Landscape. Marginal cost landscapes can be defined in a similar manner as in (6.3) by fixing all but two parameters of K to the ground truth. Their strong convexity all over the parameter domain in Figure 6.19 again suggests marginal strong convexity the the cost function associated to Design (2D).

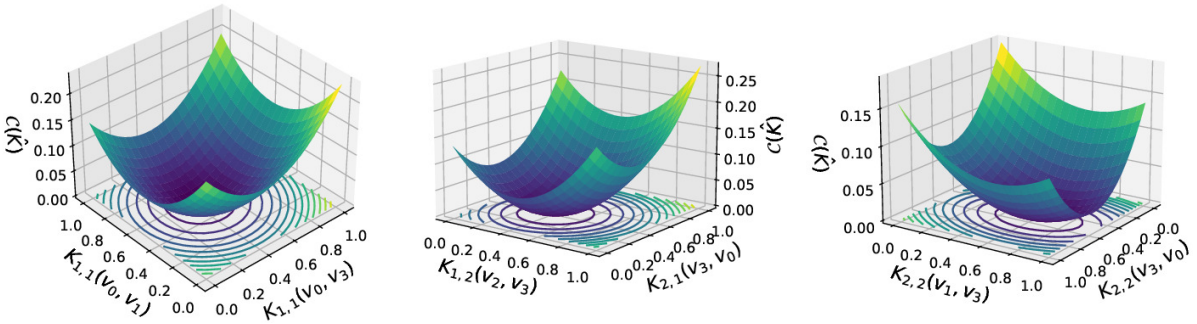


Figure 6.19.: Three marginal cost landscapes under Design (2D).

Similarly, the marginal strong convexity can be observed by plotting the value of $\mathcal{C}(K_\lambda)$

on lines $(K_\lambda)_\lambda$ through the true parameter, as in (6.4), in Figure 6.20.

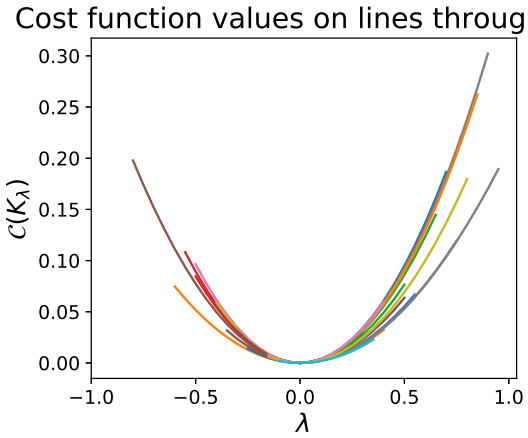


Figure 6.20.: Value of $\mathcal{C}$ on lines through $K_\star$ in 30 randomly chosen directions under Design (2D).

Parameter Reconstruction. As in the 1D case, a cost function minimization succeeds to reconstruct the true parameter in very few steps, with exponential convergence. The convergence of the single parameters is displayed in Figure 6.21 in linear and logarithmic scale. As in Figure 6.16, the peaks in the logarithmic plot originate from changes in the sign of the parameter difference during convergence.

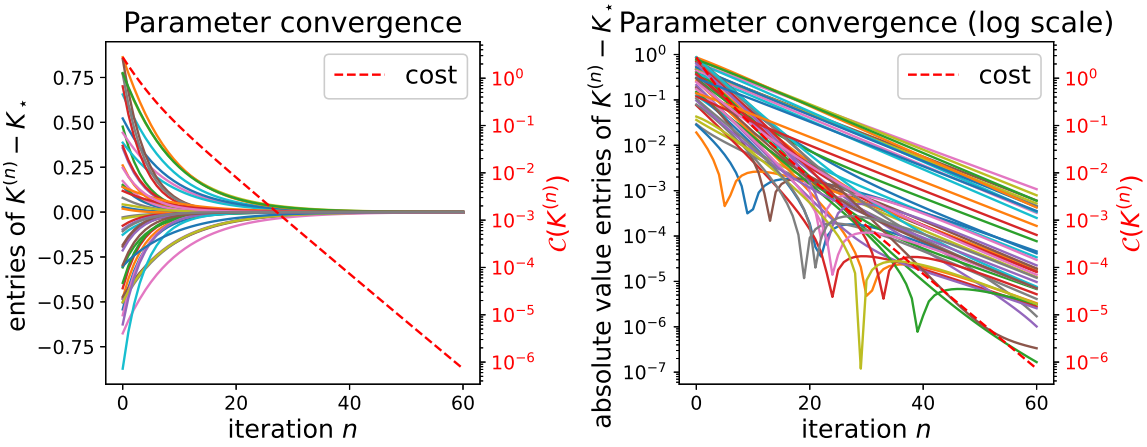


Figure 6.21.: Convergence of the values $K_r^{(n),j,i}$ to those $K_{\star,r}^{(n),j,i}$ of $K_\star$ in linear (left panel) and logarithmic scale (right panel) under Design (2D). The red dashed line depicts the cost of the chosen parameter $K^{(n)}$ in the n-th iteration.

7

Conclusion of Part I

In this part, we laid out the approach to derive an experimental design from a theoretical proof on a concrete example, the inverse problem of recovering the kinetic chemotaxis tumbling kernel from macroscopic data which has not been considered in literature before, underpinned by numerical examples.

Non-Parametric Structural Identifiability. After introducing the kinetic forward model and the inverse problem under investigation, structural identifiability of the non-parametric kernel $K \in \mathcal{A}_K^{\mathrm{cont}}$ from the corresponding input-to-output map is established, where, for the first time, unique identifiability of the kinetic kernel from macroscopic data could be shown without further simplifications on K, owed to the tight control over the experimental setup. Based on the singular decomposition framework, a rigorous, constructive proof could be given which relied on the detailed construction of a sequence of experimental setups with increasing singularity in the initial data and measurement test function and a simultaneously vanishing measurement time, under a specifically designed geometry. These setups triggered microscopic information on K in the measurement, and allowed reading off the value of K at one design-determining evaluation point $(x_{\mathrm{tumb}}, v_{\mathrm{out}}, v_{\mathrm{in}}) \in \mathbb{R}^d \times W$ from the data.

Finite Data Parametric Setting. Because it is not feasible, neither experimentally nor computationally, to run a sequence of vanishing time singular input data experiments for each evaluation point of K, the next project studied the simultaneous reduction of the data dimension and the dimension of the admissible set to a finite data parametric setting. A delicate balancing between the loss of information on the parameter in the data and the additional injection of a-priori knowledge through the admissible set is required to sustain identifiability over this process, and we laid out how the theoretical proof guided our choices: the pointwise reconstruction suggested adopting a locally constant form of K which allowed relaxing the singularity and short time requirements for the designs. This discretization form of K is more general than frequently encountered physical insight based parametric forms and directly connects the infinite data non-parametric inverse problem with its finite dimensional parametric counterpart. Similar techniques as in the

non-parametric identifiability proof allowed us to analytically demonstrate sensitivity of the found design w.r.t. the parametric K in dimension $d = 1$, entailing local strong convexity of the quadratic cost function and thus good numerical reconstruction results. This was contrasted by studying the deterioration of the inversion setting under a decay of data diversity analytically and numerically, demonstrating the need for well chosen experimental designs.

Comments and Further Directions. The presented application of the "relaxation of theory" approach to experimental design benefitted greatly from the constructive nature of the theoretical proof based on the explicit construction of experimental designs. It emphasized the duality between the discretization of the admissible set and the experimental design, which should be constructed such that the information in its data compliments the remaining degrees of freedom in the parameter.

While an analytical proof of this sensitivity of the proposed relaxation based Design (2D) w.r.t. the parameters in $\mathcal{A}_K^{\text{fin}}$ is remains open for higher spatial dimensions such as $d = 2$, its similarities to Design (1D), together with numerical experiments, make us confident that such a proof can be established as an extension of the 1D proof under mild additional efforts.

The shown analysis and numerics were presented in a noise free setting, which allowed us to isolate the influence of the experimental design on the compression of the information on K, and its relation to the admissible set. However, practical inverse problem typically suffer from data noise which can distort the cost landscapes and challenge parameter reconstruction. This entails the investigation of stability of the inverse problem. On a theoretical side, it amounts in a characterization of the admissibly type of noise in the data, and a quantification of its amplification by the inverse problem through suitable norms, which can also be approached by the singular decomposition technique [SU03, BJ08], assuming again access to the infinite dimensional input-to-output map as data. This can be complemented by the numerical application of practical identifiability techniques [SM23, LDM22] to study the influence of noise on the reconstruction with a fixed experimental design, as briefly discussed in Section 2.4.3. Investigation of the relation of infinite dimensional designs and their finite counterparts in terms of stability could lead to further insight into the behaviour of well-posedness under the transition to the finite dimensional problem.

In contrast to the presented approach to experimental design, the Bayesian optimal experimental design framework [APSG16] automatically embeds data noise by constructing designs that minimize the uncertainty in the parameter reconstruction under noisy data.

Moreover, noisy data may require more advanced inversion techniques to mitigate computational artifacts. The flexibility of the inversion framework, as discussed in Section 6.1.3, supports such adaptations.

Another interesting question when dealing with inverse problems related to kinetic forward models refers to their multiscale behaviour. As described in Section 3.1.3, the kinetic forward model (Ch) is linked to other types of descriptions of the chemotaxis phenomenon, for instance the PKS model on the macroscopic scale through a scaling limit

[CMPS04, EHPS24]. It has been shown in [HKLT21] that this limit translates to the Bayesian reconstruction of the kinetic tumbling kernel K, however, as shown for the stationary radiative transport equation in [CLW18b, NLS20, LNW22], the limit might still be ill-posed, and a thorough investigation of the limiting problem remains open. Moreover, the multiscale limit is expected to also affect experimental designs, which were clearly based on the finite speed of propagation and rare tumbling so far, both of which is no longer true in the diffusion limit. Especially if the true scaling is unknown, this raises the question for asymptotically sensitive experimental designs.

Having chosen the experimentally easier accessible macroscopic data for the reconstruction of the chemotaxis tumbling kernel, we hope that eventually our work lays the foundation of reliable estimation of the kinetic dynamics from real data. The insights into experimental design and its interplay with a-priori assumptions, together with the development of a basic reconstruction algorithm to produce the numerical results, provide a first step in this direction. Nonetheless, it is still a long route to apply the inversion strategy to real experimental data: the realizability of the proposed experimental design in the lab setting would have to be checked and potentially adjusted, and the robustness of the inversion algorithm w.r.t. measurement noise should be investigated and possibly improved through regularization techniques. These steps require detailed insight into the practical procedures, and call for a close collaboration between mathematicians and practitioners. Ultimately, reconstructing the parameter under different factors of influence might lead to biological insight into first principles on how bacteria react to these factors on a quantitative level.

Part II.

Sampling Approach

8

Experimental Design through Sampling

A typical approach to solve inverse problems is to minimize the quadratic deviation

$$\mathcal{C}_D(p) = \|F_D(p) - y\|_2^2$$

of the synthetic measurements $F_s(p)$ from the experimentally observed data y_s for all experimental setups s in the design D, in which case local strong convexity of this cost function around the ground truth parameter is a particularly desirable property, as described in Section 2.2.

Structural identifiability, i.e. identifiability under access to the full input-to-output map, often implies this local strong convexity - or equivalently sensitivity based identifiability according to Proposition 2.11 - under access to a large, possibly infinite number of noise-free data $\{y_s\}_{s\in D^{\text{full}}}$, generated by a large experimental design $D^{\text{full}} \in \mathcal{P}(\mathcal{D})$. If the parameter to be reconstruction only attains $Q << |D^{\text{full}}|$ degrees of freedom, it seems reasonable that such a large number of experiments might not be necessary for reconstruction, as suggested for instance by Proposition 2.13, nor may it be feasible, neither experimentally nor computationally. Selecting a small finite number $L \geq Q$ of informative experimental setups $s_1, ..., s_L$ such that the inverse problem corresponding to the finite design $D_L = \{s_1, ..., s_L\}$ is still locally cost function identifiable, is thus a core challenge when designing real world experiments, that is classically addressed by optimal experimental design techniques, as introduced in Section 2.5.

In this chapter, we relax the optimality condition and our goal is to propose a design mechanism in the realm of sensitivity based methods, that yields a sufficient experimental design in the sense that it ensures local strong convexity of the quadratic cost function. Through regarding the design process as a down-sampling task to select $D_L \subset D^{\text{full}}$, the main question is how to choose the sampling distribution in order to preserve cost function convexity under a small number of experimental setups L. By translating local strong convexity into positivity of the Gauss Newton Hessian $H_p\mathcal{C}_{D^{\text{full}}}(p_\star) = J^T J$ of the quadratic cost function by means of Proposition 2.11, we find the answer to this question in the seemingly unrelated area of Randomized Numerical Linear Algebra (RNLA), that was originally developed to analyse large data that exceed RAM capabilities and make big data applications feasible [Mah16, MT20, W$^+$14, HB21]. A matrix sketching algorithm

[Mah16, MT20] leverages importance down-sampling of rows of J, encoding in our case the experimental setups $s \in D^{\text{full}}$, to approximate the matrix-matrix product $J^T J$. Its quantitative guarantees on the approximation quality translate to identifiability guarantees under high probability.

Practical application then requires invoking a sampling algorithm, and we propose to adopt algorithms from Bayesian posterior sampling, given that the same difficulties as described in Section 2.3.1 arise in the design sampling context. In particular, ensemble based methods are deployed for their efficiency in a numerical example for the sensor placement problem related to the Schrödinger potential reconstruction, which illustrates the capability of our method to greatly improve convexity of the quadratic cost landscape.

In contrast to the very model specific approach and techniques that were exploited in the previous Part I to construct an identifiability guaranteeing experimental design for the kinetic chemotaxis kernel reconstruction problem, this approach uses the model as a black box and requires no deep insight, making it applicable to a large class of inverse problems.

This project emerged from a collaboration with Christian Klingenberg and Qin Li, summarized in the preprint [HKL24], which has been submitted to peer review.

Remark 8.1 (Difference to [HKL24]). The method presented in this work extends the results from [HKL24] to settings where the full data experimental design D^{full} describes an infinite set of experimental setups, possibly even a subset of an infinite dimensional function space, as frequently encountered as the domain of the input-to-output map. In this sense, the starting point of this method is put even closer to the structural identifiability setting. On a technical side, we introduce a (probability) measure on D^{full} in order to reasonably define the quadratic cost function. The above mentioned adaptations then require an extension of the randomized matrix multiplication algorithm to quasimatrices. Moreover, the theoretical guarantee on sensitivity under a sufficiently large sample size is slightly extended to additionally guarantee a conditioning close to the full data setting. Numerically, the example for the Schrödinger potential reconstruction is extended to a sensor placement and source term design problem to demonstrate the capabilities of the method to sample also forcing data and work in contexts where the full sampling distribution as well as its normalization constant are not available.

Outline. The subsequent quick review of related literature is followed by the extension of the matrix sketching algorithm from RNLA to quasimatrices in Section 8.1. This serves as a prerequisite for setting up the general design sampling program in Section 8.2, that is initiated by describing the starting point of a sensitivity based identifiable inverse problem corresponding to a large, noise-free data set in Section 8.2.1, followed by the idea to consider experimental design as a sampling task, and a translation of the importance sampling strategy and its theoretical guarantees to the design sampling context in Section 8.2.3. The general program is concluded by practical considerations around the choice of sampling algorithms and the addition of a greedy mechanism in Sections 8.2.4 and 8.2.5. The potential of the developed program to improve cost function convexity is illustrated on the sensor placement problem for the inverse Schrödinger potential reconstruction problem in Section 8.3, where importance sampling distributions are studied in Section 8.3.1,

before the improvements in cost function convexity are demonstrated in Section 8.3.2. An extension of the setting to the sensor placement and forcing data design problem verifies applicability of the method when the full sampling landscape is not available. The chapter is concluded in Section 8.4.

Novelty of this Work. The main contribution of this work lies in the change of perspectives: instead of searching the optimal design, we search for a sufficient design, that describes the inverse problem comparably as well as the full data setting. This allows us to relax the methodology from optimization to sampling. The choice of an efficient sampling strategy is key to render even a low number of experimental setups and, hence, low experimental cost, sufficient and we propose a model-informed importance sampling strategy that optimally preserves the convexity properties of the full data cost in the down-sampled problem. Sampling from these non-standard distributions then requires more sophisticated samplers, and we propose a pipeline to apply our strategy in practice.

Related Literature. After highlighting similarities to standard optimal experimental design techniques, we collect literature connecting experimental design with randomized linear algebra approaches.

Optimal Experimental Design. Our approach is rather 'qualitative' in comparison the 'quantitative' optimal experimental design methods in literature that determine a design as the minimizer of certain optimality criteria that measure uncertainty in the reconstruction by means of the conditioning of the Fisher information matrix or Bayesian posterior variance matrix [Kie59, Mit00, BSE+09, PKG+18, AC22], as described in Chapter 2. For a review, the reader is referred to [HJM24, Ale21].

By (2.19) the Hessian $H_p\mathcal{C}_{D^{\text{full}}}(p_\star) = J^T J$ coincides (up to a constant) with the Fisher information matrix FIM for the linearized inverse problem under an additive i.i.d. Gaussian noise model, and therefore our objective is to preserve - instead of optimize - its conditioning through sampling, as it encodes the convexity of the cost $\mathcal{C}$.

Moreover, the implicit linearization of the inverse problem through sensitivity methods is also in line with the approach to study the linearized problem for design questions, or equivalently a Laplace approximation of the Bayesian posterior around the maximum a posteriori point [BTGMS13, LSTW13, LMT15, APSG16], yielding locally optimal designs, so to circumvent the challenges of optimal experimental design for non-linear problems.

Random Matrix Sketching and Inverse Problem Design. The application of random matrix sketching to design tasks in inverse problems is currently at its fancy. In compressed sensing literature, the usefulness of sparse, model-independent random measurement matrices to recover a low rank signal in a generic basis has been known for a long time, as summarized in [MBKB18].

Model dependent sampling strategies were used in [BTLZNn22] for design selection in electrical impedance tomography based on sketching the input-to-output map. Model based information entered the sparse uniform random sketching through part-wise sketching on the basis of a specific rank structure decomposition, and the strategy was constructed in such a way that a subsequent matrix completion step could successfully re-

construct the input-to-output map, yielding astonishing numerical reconstruction results. In [OHC24], a sparse random mask for MRI data acquisition was constructed, based on a data- and model-adapted design parameter. This design significantly improved posterior images in Bayesian reconstructions, as compared to hand crafted designs.

Moreover, a branch of operator learning research investigates the learning of elliptic solution operators from random input-solution pairs [BT23, BHT23], based on a randomized singular value decomposition, and a hierarchical low rank structure of the solution operator, as exploited already in [BTLZNn22].

More generally, random matrix sketching finds numerous applications in enhancing the efficiency of solving linear inverse problems through preconditioning, for instance for wave form inversion [AAGO21]. In [CLNW20] and [CJ21], the authors highlighted advantages of choosing a structure of the random sketching matrices that is consistent to the forward matrix when solving the linear inverse problem or a constrained least squares optimization, respectively. As the sketching matrices in these examples were non-sparse, no further implications on the experimental design could be drawn.

Another application of sketching to inverse problems was developed in [JLNS24] in a setting very similar to this work, where the authors also utilize positive definiteness (and the conditioning) of the Gauss-Newton Hessian of the quadratic cost in the noise free setting to study well-behavedness of the inverse problem. Assuming a fixed experimental design with fewer data than the dimension of the parameter to reconstruct, their starting point differs significantly from ours, and therefore the goal of their work is to gain insight into the dimension of a random subspace reduction of the admissible set which yields a convex cost minimization problem - instead of the experimental design, they sketch the admissible set. As this reduction can be expressed as a random sketching of the underdetermined full-admissible-set Gauss-Newton cost Hessian, RNLA methods allow them to give high probability bounds on the conditioning of the reduced-admissible-set Gauss-Newton Hessian, and numerical examples validate their approach.

8.1. Prerequisites: Randomized Quasimatrix Multiplication

In this section, an extension of the randomized matrix multiplication algorithm from randomized linear algebra [Mah16, MT20] to quasimatrices, with possibly an infinite row dimension, shall be derived. This is necessary to justify our sampling approach in case of an infinite full data design D^{full}.

Quasimatrices. Quasimatrices are linear operators between a finite dimensional vector space and an infinite dimensional vector space. As such, they generalize tall but skinny matrices. For a given domain D, a probability measure μ on D and a natural number Q, the $(D,\mu)\times Q$ quasimatrices J is of the form $J = (j_1, ..., j_Q) \in (L^2(D;\mu))^Q$ with $L^2(D;\mu)$ functions j_l in its columns and possibly an infinite number of 'rows' $J_s = (j_1(s), ..., j_Q(s)) \in \mathbb{R}^Q$, denoting the evaluation of J at a fixed $s \in D$. Many matrix operations can be generalized to quasimatrices through exchanging the column scalar product by the $L^2(D;\mu)$ inner

product. For instance, the quasimatrix product $J^T J$ is defined as

$$\mathbb{R}^{Q\times Q} \ni J^T J \quad = \quad \left(\langle j_l, j_m \rangle_{L^2(D;\mu)} \right)_{l,m=1}^{Q} \quad = \quad \int_D J_s \otimes J_s \, \mathrm{d}\mu(s), \tag{8.1}$$

and the Frobenius norm can be defined via

$$\|J\|_F^2 = \int_D \|J_s\|_2^2 \, \mathrm{d}\mu(s).$$

A short introduction to quasi matrices may be found in [TT15, BT23, Ste98].

Generalizing Radomized Matrix Multiplication to Quasimatrices. In what follows, we extend the derivations in [Mah16, MT20] to approximate the matrix product $\mathtt{J}^T\mathtt{J}$ for a tall matrix $\mathtt{J}$ through row sampling, to the quasimatrix context. These approaches exploit the structure of $J^T J$ as an integral over the 'tall' direction of the quasimatrix on the RHS of (8.1).

Standard Monte Carlo. A standard approach would be to apply a Monte Carlo method and draw i.i.d. samples r_l of a random variable η according to the distribution μ, and approximate the matrix product by the uniformly weighted sum

$$J^T J = \int_D J_r \otimes J_r \, \mathrm{d}\mu(r) = \mathbb{E}_{\eta\sim\mu}(J_\eta \otimes J_\eta) \quad \approx \quad \frac{1}{L}\sum_{l=1}^{L} J_{r_l} \otimes J_{r_l} = \hat{J}^T \hat{J}, \tag{8.2}$$

for some $L \in \mathbb{N}$, justified by the law of large numbers, where $\hat{J}$ denotes the sketched random matrix whose l-th row is given by $\hat{J}_{l:} = L^{-1/2} J_{r_l}$. This method typically requires a large number of samples for good accuracy.

Importance Sampling. Importance sampling often requires a smaller sample size to achieve the same accuracy as standard Monte Carlo [MT20]. The basic idea behind this approach is to transform the sampling measure through introducing a density $\rho \in L^1(D,\mu)$ with $\int_D \rho_s \, \mathrm{d}\mu(s) = 1$, that vanishes in s only if $J_s = 0$, such that

$$J^T J = \int_D (\rho_s^{-1} J_s \otimes J_s)\rho_s \, \mathrm{d}\mu(s) = \mathbb{E}_{\zeta\sim\rho\mu}(\rho_\zeta^{-1} J_\zeta \otimes J_\zeta) \quad \approx \quad \frac{1}{L}\sum_{l=1}^{L} \rho_{s_l}^{-1} J_{s_l} \otimes J_{s_l} = \tilde{J}^T \tilde{J}$$

by the law of large numbers, where the l-th row of sketched random matrix $\tilde{J}$ reads

$$\tilde{J}_{l:} = (\rho_{s_l} L)^{-1/2} J_{s_l} \qquad \text{for } l = 1, ..., L, \quad \text{and random draws} \qquad s_l \sim \rho\mu. \tag{8.3}$$

This allows more flexibility in sampling, and a clever choice of ρ as the relative size of the s-th 'row'

$$\tilde{\rho}_s \quad := \quad \frac{\|J_s\|_2^2}{\|J\|_F^2} \tag{8.4}$$

minimizes the expected quadratic deviation $\mathbb{E}(\|J^T J - \tilde{J}^T \tilde{J}\|_F^2)$ of $\tilde{J}^T \tilde{J}$ from its mean $\mathbb{E}(\tilde{J}^T \tilde{J}) = J^T J$, in complete analogy to the finite dimensional case [Mah16, MT20]. Similarly, an application of concentration bounds yields closeness in Frobenius norm with high probability:

Theorem 8.2 (Extension of [Mah16, Thm.7]). *Let J be a $(D;\mu) \times Q$ quasimatrix with finite Frobenius norm $\|J\|_F < \infty$, and fix a natural number $L << |D|$, and consider a probability density ρ on D with $\int_D \rho_s \,\mathrm{d}\mu(s) = 1$ such that there exists a positive number $\beta \in (0,1]$ for which*

$$\rho_s \geq \beta \tilde{\rho}_s \qquad \text{for } \mu \text{ almost all } s \in D, \tag{8.5}$$

and let $\tilde{J}$ be constructed as in (8.3). Then $\tilde{J}^T\tilde{J}$ approximates J^TJ with high accuracy, with high probability: For any failure rate $\delta \in (0,1)$,

$$\mathbb{P}\left(\|J^TJ - \tilde{J}^T\tilde{J}\|_F \leq \frac{1+\sqrt{8\beta^{-1}\log(\delta^{-1})}}{\sqrt{\beta L}} \|J\|_F^2 \right) \geq 1-\delta.$$

Note that the optimal sampling density is given by $\rho = \tilde{\rho}$, where $\beta = 1$ minimizes the error sampling error in Frobenius norm. It suggests biasing the sampling towards those elements $s \in D$ for which the evaluation of J in s attains a big norm $\|J_s\|_2^2 \propto \tilde{\rho}$ in relative terms - in a finite dimensional setting this means a big relative row size. Introducing smaller β allows for more flexibility in the choice of the sampling distribution, at the cost of accuracy, sample size or occurrence probability. The order $\sqrt{\log(\delta^{-1})/L}$ of the sampling error in L could be anticipated from the central limit theorem.

8.2. General Program

In this section we present the sampling strategy at the core of this project. Starting with an over-sampled setting for which identifiability holds - for instance established in structural identifiability studies -, experimental design can be regarded a down-sampling task to select the meaningful designs, and the sampling distribution is derived from a sketching algorithm for the Gauss-Newton Hessian encoding the cost function convexity. The program is completed by practical considerations about the choice of sampling algorithms, and the suggestion to combine the method with a greedy selection mechanism to further improve cost convexity.

8.2.1. Starting Point

Consider a finite dimensional admissible set $\mathcal{A} \subset \mathbb{R}^Q$ for the inverse problem

$$\text{find} \quad p \in \mathcal{A} \qquad \text{s.t.} \qquad (y_s)_{s\in D} = (F_s(p))_{s\in D} \tag{8.6}$$

associated to an experimental design D. The sensitivity analysis framework, as introduced in Section 2.4.2, can be extended to potentially infinite experimental designs D, when defining the cost by means of a fixed probability measure μ on D as

$$\mathcal{C}_{D,\mu}(p) = \frac{1}{2}\|F(p) - y\|^2_{L^2(D,\mu)} = \frac{1}{2}\int_D (F_s(p) - y_s)^2 \,\mathrm{d}\mu(s).$$

If $\mathcal{C}_{D,\mu}$ attains sufficient regularity, then, in analogy to the finite design setting, its Hessian

$$\mathbb{R}^{Q\times Q} \ni H_p\mathcal{C}_{D,\mu}(p_\star) = J^T J = \int_D J_s \otimes J_s \,\mathrm{d}\mu(s),$$

where J is a $(D,\mu)\times Q$ quasimatrix and $J_s = \nabla_p F_s \in \mathbb{R}^Q$ denotes its value at experimental setup $s \in D$, is informative about local strong convexity of $\mathcal{C}_{D,\mu}$, as an obvious extension of Proposition 2.11 shows. Again, a global characterization of the landscape is hard for nonlinear inverse problems.

Remark 8.3. The measure $\mu(s)$ might for instance indicate the difficulty of preparing the an experimental setup s, or its likelihood of appearance. For finite designs D, the uniform measure $\mu(s) = \frac{1}{|D|}$ will be deployed without further specification. A probability measure is taken for notational convenience, but the theory below still holds true for non-zero finite measures, when normalization is conducted where necessary.

We pose the following assumptions to further characterize the setting of our inverse problems (8.6).

Assumption 8.4. Assume that there exists a (possibly infinite) design D^{full}, called full data design, and a probability measure μ on D^{full} such that

(A1) the noise free data $y = F(p_\star)$ is generated by an underlying ground truth parameter $p_\star$ under design D^{full},

(A2) the single experiment forward maps F_s are C^2 in p in a small neighbourhood of $p_\star$, with uniformly continuous Hessians $H_p\mathcal{C}_{\{s\}}$ for all $s \in D^{\text{full}}$, and

(A3) the Hessian of the full data cost $\mathcal{C}^{\text{full}}(p) := \mathcal{C}_{D^{\text{full}},\mu}(p)$ at the global optimizer is positive definite, i.e. $H_p\mathcal{C}^{\text{full}}(p_\star) \succ 0$.

Assumption (A1) summarizes the setting that we operate in: to study solely experimental design according to source (II) of non-identifiability, noise is excluded at this point of consideration. Moreover, it ensures that the true parameter $p_\star$ is a global minimum of the cost function.

As discussed below Proposition 2.11, regularity assumptions on F and continuity of the Hessian in (A2) are rather mild and often satisfied, in particular for PDE inverse problems, where forward well-posedness often includes regularity of the solution w.r.t. the parameters, which then translates to regularity of the quadratic cost function and thus its Hessian, as shown in Lemmas 3.4 and 5.3.

Positive definiteness of the Hessian at the ground truth in assumption (A3) ensures sensitivity based identifiability, if the full data under D^{full} is accessed. This implies local structural identifiability and excludes the worst case scenario, where any data is uninformative about the parameter. Vice versa, if structural identifiability holds, i.e. the parameter is identifiable from the input-to-output map, then chances are good that a probability measure μ on the (possibly infinite dimensional) domain of the input-to-output map, the set $\mathcal{D}$ of all considered experimental designs, can be constructed such that (A3) holds. As we do not want to dive into the technicalities of infinite dimensional probability measures, we keep this discussion on an intuitive level.

Together with (A2), a straightforward extension of Proposition 2.11 for possibly infinite designs in the realm of quasimatrices then ensures local strong convexity of the cost function $\mathcal{C}^{\text{full}}$ at $p_\star$.

8.2.2. Experimental Design as Sampling

In this setting, the question of experimental design can be approached through searching a small subset of experimental setups $D_L \subset D^{\text{full}}$ of finite size $L \geq Q$, according to Proposition 2.13, that still generates a sensitivity based identifiable inverse problem. Experimental design thus becomes a down-sampling problem, with the goal to preserve positivity (or alternatively the conditioning) of Gauss Newton Hessian of the quadratic cost

$$H_p\mathcal{C}^{\text{full}}(p_\star) = \int_{D^{\text{full}}} J_s \otimes J_s \,\mathrm{d}\mu(s) \succ 0.$$

The main challenge remains to define a sampling distribution that achieves preservation under a small number L of experimental setups.

Remark 8.5. Our approach is can be regarded as a generalized preconditioning strategy: In case of a finite D^{full}, we are searching a sparse preconditioning matrix $M \in \mathbb{R}^{L\times|D^{\text{full}}|}$, with only one non-zero entry in each row, for which the preconditioned cost $\|M(Jp-y)\|_2^2$ attains a similar local convexity structure as the unconditioned problem. This shape of M extracts single rows of J and allows linking the preconditioner to experimental setup selection. Statistical methods allow us to build M efficiently and to generalize to quasimatrices.

8.2.3. Importance Sampling and Matrix Sketching

The quasimatrix sketching approach from Section 8.1 suggests to sample s_l from a distribution $\rho\mu$ where ρ approximates the relative size importance density $\tilde{\rho}$ in (8.4), so to obtain the approximation

$$H_p\mathcal{C}^{\text{full}}(p_\star) = \int_{D^{\text{full}}} (\rho_s^{-1} J_s \otimes J_s)\rho_s \,\mathrm{d}\mu(s) \quad \approx \quad \frac{1}{L}\sum_{l=1}^{L} \rho_{s_l}^{-1} J_{s_l} \otimes J_{s_l} = H_p\tilde{\mathcal{C}}(p_\star).$$

Note that the sample sum on the RHS can be regarded as the Gauss-Newton Hessian of the weighted quadratic cost

$$\tilde{\mathcal{C}}(p) = \frac{1}{2L}\sum_{l=1}^{L} (\rho_{s_l})^{-1}(F_{s_l}(p) - y_{s_l})^2 \quad \text{associated to design} \quad D_L = \{s_l\}_{l=1}^L \text{ with } s_l \sim \rho\mu. \tag{8.7}$$

Hence, if $H_p\tilde{\mathcal{C}}(p_\star)$ is sufficiently close to $H_p\mathcal{C}^{\text{full}}(p_\star) \succ 0$, then Lipschitz continuity of eigenvalues w.r.t. matrix perturbation [HJ85] ensures positivity of $H_p\tilde{\mathcal{C}}(p_\star) \succ 0$, i.e. local strong convexity of cost function $\tilde{\mathcal{C}}$ and sensitivity based identifiability of the respective inverse problem.

Theoretical guarantees on the matrix approximation quality with high probability under a sufficiently high number L thus translate to local identifiability guarantees for experimental setups in our context, as summarized by the following adaptation of Theorem 8.2.

Theorem 8.6. *Consider an inverse problem (8.6) that satisfies Assumption 8.4. Moreover, assume that J is bounded in Frobenius norm $\|J\|_F < \infty$, and consider the cost function $\tilde{\mathcal{C}}$ from (8.7), where D_L is sampled from D^{full} by i.i.d. draws with a sampling distribution $\rho\mu$ on D^{full}, whose density ρ satisfies $\int_{D^{\text{full}}} \rho \, \mathrm{d}\mu = 1$ and*

$$\rho \geq \beta\tilde{\rho} = \beta \frac{\|J_s\|_2^2}{\|J\|_F^2} \tag{8.8}$$

for some $\beta \in (0, 1]$. Then, if L is sufficiently large, the data under D_L is sensitive to the parameter and $\tilde{\mathcal{C}}$ is locally strongly convex at the ground truth parameter $p_\star$ with high probability.

To be more precise, denoting by $\lambda_{\min}^{\text{full}}, \lambda_{\max}^{\text{full}}, c^{\text{full}}$ and $\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max}, \tilde{c}$ the minimal and maximal eigenvalue and condition number of $H_p\mathcal{C}^{\text{full}}(p_\star)$ and $H_p\tilde{\mathcal{C}}(p_\star)$, respectively, then for any failure probability $\delta \in (0, 1)$ and accuracy $\varepsilon \in (0, \lambda_{\min}^{\text{full}})$, a choice of

$$L \geq \|J\|_F^4 \frac{(1 + \sqrt{8\beta \log(\delta^{-1})})^2}{\beta\varepsilon^2}$$

is sufficient to ensure that with probability at least $1 - \delta$, the cost function $\tilde{\mathcal{C}}$ is locally strongly convex at $p_\star$ with minimum Hessian eigenvalue $\tilde{\lambda}_{\min} \geq \lambda_{\min}^{\text{full}} - \varepsilon > 0$ and the condition number of its Hessian is close to the full data setting $\tilde{c} \leq (c^{\text{full}} + \frac{\varepsilon}{\lambda_{\min}^{\text{full}}})(1 - \frac{\varepsilon}{\lambda_{\min}^{\text{full}}})^{-1}$.

The importance density $\tilde{\rho} \propto \|J_s\|_2 = \|\nabla_p F_s(p_\star)\|_2$ suggests that the sampling distribution is skewed towards favouring those experimental setups s, that are the most sensitive w.r.t. the parameter locally in $p_\star$, in the sense that their gradient is large and thus a small deviation in the parameter, in a suitable direction, will have a relatively big impact on the forward map. With this and Assumption (A3) in mind, it makes sense that the down-sampled inverse problem is also likely to be sensitive to the parameter entries, if a sufficient number of experiments can be conducted.

Although this theorem provides an explicit bound for this number L, the bound is of limited use in practical applications, where typically $\|J\|_F$ as well as $\lambda_{\min}^{\text{full}}$ are unknown due to the high computational effort linked to their computation. Furthermore, the bound on L is a worst case bound, and we expect that a significantly lower number of samples will already perform well in many cases. In this sense, the result must rather be understood as an 'existence' result. This is good news for applications where the number of sensors is frequently constrained by the experimental restrictions or economical considerations.

Again, approximating $\tilde{\rho}$ by some $\rho \geq \beta\tilde{\rho}$ with $\beta < 1$ is possible, but might entail a higher sample size, i.e. a larger number of experiments, or a smaller success probability.

Proof. By an extension of Proposition 2.11, local strong convexity of $\tilde{\mathcal{C}}$ follows from $H_p\tilde{\mathcal{C}}(p_\star) \succ 0$. Again, continuity of the eigenvalues w.r.t. matrix perturbation [HJ85] provides the lower bound

$$\tilde{\lambda}_{\min} \quad \geq \quad \lambda_{\min}^{\text{full}} - |\lambda_{\min}^{\text{full}} - \tilde{\lambda}_{\min}| \quad \geq \quad \lambda_{\min}^{\text{full}} - \|H_p\mathcal{C}^{\text{full}}(p_\star) - H_p\tilde{\mathcal{C}}(p_\star)\|_F \, .$$

Positivity of the first term on the RHS holds by Assumption (A3) on the positivity of $H_p\mathcal{C}^{\text{full}}(p_\star) \succ 0$, and it remains to bound the Hessian difference term. By construction of $\tilde{\mathcal{C}}$, Theorem 8.2 can be applied and it shows that with probability at least $1-\delta$

$$\|H_p\mathcal{C}^{\text{full}}(p_\star) - H_p\tilde{\mathcal{C}}(p_\star)\|_F = \|J^TJ - \tilde{J}^T\tilde{J}\|_F \leq \frac{1+\sqrt{8\beta^{-1}\log(\delta^{-1})}}{\sqrt{\beta L}}\|J\|_F^2.$$

The bound for L then follows from bounding the above by ε so to achieve $\tilde{\lambda}_{\min} \geq \lambda_{\min}^{\text{full}} - \varepsilon > 0$, and the bound for the condition number $c(H_p\tilde{\mathcal{C}})$ follows by analogous estimations for the maximal eigenvalue $\tilde{\lambda}_{\max}$. □

Note that another viable way, as suggested by the Monte Carlo method, is to simply draw i.i.d. random samples $r_l \in D^{\text{full}}$ from the distribution μ, and approximate the Hessian through (8.2). This approach leads to a quadratic cost function $\mathcal{C}_{D_L}(p) = \frac{1}{L}\sum_{l=1}^{L}(F_{r_l}(p) - y_{r_l})^2$. However, the method typically requires a large number of samples for similar accuracy of the Hessian approximation under the same probability, and hence to ensure local identifiability. Keeping in mind that experimentation is often expensive, we base our analysis on importance sampling that often requires a smaller sample to attain the same approximation accuracy, as described in Section 8.1 - at the cost of a more complex sampling distribution that then requires sophisticated sampling algorithms.

Remark 8.7. So far the cost function did not involve any regularization term $R(p)$. A regularized cost function of the form $\mathcal{C}_R = \mathcal{C}(p) + R(p)$ will lead to similar results, with small modifications, as long as the regularizer R attains sufficient regularity.

8.2.4. Choice of Sampling Algorithms

In the practical context, drawing samples from non standard distributions is challenging. Hence, we share some practical considerations about the implementation of sampling.

Sampling from an Approximate Distribution. Theorem 8.6 leaves flexibility of using an approximation ρ to the optimal sampling density $\tilde{\rho}$ with $\beta < 1$. In certain situations, insight into the forward model might allow the derivation of an approximation ρ. If, for instance, one was able to show that the $\|J_s\|_2$ are uniformly bounded from above by a moderately large constant, this would suggest a uniform density as a suitable approximation. As could be observed in our numerical test in Figure 8.6, this can yield satisfactory results. In general, however, these simple densities might not capture a complex $\tilde{\rho}$ well, and deriving a good approximation with reasonably large β might be challenging itself and require model insight, as well as analytical manipulations. Sampling will still require calling specific sampling algorithms for more complex approximation distributions $\rho\mu$, depending on their specific properties.

Sampling from the Importance Sampling Distribution. The alternative is (to attempt) direct sampling from the importance sampling distribution $\tilde{\mu} := \tilde{\rho}\mu$. Because this distribution is typically highly non-standard and very model dependent, as can be observed in

Figures 8.2 and 8.3, this requires more sophisticated sampling techniques. We propose the application of samplers developed for Bayesian posterior sampling, as briefly introduced in Section 2.3.1, for two simple reasons:

- The structure of the sampling distribution $\tilde{\mu} = \tilde{\rho}\mu$ resembles that of a Bayesian posterior distribution in (2.7), and
- the same challenges as for Bayesian posterior sampling arise: the computation of the density $\tilde{\rho} \propto \|J_s\|_2 = \|\nabla_p F_s(p_\star)\|_2$ requires solving the forward (and adjoint) model, and is often expensive - especially for PDE models, which restricts computational capabilities to a few evaluations, and suggests avoiding the calculation of the normalization constant $\|J\|_F$, calling for solvers that satisfy (P1)-(P2).

Both, classical MCMC methods as well as the newly developed ensemble samplers can be deployed, and the concrete choice of the sampling algorithm is usually affected by many other factors in addition to their advantages and disadvantages in computational cost and theoretical convergence guarantees, as described in Section 2.3.1. For instance, a discrete or function space structure of D^{full} would require specific extension of the standard algorithms that were originally developed on continuous domains, as available for MCMC methods in [RG22, BS09], for instance. Moreover, external factors such as the desired balance between computational cost of the experimental design sampling and experimental cost might provide guidance, whether a cheaper but possibly less accurate computation with gradient-free, ensemble methods should be favoured over expensive MCMC methods with well established theoretical high accuracy guarantees.

Although the presentation of the ensemble samplers in Section 2.3.1 allows direct application to our design sampling strategy in case of a probability measure $\mu(s) = q(s)\,\mathrm{d}s$ on a continuous D^{full} with sufficiently regular Lebesgue density q by setting $\Phi(s) = -\log(\|J_s\|_2^2 q(s))$, it should be mentioned that the approximation of the gradient by the difference term may not be very accurate, if J_s is not almost linear in s or the sample size is small. Due to the lack of exhaustive non-asymptotic convergence results, convergence of the sample distribution to $\tilde{\mu}$ can in general not be guaranteed.

However, sampling most accurately from $\tilde{\mu}$ in order to obtain high alignment with the original Gauss-Newton Hessian is not our primary goal, but rather improving the convexity of the weighted quadratic cost $\tilde{C}$. In that sense, ensemble methods might still yield satisfactory results in numerical experiments as observed in Section 8.3.2. Intuitively, these cases can be understood as examples where the approximate sampling distribution $\rho\mu$ in Theorem 8.6 is generated through (non optimal) sampling by the ensemble methods, whose target distribution is still somewhat informed by the model, even if it does not coincide with the optimal $\tilde{\mu} = \tilde{\rho}\mu$.

8.2.5. Greedy Sampling

In order to leverage the previous thought, the samplers will be combined with a selection process that accepts or rejects newly proposed samples according to a criterion that

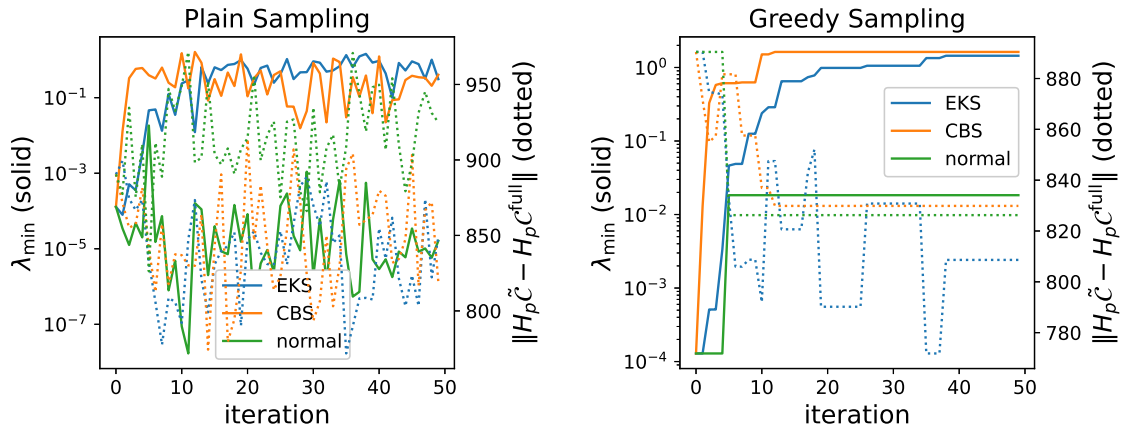


Figure 8.1.: Evolution of the minimal eigenvalue $\lambda_{\min}$ (solid lines) and deviation of the down-sampled Hessians to the full Hessian $\|H_p\tilde{\mathcal{C}} - H_p\mathcal{C}^{\text{full}}\|_F$ (dotted) of inverse problem (8.10) for designs D_L generated by three different sampling strategies: EKS (blue), CBS (orange) sampling, as well as repeated random sampling from the initial normal distribution (green), without (left panel) and with (right panel) greedy mechanism. All three sampling methods share the same initial configuration.

measures cost function convexity. Similar rejection mechanisms are well established and have been successfully implemented to improve certain quantities of samplers, for instance to remove the bias in MCMC samples by adding an additional Metropolis-Hastings step [SWZ23].

For the purpose of this work, we choose another criterion that evaluates convexity of the down-sampled cost function, and update the sample only if the newly proposed sample improves local convexity, measured for instance in terms of the minimum eigenvalue or the conditioning of the Gauss-Newton Hessian - in analogy to the E- or K- optimality criteria of optimal experimental design in Section 2.5. Numerical examples show that this minimal additional effort significantly improves the convexity of the down-sampled cost function, up to a degree where it even exceeds the full data setting in Figures 8.5, 8.6 and 8.9.

This mechanism can be regarded an early stopping mechanism for the sampling algorithm at favourable configurations, to cope with the oscillations in the eigenvalue structure introduced by the random samplers, as observed in Figure 8.1.

8.3. Application to the Schrödinger Potential Reconstruction

The performance of our method will be demonstrated on a specific example: the inverse problem of reconstructing a non negative potential $p \in C^\infty_+(X)$ from pointwise evaluations of the solution u_p to the steady state Schrödinger equation on the quadratic spatial domain $X = (-1, 1)^2 \subset \mathbb{R}^2$:

$$\begin{aligned} (-\Delta + p)u_p &= \gamma \qquad \text{for } x \in X, \\ u_p &= 0 \qquad\quad \text{on } x \in \partial X. \end{aligned} \tag{8.9}$$

Vanishing boundary conditions and a constant source term $\gamma = 10^4$ complement the PDE. The large size of γ is chosen to facilitate readability of subsequent values, and smoothness of p is chosen for convenience.

This model describes diffusion of a density u_p and decay according to a (reaction) rate, given by the potential p. The source term corresponds to a continuous injection of mass into the system. The model belongs to the broad class of reaction diffusion models that describe various phenomena, for instance the chemical concentration dynamics in a chemical reaction process [DAB03], the spread of diseases [CM81], or the propagation of a population density, e.g. of bacteria, under reproduction and death [KS70], each of which attains its own interpretation for the parameter p.

Existence of a unique solution $u_p \in C^\infty_+(X)$ to (8.9) that is positive $u_p > 0$ in X follows from standard elliptic theory [Bre11, Eva22]. Details are placed in Proposition A.2.2 in Appendix A.2.

The inverse problem thus reads

$$\text{find } p \in \mathcal{A}_p \qquad \text{such that} \qquad F_x(p) := u_p(x) = y(x) \quad \text{for all } x \in D, \tag{8.10}$$

where the design $D \subset X$ collects all locations at which u_p is tested, and $\mathcal{A}_p$ denotes the chosen admissible set.

Lemma 8.8. *The inverse problem (8.10) is structurally identifiable, even for large admissible sets such as $\mathcal{A}_p = C^\infty_+(X)$.*

Proof. If u_p can be observed on the full domain, i.e. $D = X$, then its derivatives can be built and p can be reconstructed point wise by the simple formula $p = \frac{\gamma + \Delta u_p}{u_p}$, where positivity of u_p holds according to Proposition A.2.2. □

Challenges arise when only a finite number of measurements of u_p are available, in which case only a finite dimensional parameter can be expected to be recovered, as suggested for instance by Proposition 2.13. Naturally, the choice of measurement locations determines identifiability: measurements on the boundary where $u_p = 0$, for instance, do not carry any information on p. The goal in this section is thus to find good measurement locations for a prescribed form of the admissible set.

Parameter discretization. Given a finite set of smooth basis functions $\{b_q : X \to \mathbb{R}\}_{q=0}^{Q-1}$, the admissible set for the parameter p collects their linear combinations that attain non negative values on X

$$\mathcal{A}_p := \left\{ p : X \to \mathbb{R}_0^+, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto p(x) = \sum_{q=0}^{Q-1} p_q b_q(x_1, x_2) \quad \text{for some } p_q \in \mathbb{R} \right\}. \tag{8.11}$$

In numerical experiments, this basis will be fixed to

$$\{b_{q_1,q_2}(x_1, x_2) = \cos(q_1 \pi x_1)\cos(q_2 \pi x_2)\}_{q_1,q_2=0}^{\sqrt{Q}-1} . \tag{8.12}$$

Experimental Setup. Noise free data $\{y(x) = F_x(p_\star) = u_{p_\star}(x)\}_{x\in D}$ shall be considered that was generated by a ground truth parameter $p_\star \in \mathcal{A}_p$, reflecting Assumption (A1).

The selection of a suitable design then translates to defining the number $L = |D_L|$ and locations $D_L \subset X$ of sensors to be placed in the domain, with the goal to obtain a locally strongly convex associated cost function.

Space Discretization. The solution to (8.9) will be accessed on an equidistant Cartesian grid $\mathtt{X} = \{\mathtt{x}_n, n = 1, ..., (N_x + 1)^2\}$, where N_x cells are placed in every direction.

Full Measurement Setup. Define the full data setup as pointwise measurements on all interior vertices $D^{\text{full}} = \mathtt{X}\backslash\partial X$, which amounts to a total number of $|D^{\text{full}}| = N = (N_x-1)^2$ measurements. The cost function reads

$$\mathcal{C}^{\text{full}}(p) = \frac{1}{2N} \sum_{x\in D^{\text{full}}} |u_p(x) - y(x)|^2 \ .$$

Computation of $J_{\hat{x}}$. Application of sampling algorithms, such as the EKS or the CBS, require evaluation of the negative log density $\Phi(\hat{x}) = -\log(\|J_{\hat{x}}\|_2^2)$ of $\tilde{\rho}$ from (8.4), which in turn amounts in computing the gradient $J_{\hat{x}} = \nabla_p F_{\hat{x}}(p_\star) = \nabla_p u_{p_\star}(\hat{x})$ at different $\hat{x} \in X$. An adjoint gradient method, as spelled out in Section 8.5, provides an efficient approach that computes the q-th entry of the gradient as

$$[J_{\hat{x}}]_q = \partial_{p_q} u_{p_\star}(\hat{x}) = \langle g_{\hat{x}}, b_q u_{p_\star}\rangle_{L^2(X)},$$

where $g_{\hat{x}}$ solves the adjoint equation corresponding to point wise measurement at location $\hat{x} \in X$

$$-\Delta g_{\hat{x}} + p g_{\hat{x}} = -\delta_{\hat{x}} \quad \text{in } X, \qquad \text{and} \qquad g_{\hat{x}} = 0 \quad \text{on } \partial X. \tag{8.13}$$

The singularity in the source reflects the singular point wise measurement of u_p. It reduces regularity of solutions and requires introduction of a new solution concept. Existence of solutions $g_{\hat{x}} \in L^1(X)$ to (8.13) in the sense of Definition A.2.3 is established in Proposition A.2.4 in Appendix A.2.

Numerically, the forward and the adjoint Schrödinger equation (8.9) and (8.13) will both be solved by a finite element method with nodal basis defined on the equidistant Cartesian grid $\mathtt{X}$. More details are provided in Appendix C.2.

8.3.1. Importance Sampling Distributions.

The optimal sampling distribution $\tilde{\mu} = \tilde{\rho}N^{-1}$ describes good positions for the sensors. It will be investigated numerically numerically, by plotting it on a fine discretization of the spatial domain X. In the four subsequent examples, the dimension of the parameter is fixed to $Q = 9$, and the parameter values $(p_{q_1,q_2})_{q_1,q_2=0,1,2}$ that characterize the chosen ground truth parameters corresponding to the basis (8.12) are displayed in Table 8.1.

Figure 8.2 demonstrates that the optimal sampling distribution $\tilde{\mu}$ admits significant dependence on the underlying ground truth parameter. Qualitative differences in the shapes of the sampling distributions are clearly observable.

System	ground truth parameter
A	$p_\star^A = \begin{pmatrix} 13.6 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 10 \end{pmatrix}$
B	$p_\star^B = \begin{pmatrix} 5.856 & 0.103 & 3.168 \\ 3.7441 & 2.493 & 1.124 \\ 0.9902 & 3.803 & 0.846 \end{pmatrix}$

System	ground truth parameter
C	$p_\star^C = \begin{pmatrix} 11 & 8.889 & 7.778 \\ 6.667 & 5.556 & 4.444 \\ 3.333 & 2.222 & 1.111 \end{pmatrix}$
E	$p_\star^E = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

Table 8.1.: Scenarios under which the optimal sampling strategy $\tilde{\mu}$ is studied. The (i, j) entry of the matrix encodes the coefficient $p_{i,j}$.

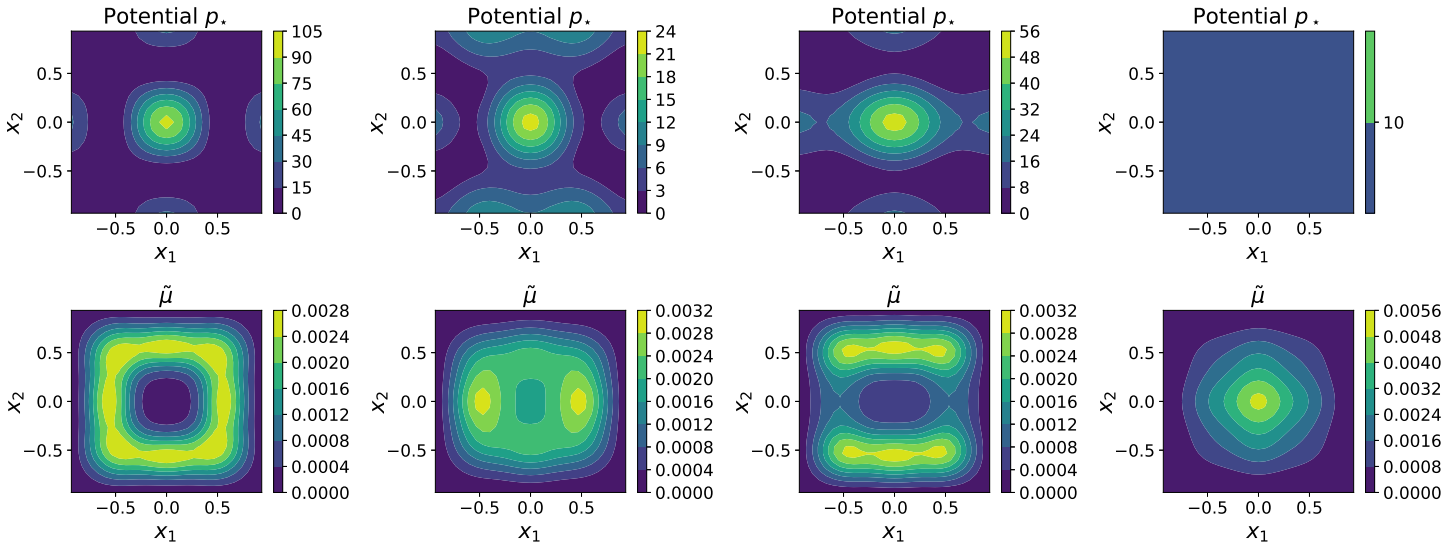
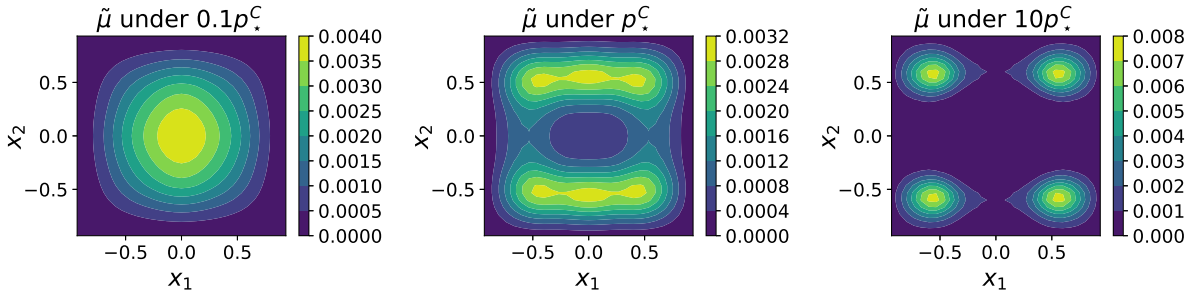
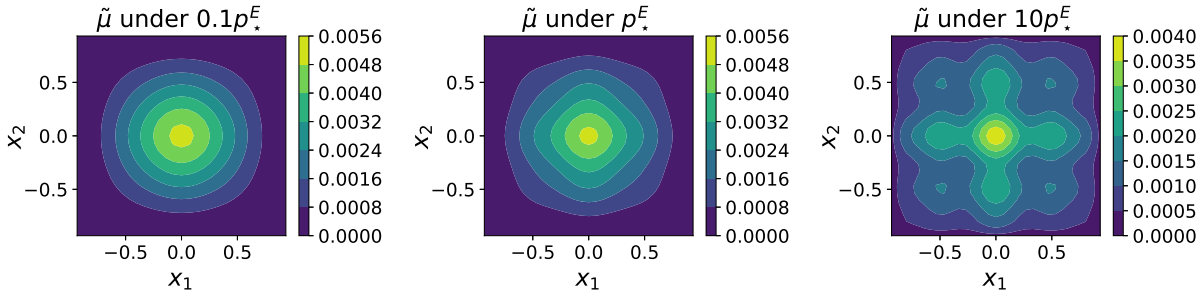


Figure 8.2.: Four different ground-truth media $p_\star$ (top row), ordered alphabetically by $p_\star^A, p_\star^B, p_\star^C, p_\star^E$ from right to left, lead to four different corresponding optimal sampling distribution $\tilde{\mu}$ (bottom row).

To test whether these qualitative differences originate from the qualitatively different shape of the ground truth parameters $p_\star$, the latter are scaled by a scaling factor $\alpha \in \{0.1, 1, 10\}$ that varies their amplitude, and the respective $\tilde{\mu}$ landscapes are plotted in Figure 8.3. Even under the same qualitative shape of $p_\star$, the sampling distribution $\tilde{\mu}$ varies significantly. It is striking that $\tilde{\mu}$ is more centered in the middle of the domain when $p_\star$ values are small, whereas interesting patterns, also reaching towards the boundaries, evolve for large scaling values.



(a) System C.



(b) System E.

Figure 8.3.: Optimal sampling distributions $\tilde{\mu}$ for scaled ground truth parameters $\alpha p_\star$ with $\alpha = 0.1$ (left), $\alpha = 1$ (center) and $\alpha = 10$ (right), where the ground truth parameters $p_\star$ from System C and E from Table 8.1 are taken.

8.3.2. Effect of Sampling

In the following, we deploy the thus developed design sampling strategy for the recovery of good sensor locations in the inverse Schrödinger potential reconstruction (8.10) as a proof of concept. Local convexity of the respective cost landscapes will be examined through the minimal eigenvalue of the Gauss-Newton Hessian, as well as its inverse condition number $c_{\text{inv}}(H) = \frac{\lambda_{\min}(H)}{\lambda_{\max}(H)}$. As an updating criterion for the greedy strategy, we chose an increase in the minimum eigenvalue.

Effect on Sensor Locations and Hessian Eigenvalues. To study the effect of sampling of the sensor locations, from now on, the ground truth parameter of system C in Table 8.1 will be fixed. The numerical discretization is set to $N_x = 30$ cells in each direction.

The starting point is marked in Figure 8.4: the red dots mark all $N = (N_x - 1)^2 = 841$ sensor locations in the full data design D^{full}, against the optimal sampling distribution $\tilde{\mu}$ in the background. The minimal eigenvalue of the Gauss Newton Hessian in this setting is strictly positive $\lambda^{\text{full}}_{\min} = 0.8 > 0$ and the inverse condition number is $8.18 \cdot 10^{-4}$, indicating that the problem is locally strongly convex. Given that $Q = 9$, it can be suspected that not all $N = 841$ sensor locations are required, and our down-sampling strategy shall be applied to find a representative subset, for which we allow only $L = 18 = 2Q$ sensor locations.

An initial guess D^{init}_{18} of narrowly centred, normally distributed sensor locations in X

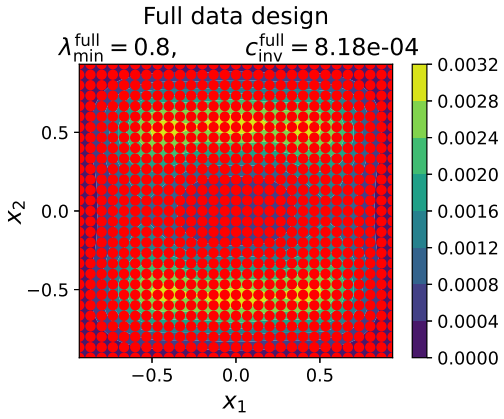


Figure 8.4.: Full Data Design D^{full}: Measurement locations (red dots) are located in all grid points. The optimal importance sampling distribution $\tilde{\mu}$ is drawn in the background.

performs poorly, as the minimal eigenvalue and the inverse conditioning of Gauss Newton Hessian degrade to $1.29 \cdot 10^{-4}$ and $1.26 \cdot 10^{-7}$, respectively.

Running greedy versions of the EKS in [GIHLS20] and the CBS from [CHSV22], as described in Sections 2.3.1, 8.2.4 and 8.2.5, to sample from the optimal sampling distribution $\tilde{\mu}$, instead, demonstrates the potential of our method to significantly improve this initial sample: after running 25 iterations, the sensor location sample has moved around the domain and the suggested locations are shown in Figure 8.5. This leads to an increase in the minimum eigenvalue and the conditioning of the corresponding Gauss Newton Hessians to $\lambda_{\min}^{EKS} = 0.99$ and $c_{\text{inv}}^{EKS} = 1.12 \cdot 10^{-3}$ and $\lambda_{\min}^{CBS} = 1.63$ and $c_{\text{inv}}^{CBS} = 1.75 \cdot 10^{-3}$, where the superscript denotes the sampling method.

Both methods are compared to a repeated greedy sampling strategy w.r.t. the initial guess distribution, to observe the influence the importance sampling distribution $\tilde{\mu}$. This strategy, though improving convexity up to a minimum eigenvalue and conditioning of the Hessian of $\lambda_{\min}^{\text{rand}} = 1.83 \cdot 10^{-2}$ and $c_{\text{inv}}^{\text{rand}} = 1.92 \cdot 10^{-5}$, clearly falls behind the performance of EKS and CBS, which confirms that the narrow centered Gaussian is not a good approximation to the importance sampling distribution $\tilde{\mu}$ as depicted in the background of Figure 8.5.

It is remarkable that the minimum eigenvalue, as well as the inverse condition number for the down-sampled Hessian, using our proposed strategy, even exceed that one of the full dataset D^{full}. This indicates that a large number of data might in fact hide more important data points and dilute the information on the parameter, hence diminishing the convexity of the cost landscape. Moreover, a better conditioning typically comes along with improved stability properties.

The evolution of the minimal eigenvalue, and the Hessian approximation error along the sampling iterations follows the behaviour in the right panel of Figure 8.1, that in fact originated from this setting (under an increased number of iterations). The resulting convexity measures of all considered designs are summarized in Table 8.2.

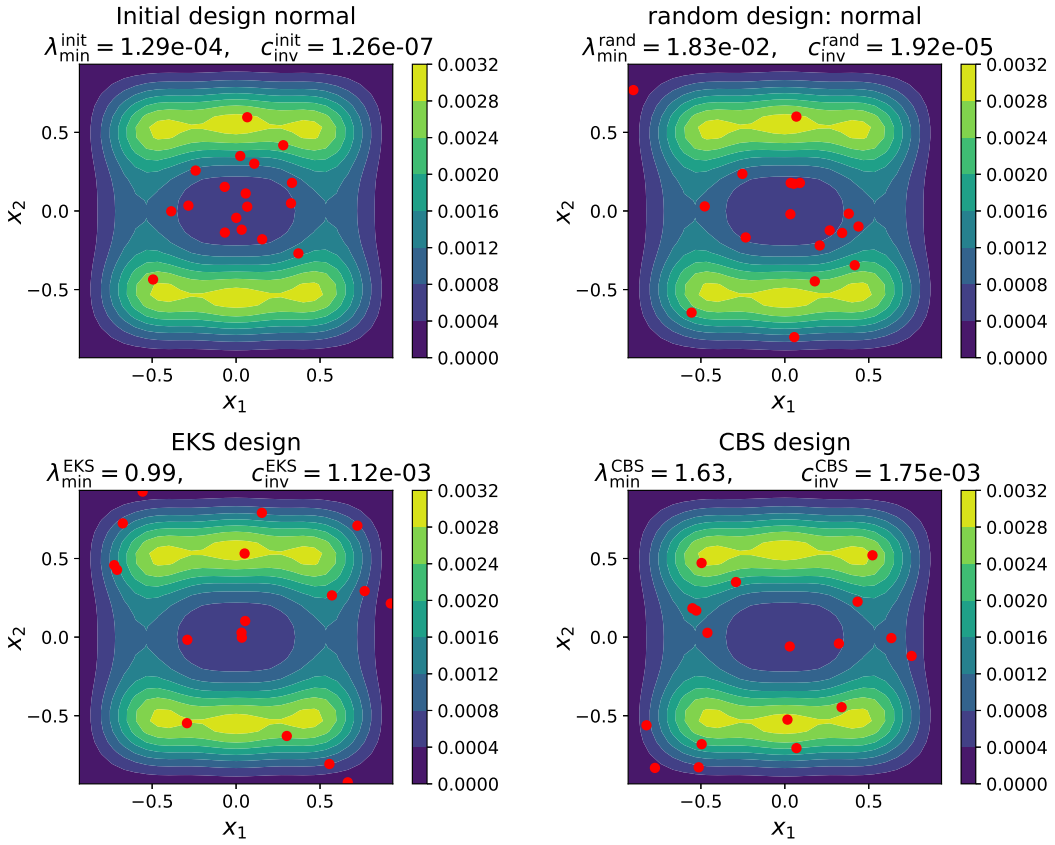


Figure 8.5.: Different designs and the convexity measures of their cost functions: sensor locations are depicted by red markers on the optimal sampling distribution in the background. Considered designs are the normally distributed initial guess (upper left panel), a greedy repeated sampling from this initial distribution (upper right), and the EKS and CBS samples from $\tilde{\mu}$ (lower left and right), each after 25 iterations.

Design D	$\lambda_{\min}$	c_{inv}
full data D^{full}	0.8	$8.18 \cdot 10^{-4}$
normal initial guess	$1.29 \cdot 10^{-4}$	$1.26 \cdot 10^{-7}$
EKS sample	0.99	$1.12 \cdot 10^{-3}$
CBS sample	1.63	$1.75 \cdot 10^{-3}$
greedy normal sampling	$1.83 \cdot 10^{-2}$	$1.92 \cdot 10^{-5}$
uniform initial guess	1.17	$1.24 \cdot 10^{-3}$
EKS sample	2.34	$2.54 \cdot 10^{-3}$
CBS sample	2.41	$2.1 \cdot 10^{-3}$
greedy normal sampling	2.73	$3.07 \cdot 10^{-3}$

Table 8.2.: Comparison of convexity measures of the cost function landscape associated to different designs. Rows below an initial guess refer to sampling starting from this initial guess.

Interestingly, in this scenario, uniformly distributed sensor locations, as depicted in Figure 8.6, already perform very well: the initial guess attains a minimum Hessian eigenvalue of $\lambda_{\text{min}}^{\text{init}} = 1.17$ and a conditioning of $c_{\text{inv}}^{\text{init}} = 1.24 \cdot 10^{-3}$. This can be anticipated, observing that the optimal importance sampling distribution $\tilde{\mu}$ is bounded from above by 0.0031. Comparing the value to a uniform distribution $\frac{1}{N} \approx 0.0012$, then a choice of $\beta = 0.383$ in (8.8) shows that the uniform distribution is a good approximation to $\tilde{\mu}$, potentially even a better one than the approximation generated by the EKS and CBS under a poorly chosen initial distribution and a non-Gaussian target. Starting from this uniform initial guess, a greedy sampling with the EKS and CBS from $\tilde{\mu}$, and repeated uniform sampling for 25 iterations further improve the minimum Hessian eigenvalue to 2.34, 2.41 and 2.73, and inverse condition numbers $2.54 \cdot 10^{-3}$, $2.66 \cdot 10^{-3}$ and $3.07 \cdot 10^{-3}$, respectively.

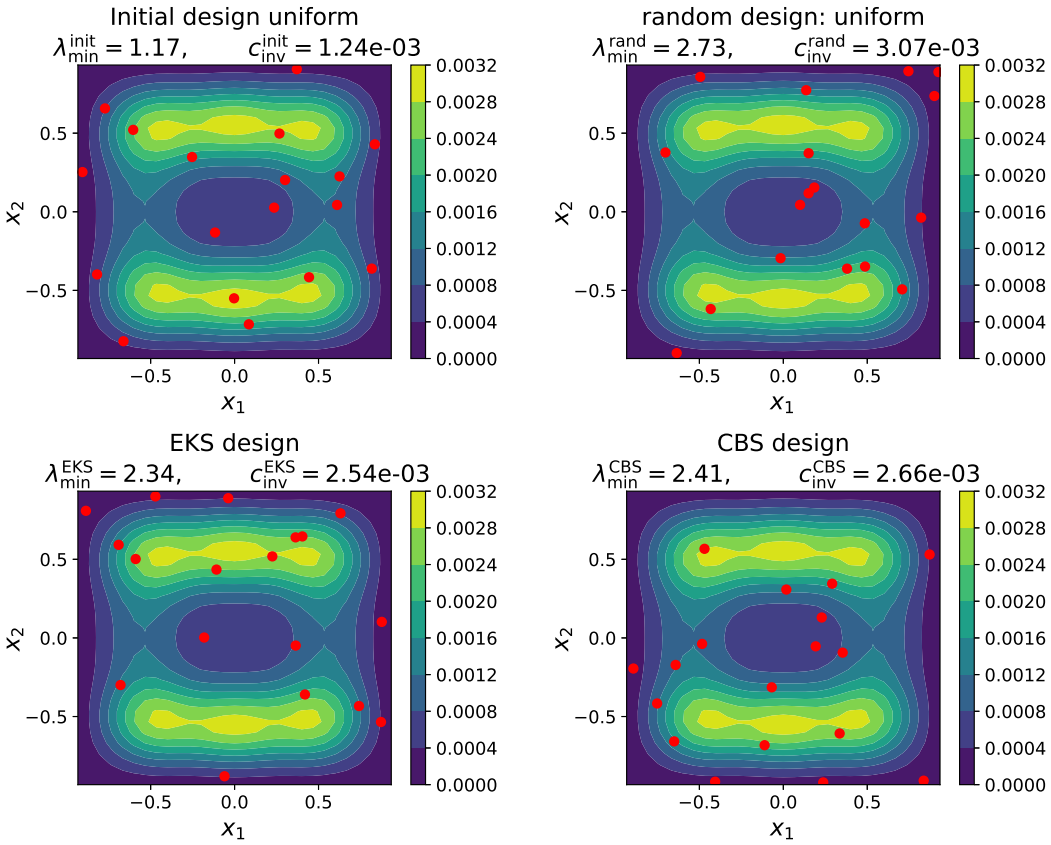


Figure 8.6.: Convexity measures for cost functions, corresponding to four different designs, whose sensor locations are depicted by red markers on the optimal sampling distribution in the background: a uniformly distributed initial guess (upper left), and three designs originating from greedy random sampling, either w.r.t. this initial distribution (upper right), with the EKS (lower left) or the CBS (lower right) from $\tilde{\mu}$, after 25 iterations each.

Effect on the Cost Function. In the previous paragraph, local convexity of the cost function around the ground truth, and its improvement by the proposed design sampling strategy, were explored by means of the minimum eigenvalue and the inverse conditioning

of the Gauss Newton Hessian. In the following, this effect shall be visualized globally in the parameter space. To plot the Q-dimensional cost function landscape, a reduced setting with a two dimensional admissible set

$$\mathcal{A}_p = \{p : X \to \mathbb{R}_0^+ \mid p(x) = p_1 \cos(\pi x_1) + p_2 \cos(\pi x_2) + 12\}$$

shall be considered and the ground truth parameter is fixed to $p_\star(x_1, x_2) = 1\cos(x_1) + 10\cos(x_2) + 12$. Its profile as well as the corresponding optimal importance sampling distribution $\tilde{\mu}$ are plotted in Figure 8.7. The different scaling of $p_\star$ in x_1 and x_2 direction leads to a stronger sensitivity of $p_\star$ w.r.t. x_2, which is also reflected in the sampling probability.

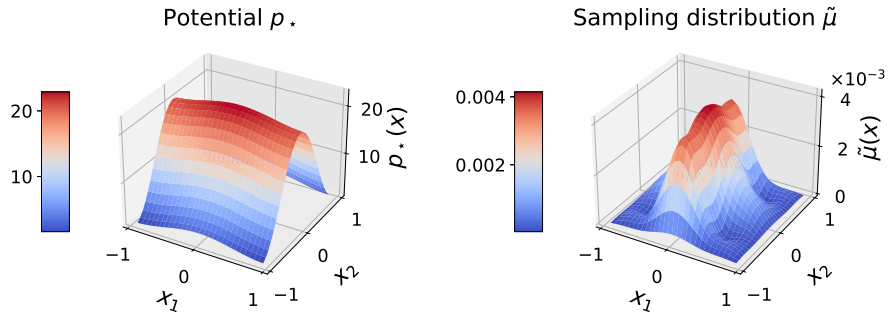


Figure 8.7.: Shape of the ground truth parameter $p_\star$ (left) and the optimal importance sampling distribution $\tilde{\mu}$ (right) in the two-dimensional setting.

The cost function landscape under the full data design D^{full}, as plotted in Figure 8.8, which demonstrates global strong convexity. In analogy to previous examples, the convexity is significantly reduced under an initial guess of $L = 6 = 3Q$ narrowly centred, normally distributed sensor location, which expresses through flatness of the cost function in one direction. An application of greedy greedy sampling from $\tilde{\mu}$ with the EKS and CBS clearly improves the global convexity of the cost landscape, as can be observed in Figure 8.9. Repeated greedy random sampling with the initial distribution also improves convexity a little, but fails to reach the full data configuration, and the flatness in the diagonal direction can not fully be mitigated.

8.3.3. Extension to Controllable Source

With a second example, we leave the well observable setting of a fixed constant source term 10^4, and instead consider a controllable function $\gamma(x) = \gamma_1 x_1 + \gamma_2 x_2 + 10$, with additional design parameter $(\gamma_1, \gamma_2) \in [-2, 2]^2$. By a slight abuse of notation, we identify the source function $\gamma(x)$ with its parameters $(\gamma_1, \gamma_2) \in [-2, 2]^2$ and denote both by γ. Then writing u_p^γ for the solution to (8.9) under the respective source γ, the inverse problem associated to an experimental design $D \subset [-2, 2]^2 \times X$ reads

$$\text{find } p \in \mathcal{A}_p \qquad \text{such that} \qquad F_s(p) := u_p^\gamma(x) = y_s \quad \text{for all } s = (\gamma, x) \in D, \tag{8.14}$$

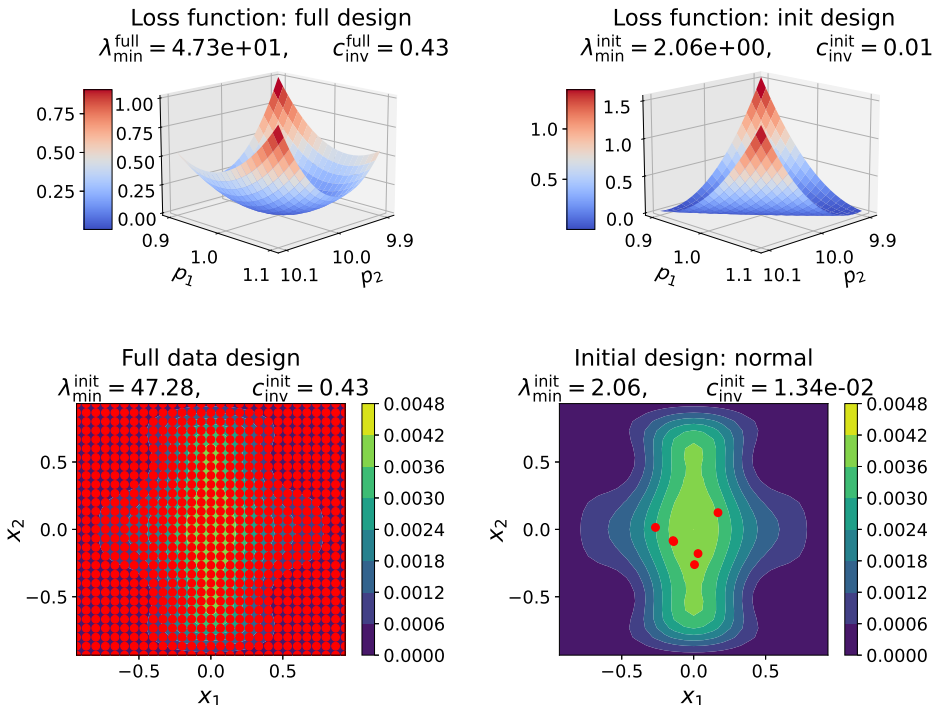


Figure 8.8.: Cost function landscapes (upper row) for different sensor locations (lower row): full data setup D^{full} (left) and normally distributed initial sensor locations D_6^{init} (right).

with admissible set $\mathcal{A}_p$ as in (8.11), and where the data $y_s := u_{p_\star}^{\gamma}(x)$ shall again be generated by a ground truth parameter $p_\star$, that will be chosen as scenario C in Table 8.1. Lemma 8.8 still provides structural identifiability of this inverse problem.

Choosing the same spatial discretization setting as previously, we can consider the full data setting $D^{\text{full}} = [-2,2]^2 \times \mathtt{X} \backslash \partial X$. Because accessing a large number of data with many different values of γ, each of which entailing a separate forward computation, is computationally costly, the computation of $\|J\|_F^2$ shall be avoided. Instead, we base the application of our sampling strategy on the unscaled sampling measure $\tilde{\mu}'(s) = \|J_s\|_2^2$. This also means that the displayed minimal eigenvalues refer to a scaled version of the cost function $\hat{\mathcal{C}}(p) = \|J\|_F^{-2}\tilde{\mathcal{C}}(p)$, and the performance of designs can not be compared to the unavailable full data setting.

Remark 8.9. Admittedly, this setting is somewhat exaggerated for the easy toy problem, given that Lemma 8.8 already suggests that access to all spatial measurements for only one source term $\gamma(x) \equiv \text{const} > 0$ is sufficient for the reconstruction. This additional model-based insight shall not enter our construction of D^{full} or the further sampling strategy, assuming an agnostic access to the inverse problem, and instead demonstrate how the algorithm can be applied, if $H_p\mathcal{C}^{\text{full}}(p_\star)$ is not available.

Figure 8.10 compares four designs and their convexity measures, where the source term parameters γ_1, γ_2 are encoded as the colour and size of the dots, respectively, and their location reflects the corresponding measurement location. The considered designs comprise:

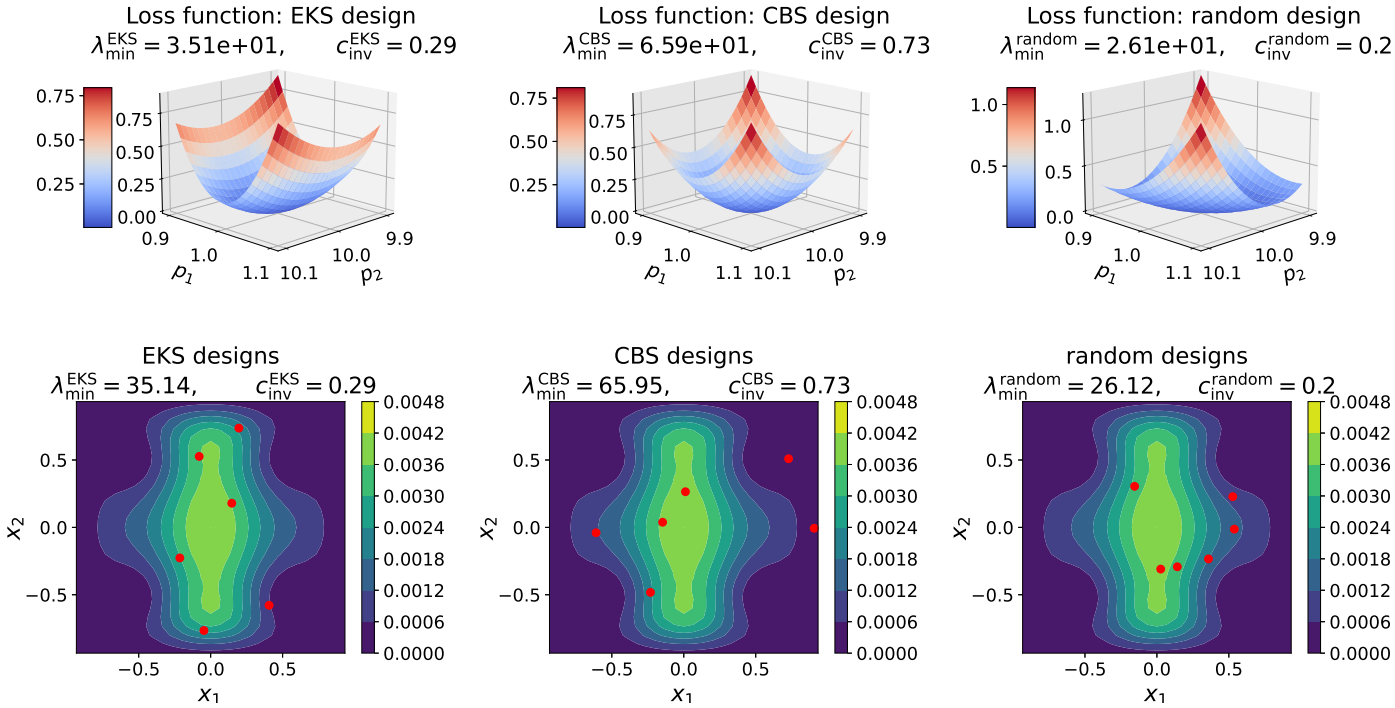


Figure 8.9.: Cost function landscapes (upper row) corresponding to different sensor locations, marked by red dots in front of the sampling distribution in the background (lower row), that were obtained from sampling according to greedy EKS (left) or CBS (middle) from $\tilde{\mu}$ and greedy repeated normal sampling (right) with the right panel of Figure 8.8 as initial setup.

- an initial guess of narrowly centered normally distributed sensor locations for experiments with randomly uniformly chosen $\gamma \in [-2, 2]^2$. It attains a very small minimum eigenvalue and inverse condition number of the Gauss Newton Hessian, compared to the other designs.

- a design derived from repeated greedy random sampling according to the initial guess distribution, which improves both convexity measures by an order of 10^3, or

- designs obtained from sampling with the EKS or the CBS according to the measure $\tilde{\mu}'$, which yields an improvement of convexity measures by five orders of magnitude, compared to the initial guess. Noting that the scale invariant inverse conditioning of the Hessian is then of the same order 10^{-3} as for the full data Gauss Newton Hessian for the fixed constant source term in the previous examples in Section 8.3.2, and knowing that fixed constant source term knowledge is sufficient for reconstruction by Lemma 8.8, this suggests that the found designs might already exhibit very similar local cost function convexity properties, and thus reconstruction performance, as the full data setting.

In total, 60 iterations were used for each sampling algorithm. The precise values of the

convexity measures for these designs, as well as the same experiment with uniformly distributed initial sensor locations are summarized in Table 8.3.

Design D	$\lambda_{\min}^D$	c_{inv}^D
normal initial guess	$2.95 \cdot 10^{-9}$	$3.27 \cdot 10^{-9}$
EKS sample	$9.34 \cdot 10^{-4}$	$1.24 \cdot 10^{-3}$
CBS sample	$6.22 \cdot 10^{-4}$	$8.4 \cdot 10^{-4}$
greedy normal sampling	$4.02 \cdot 10^{-6}$	$4.82 \cdot 10^{-6}$
uniform initial guess	$4.06 \cdot 10^{-4}$	$5.38 \cdot 10^{-4}$
EKS sample	$1.6 \cdot 10^{-3}$	$2.06 \cdot 10^{-3}$
CBS sample	$9.18 \cdot 10^{-4}$	$1.15 \cdot 10^{-3}$
greedy uniform sampling	$2.06 \cdot 10^{-3}$	$2.75 \cdot 10^{-3}$

Table 8.3.: Comparison of local convexity measures of the cost function landscapes under different designs. Rows below an initial guess refer to sampling starting from this initial configuration.

8.4. Conclusion

In this chapter, we proposed a mechanism to transition from a well identifiable parametrized inverse problem with much more data than parameters to reconstruct, to a small data inverse problem, by leveraging a sampling strategy to select informative data points. A shift of paradigm from searching an optimal experimental design to a sufficient one allows us to change the methodology from optimization to sampling. By leveraging the Gauss-Newton Hessian in a quadratic cost minimization framework as an indicator of identifiability, its specific structure allows adopting a well-studied matrix sketching algorithm from RNLA to propose an importance sampling distribution that selects experimental setups whose data is particularly sensitive w.r.t. the parameter. Via extending RNLA results to infinite dimensional quasimatrices, theoretical identifiability guarantees under high probability can be stated. To execute sampling from the model-informed importance sampling distribution in applications, we propose Bayesian posterior sampling algorithms, if no easy approximate distribution can be found, in combination with a greedy mechanism. This strategy is applicable to a broad class of experimental design tasks in inverse problems, and does not require deep model insight. Its power has been illustrated in numerical examples for the Schrödinger potential reconstruction problem.

We expect that the performance of our method can be improved when introducing advanced cooling schemes for hyperparameters or stopping rules for iterative samplers instead of the heuristically chosen values in our example, or more advanced samplers, for instance the polarized CBS [BRW24] which is able to cope with multi modal distributions.

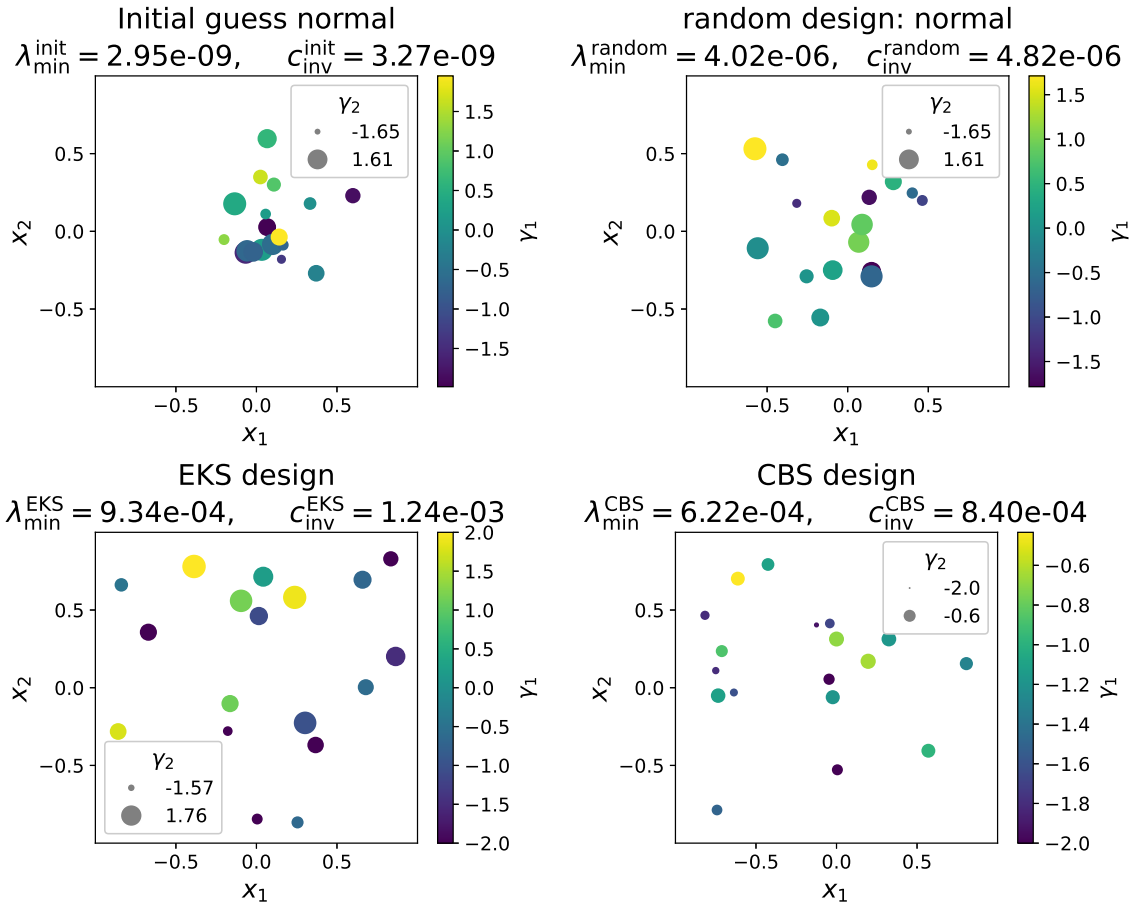


Figure 8.10.: Four different designs, with sensor locations given by the dot locations, and γ_1, γ_2 values encoded in colour and size of the dots, and their convexity measures: normally distributed initial sensor location guess with uniformly distributed γ_1, γ_2 (upper left), greedy repeated sampling w.r.t. this distribution (upper right), greedy EKS (lower left) and CBS (lower right) w.r.t. $\tilde{\mu}'$, after 60 iterations each.

The practical advantage of these methods should be evaluated while keeping in mind that optimal sampling is not the main focus of this project, but rather improving the cost function convexity.

A challenge in the application of our method to real inverse problems lies in the dependence of the sampling distribution $\tilde{\mu}$ on a usually unknown ground truth parameter $p_\star$. Because this is a commonality among all experimental design tasks, several strategies have been proposed to solve this problem [LMT15, Ale21, HJM24, BBKS00], as described in Section 2.5, the easiest being to derive the design based on the MAP point. Alternatively, designs are often based on a mean design criterion which is averaged over the parameter space possibly weighted by a prior distribution, or on a constantly updated parameter reconstruction in sequential optimal experimental design. We expect these strategies to be transferable to our design sampling setting, and even see synergies between a greedy sequential approach and the sampling algorithm: because sampling requires computation

of the gradient $J_s = \nabla_p F_s(p_\star)$, anyways, performing one gradient descend parameter update after each design iteration is basically for free and can readily be integrated in the design process.

Additionally, the relation between our sampling approach and existing methods should be further investigated, where we see similarities to

- classical optimization based design methods that utilize similar optimality criteria as deployed in the greedy mechanism, which already adds a certain portion of optimization to the method - for instance E- or K-optimality investigating the minimum eigenvalue or conditioning, respectively,
- measure based optimal experimental design that seeks to optimize a sensor location measure [NPVW19] which might be related to the transition from μ to $\tilde{\mu} = \tilde{\rho}\mu$, and
- least squares preconditioning, as mentioned in Remark 8.5.

Furthermore, it is worth mentioning that measurement noise is naturally taken into account in Bayesian optimal experimental design contexts, whereas an addition to our context is more involved, and only intuitively suggested in the spirit of sensitivity analysis, where a good cost function convexity leads to small confidence region for the parameter. A detailed investigation is left to future work.

Finally, we hope that our strategy can help bridging the gap between inverse problem theory, where structural identifiability analysis assumes access to the full input-to-output map, and practical finite data implementation: after discretizing the parameter, the found experimental design suggests which experiments shall be conducted for a successful reconstruction. A combination with a parameter sketching idea, as adopted in [JLNS24], might yield the missing step from an infinite dimensional non-parametric to the finite dimensional parametric admissible set to achieve an efficient sketching of the whole input-to-output map. This approach could then be compared to the strategy proposed in [BTLZNn22], and complemented by a similar matrix completion step to further improve numerical reconstruction.

8.5. Derivation of the formula for $\nabla_p u_p(\hat{x})$

Application of the EKS and CBS, as well as other sampling algorithms, requires computation of the log density $\Phi(\hat{x}) = -\log(\|J_{\hat{x}}\|_2^2) = -\log(\|\nabla_p u_{p_\star}(\hat{x})\|_2^2)$ and hence the gradient $\nabla_p u_{p_\star}(\hat{x})$. An adjoint based formula will be derived in the following, in analogy to the proof of Lemma 5.5 in Section 5.3.1.

For a fixed $\hat{x} \in X$, the Lagrangian considers the measurement $u_p(\hat{x})$ as a measurement of a generic function $u \in C^\infty(X)$ at $\hat{x}$ under the constraint that $u = u_p$ solves (8.9). The corresponding Lagrangian function can be defined as

$$\mathbb{L}_{\hat{x}} : \mathcal{A}_p \times C^\infty(X) \times L^1(X) \to \mathbb{R}, \qquad \mathbb{L}_{\hat{x}}(p, u, g) = u(\hat{x}) + \langle g, -\Delta u + pu - \gamma \rangle_{L^1(X), L^\infty(X)}$$

with Lagrange multiplier $g \in L^1(X)$, and where $\langle \cdot, \cdot \rangle_{L^1(X),L^\infty(X)}$ denotes the duality bracket on $L^1(X) \times L^\infty(X)$. By (8.9), inserting $u = u_p$ yields the measurement $\mathbb{L}_{\hat{x}}(p, u_p, g) = u_p(\hat{x})$. The chain rule then shows

$$\left.\frac{\partial u_p(\hat{x})}{\partial p_q}\right|_{p=\hat{p}} = \left.\frac{\partial \mathbb{L}_{\hat{x}}}{\partial p_q}\right|_{\substack{p=\hat{p} \\ u=u_{\hat{p}}}} + \left.\frac{\partial \mathbb{L}_{\hat{x}}}{\partial u}\right|_{\substack{p=\hat{p} \\ u=u_{\hat{p}}}} \left.\frac{\partial u_p}{\partial p_q}\right|_{p=\hat{p}}$$

for arbitrary $g \in L^1(X)$ and $q = 0, ..., Q-1$, in particular for such $g = \hat{g}$ for which $\partial \mathbb{L}_{\hat{x}} / \partial u = 0$. Using the parametric form of p according to (8.11), the partial derivative then becomes

$$\left.\frac{\partial u_p(\hat{x})}{\partial p_q}\right|_{p=\hat{p}} = \left.\frac{\partial \mathbb{L}_{\hat{x}}}{\partial p_q}\right|_{\substack{p=\hat{p} \\ u=u_{\hat{p}} \\ g=\hat{g}}} = \left.\frac{\partial \langle g, pu \rangle}{\partial p_q}\right|_{\substack{p=\hat{p} \\ u=u_{\hat{p}} \\ g=\hat{g}}} = \left.\frac{\partial \sum_{q'} p_{q'} \langle g, b_q u \rangle}{\partial p_q}\right|_{\substack{p=\hat{p} \\ u=u_{\hat{p}} \\ g=\hat{g}}} = \langle \hat{g}, b_q u_{\hat{p}} \rangle_{L^1(X),L^\infty(X)} .$$

In order to assemble the gradient according to the above formula, it remains to compute $\hat{g}$ that satisfies

$$0 = \left.\frac{\partial}{\partial u} \mathbb{L}_{\hat{x}}\right|_{g=\hat{g}} = \frac{\partial}{\partial u} \left[u(\hat{x}) + \langle \hat{g}, -\Delta u + pu \rangle \right] .$$

According to Definition A.2.3, the inner bracket vanishes for all $u \in C^\infty(X)$, if $\hat{g} = g_{\hat{x}}$ is the unique solution to the adjoint equation (8.13), with Proposition A.2.4 showing its existence.

Discussion

9

Discussion

In inverse problems, reducing the amount of data, for instance when going over from the theoretical input-to-output map to a practically feasible finite experimental design, is a delicate business, and a careful selection of the 'important' data, potentially combined with a compatible injection of a-priori knowledge on the parameter to bridge the missing information in the data, is necessary to sustain identifiability properties. The problem is classically treated by first choosing a parametric form of the model parameter and then applying methods of optimal experimental design [HJM24, APSG16] to find that experimental design that minimizes uncertainty in the parameter reconstruction.

In this work, we laid out two further approaches to sustain sensitivity based identifiability in the finite data setting, when the infinite data inverse problem is structurally identifiable.

Relaxation of Theory Approach. The first approach implemented the full transition from the non-parametric input-to-output map to the parametric finite experimental design, and emphasized the benefits of a compatible construction of the parameter discretization in the admissible set and the experimental design, as derived by a relaxation of a suitable theoretical structural identifiability proof.

Such a proof was developed for the inverse problem related to the kinetic chemotaxis model, where the tumbling kernel shall be reconstructed from macroscopic measurements. Following the singular decomposition technique, a delicate construction of sequences of experimental designs allowed to directly read off the tumbling kernel from the measurement and thus provided structural identifiability of a non-parametric tumbling kernel under access to the full input-to-output map. The design injected singular initial data to trigger the microscopic information in the system, which could then be tracked by a suitable singular short time measurement strategy, allowing the direct reading of the point evaluation of the tumbling kernel evaluation from the respective measurement.

In order to transfer identifiability to the finite data setting, we leveraged this construction to build a suitable experimental setting, consisting of a compatible pair of a finite dimensional admissible set, and a finite experimental design. Inspired by the point-wise reconstruction of the tumbling kernel from the experimental design that becomes locally

concentrated in the reconstruction point, we discretized the shape of K to a step function in space and velocity, reducing it to a finite dimension. The point wise value $K(x, v, v')$ to be recovered was hence attained at a larger area, leaving more space for the experimental setup to develop, and softening the requirement to reach the singular initial datum and vanishing measurement time limit. Moreover, a similar strategy as deployed in the theoretical structural identifiability proof allowed us to analytically prove sensitivity based identifiability for this design. Numerical examples quantified the sensitivity of this design at a high level through the strong convexity of the corresponding quadratic cost function, leading to fast and accurate reconstructions.

Despite the mathematical elegance in the unified treatment and methodology in both, the non-parametric theoretical inverse problem, as well as the parametric finite data setting, this approach, in general, suffers from the requirement for a constructive proof that already indicates an experimental design construction. In many cases, this is out of reach. We thus proposed another approach, which treats the model as a black box and requires not further insight, and is thus applicable to generic parameter identification problems.

Sampling Approach. The sampling approach assumed that the parametric form of the parameter to be inferred has been fixed beforehand, as adaptation to the model is not possible without further insight. Then, being rooted in the sensitivity based identifiability framework, local identifiability was expressed in terms of positivity or conditioning of the Gauss-Newton Hessian of the quadratic cost function. Its matrix product structure allowed utilizing a matrix sketching algorithm from RNLA to derive an importance sampling distribution of the experimental setups, encoded as rows of the sensitivity matrix, according to their sensitivity, and theoretical high probability guarantees on sensitivity and conditioning of the Gauss-Newton Hessian of the thus down-sampled design under a sufficiently high number of data carried over from RNLA theory. To realize sampling from these non-standard distributions, we proposed Bayesian posterior samplers in combination with a greedy strategy based on sensitivity measures, which proved very effective for of sensor placement (and source design) problem for the Schrödinger potential reconstruction.

Conclusion. In summary, these two techniques both represent a perspective onto experimental design construction that differs from optimal experimental design, as it relaxes the objective of obtaining an optimal design to finding a sufficient design that provides identifiability. Numerical experiments underline that both approaches describe viable alternatives.

The first approach is in line with the wide-spread mentality to discretize a problem as late as possible, and instead base theory and the setup of methods on the infinite dimensional setting [Stu10, Ale21]. This ensures that the finite parametric inverse problem attains a reasonable limit, as the discretization becomes arbitrarily fine, and renders our found design adaptive to the parameter discretization, at the cost of potentially very fine measurement requirements. As mentioned before, this approach suffers from a lack of availability of suitable theoretical results to be relaxed, and the derivation of constructive proofs, if possible, may entail significant additional efforts [FSU19].

The second approach, instead, is applicable to a large class of inverse problems, in particular those with non-constructive uniqueness proofs. It is closer to the optimal experimental design setting, as the parametric form is prescribed a priori, which allows an efficient 'black box' application. However, this hides the intricate balance between a priori information in the admissible set and the experimental design, which - if designed compatibly as in the first approach - might yield even more powerful inversion settings.

Outlook. We see many possible directions for future research: a combination of the design sampling strategy with a sketching of the parameter information, as proposed in [JLNS24], would lift the method to a sketching of the full infinite-to-infinite dimensional input-to-output map for a non-parametric model parameter and could establishing the full link between the theoretical non-parametric and the parametric finite data setting for this approach. Because the design sketching depends on the sensitivity and thus on the chosen parametric form of the model parameter, this might lead to automatic compatibility of admissible set and experimental design. Executing the sampling and the derivation of theoretical guarantees, however, can become very involved. We leave this for future work. The resulting sketching mechanism can then be compared to the strategy proposed in [BTLZNn22] for the electrical impedance tomography problem, and, as suggested in this publication, be combined with a matrix completion algorithm to build the bridge back from the finite setting to the non-parametric inverse problem and improve reconstruction.

A shortcoming of the thus presented methods is their dependence on the typically unknown ground truth parameter - where independence of the found designs (1D) and (2D) follows rather coincidentally from the transport nature of the model, and will in general not hold for other types of models. For the sketching approach, well-established strategies from optimal experimental design [HJM24] can be applied to tackle this problem, whereas it is obvious how to proceed in the model-based relaxation approach.

Undoubtedly, the numerical methods in this work leave room for further improvements: the implementation of the sketching approach could benefit from samplers with better capabilities to sample from multi modal distributions [BRW24, MS21], improved criteria for the greedy mechanism, or hyperparameter tuning might. On the other hand, convergence of the reconstruction algorithms, even under weak convexity, can be improved when more advanced optimization methods are invoked. However, referring back to Lanczos [Lan96], even the most sophisticated methods cannot cure intrinsic ill-posedness of the problem, suggesting that these methods can never replace experimental design to procure an informative data.

In addition to that, it should be noted that both considered approaches did not take measurement noise into account, but solely select designs according to their sensitivity. Optimal experimental design, instead, naturally embeds measurement noise through its consideration of uncertainty of the reconstruction as an objective. Nonetheless, sensitivity is in some sense related to the robustness of the reconstruction against noise, for instance by consideration of standard quadratic confidence intervals. Elaborating this connection for the presented approaches might help understanding the relation of the design sampling approach to the optimal experimental design methodology could reveal additional insight

into how stability is propagated to the finite dimensional setting under the 'relaxation of theory' approach.

Finally, we hope that experimental design - be it qualitative or optimal - will lead to more interaction between the experimenters and mathematicians involved in inverse problem solving, to disseminate a holistic perspective on inverse problems and instigate a collaborative design process where a-priori parameter information and data information is well balanced and takes measurement restrictions into account, to obtain meaningful reconstructions of the parameters and gain deeper insight into the world that we are surrounded by.

Appendices

A

Existence Theory

A.1. Existence Results for the Kinetic Chemotaxis Equation via Semigroup Theory

This section provides existence theory of the forward and adjoint equations, through application of semigroup theory. For a detailed introduction, the reader is referred to [EN01, Paz12]. For both equations, the solution concept of mild solutions will be adopted.

Definition A.1.1 (Mild and classical solutions; [EN01])**.** Consider the inhomogeneous abstract Cauchy problem with initial data $x \in \mathcal{X}$ in a Banach space $\mathcal{X}$ and inhomogeneity $h \in L^1([0,T];\mathcal{X})$

$$\begin{aligned} \dot{u}(t) &= \mathcal{A}u(t) + h(t) \qquad \text{for } t \in [0,T], \\ u(0) &= x, \end{aligned} \tag{iACP}$$

where the operator $\mathcal{A} : \mathcal{D}(\mathcal{A}) \to \mathcal{X}$ generates a strongly continuous semigroup $\mathsf{S}(t) : \mathcal{X} \to \mathcal{X}$ on $\mathcal{X}$ and is defined on $\mathcal{D}(\mathcal{A}) := \{x \in \mathcal{X} \mid \mathcal{A}x \in \mathcal{X}\}$. Then the function $u \in C^0([0,T];\mathcal{X})$ defined by

$$u(t) := \mathsf{S}(t)x + \int_0^t \mathsf{S}(t-s)h(s)\,\mathrm{d}s \tag{A.1.1}$$

is called the (unique) mild solution to (iACP). If, additionally, $x \in \mathcal{D}(\mathcal{A})$ and $h \in W^{1,1}([0,T];\mathcal{X})$, then $u \in C^1([0,T];\mathcal{X}) \cap C^0([0,T];\mathcal{D}(\mathcal{A}))$ is called a classical solution to (iACP).

Note that for classical solutions u, their derivatives appearing in the equation (iACP) exist in $\mathcal{X}$, whereas they might not be well-defined for mild solutions. The concept is transferable to final value problems with negative in time propagation.

A.1.1. Forward Problem Existence

Consider the kinetic initial value problem

$$\begin{aligned} \partial_t f + v \cdot \nabla_x f &= \alpha \mathcal{L}(f) - \sigma f + h, && \text{for } (t,x,v) \in [0,T] \times \mathbb{R}^d \times V, \\ f(t=0,x,v) &= \phi(x,v), && \text{for } (x,v) \in \mathbb{R}^d \times V, \end{aligned} \tag{A.1.2}$$

where $\alpha \in \{0,1\}$ and $\mathcal{L}$ and σ are defined as in (3.1) for $K \in \mathcal{A}_K$ as in (3.3). Let a source term h and an initial condition ϕ be given.

In this section, the proof of existence of a solution to (A.1.2) is presented and it is shown that the solution preserves positivity of initial data and source term. These results are well-established in literature and can be found e.g. in [CMPS04, Maj97]. Their derivation by standard arguments is described in the following for sake of completeness.

A.1.1.1. Existence

The proof of existence is based on semigroup theory. Two decompositions of the full differential operator $\mathcal{A} = -v \cdot \nabla - \sigma + \alpha\mathcal{L}$ describing (A.1.2) will be used, defined by the operators

$$\begin{aligned} \tilde{\mathcal{T}} : u \mapsto -v \cdot \nabla_x u - \sigma u \qquad &\text{and} \qquad \mathcal{L} : u \mapsto \mathcal{L}(u), \qquad &&\text{and} \\ \mathcal{T} : u \mapsto -v \cdot \nabla_x u \qquad &\text{and} \qquad \mathcal{K}_\alpha : u \mapsto -\sigma u + \alpha\mathcal{L}(u). \end{aligned} \tag{A.1.3}$$

The operators $\mathcal{T}, \tilde{\mathcal{T}} : \mathcal{D}(\mathcal{T}) \to \mathcal{X}$ account for transport and transport with decay at rate σ, respectively. In contrast to that, the operators $\mathcal{L}, \mathcal{K}_\alpha : \mathcal{X} \to \mathcal{X}$ account for the gain due to tumbling into the given velocity, and the overall effect of tumbling, respectively. The first decomposition $\mathcal{A} = \tilde{\mathcal{T}} + \alpha\mathcal{L}$ provides advantages for bounding the solution f, whereas the second one $\mathcal{A} = \mathcal{T} + \mathcal{K}_\alpha$ is beneficial when the influence of the tumbling kernel K shall be separated, as in Chapter 5.

Lemma A.1.2. *Fix $1 \le p < \infty$ and consider $\mathcal{X} := L^p(\mathbb{R}^d \times V)$. The operators $(\mathcal{T}, \mathcal{D}(\mathcal{T}))$, $(\tilde{\mathcal{T}}, \mathcal{D}(\mathcal{T}))$ generate strongly continuous semigroups $(\mathsf{T}(t))_{t\ge0}, (\tilde{\mathsf{T}}(t))_{t\ge0}$ on $\mathcal{X}$ defined by*

$$\mathsf{T}(t) : u(x,v) \mapsto u(x - vt, v), \qquad \text{and} \qquad \tilde{\mathsf{T}}(t) : u(x,v) \mapsto e^{-\int_0^t \sigma(x-vs,v)\,\mathrm{d}s} u(x-vt,v) \tag{A.1.4}$$

that are bounded $\|\mathsf{T}(t)\|, \|\tilde{\mathsf{T}}(t)\| \le 1$ in operator norm. The operator $(\mathcal{A} = \tilde{\mathcal{T}} + \alpha\mathcal{L}, \mathcal{D}(\mathcal{T}))$ generates a strongly continuous semigroup $(\mathsf{S}_\alpha(t))_{t\ge0}$ on $\mathcal{X}$ that is bounded by $\|\mathsf{S}_\alpha(t)\| \le e^{\alpha|V|C_K t}$.

Proof. The semigroup properties of $\mathsf{T}(t), \tilde{\mathsf{T}}(t)$ and their boundedness by 1 are obvious. Strong continuity translates from the dense subset $C_c^1(\mathbb{R}^d; L^p(V))$ to $\mathcal{X}$. Furthermore, they are generated by $\mathcal{T}, \tilde{\mathcal{T}}$ by the difference quotient form of weak directional derivatives. Linearity of the operator $\mathcal{L}$ and its boundedness

$$\|\mathcal{L}(u)\|_{\mathcal{X}} = \left\| \int_V K(x,v,v') u(x,v')\,\mathrm{d}v' \right\|_{\mathcal{X}} \le C_K |V| \|u\|_{\mathcal{X}} \qquad \Rightarrow \qquad \|\mathcal{L}\| \le C_K |V|$$

by Jensen's inequality then allow application of the bounded perturbation theorem [EN01, Th. III.1.3] to see that $\tilde{\mathcal{T}} + \alpha\mathcal{L}$ generates a strongly continuous semigroup $(\mathsf{S}_\alpha(t))_{t\ge0}$ with $\|\mathsf{S}_\alpha(t)\| \le e^{\alpha\|\mathcal{L}\|t}$. □

Existence and boundedness of a mild solution to (A.1.2) in $C^0([0,T];L^p(\mathbb{R}^d\times V))$ immediately follows by construction [EN01, Cor. IV.7.6]

$$f(t)=\mathsf{S}_\alpha(t)\phi+\int_0^t \mathsf{S}_\alpha(t-s)h(s)\,\mathrm{d}s. \tag{A.1.5}$$

In fact, this solution lives in $C^0([0,T];L^1\cap L^\infty(\mathbb{R}^d\times V))$ under sufficiently regular data.

Corollary A.1.3 (Forward Existence)**.** *Let $T>0$ and initial data $\phi\in\mathcal{Y}:=L^1\cap L^\infty(\mathbb{R}^d\times V)$ and a source term $h\in L^1([0,T];\mathcal{Y})$ be given. Then initial value problem (A.1.2) attains a unique solution $f\in C^0([0,T];\mathcal{Y})$ in the sense of (A.1.5), that is bounded by*

$$\|f(t)\|_{\mathcal{Y}}\le e^{\alpha|V|C_K t}\left(\|\phi\|_{\mathcal{Y}}+\int_0^t\|h(s)\|_{\mathcal{Y}}\,\mathrm{d}s\right). \tag{A.1.6}$$

If additional regularity of initial data $\phi\in\mathcal{D}_{\mathcal{Y}}(\mathcal{T}):=\{u\in\mathcal{Y}\mid v\cdot\nabla u\in\mathcal{Y}$ for a.a. $x\in\mathbb{R}^d\}$ and source term $h\in W^{1,1}([0,T];\mathcal{Y})$ is given, then f is the classical solution of (A.1.2) and attains additional regularity $C^1([0,T];\mathcal{Y})\cap C^0([0,T];\mathcal{D}_{\mathcal{Y}}(\mathcal{T}))$.

Proof. Hölder's inequality gives $\|u\|_p\le\|u\|_{\mathcal{Y}}$ and thus $\mathcal{Y}\subset L^p(\mathbb{R}^d\times V)$ for all $1\le p<\infty$. From the previous considerations it is clear that the mild $L^p(\mathbb{R}^d\times V)$ solution f defined in (A.1.5) satisfies the uniform bound

$$\|f(t)\|_{L^p(\mathbb{R}^d\times V)}\le e^{2|V|C_K t}\|\phi\|_{\mathcal{Y}}+\int_0^t e^{2|V|C_K(t-s)}\|h(s)\|_{\mathcal{Y}}\,\mathrm{d}s.$$

For $p=1$ and as $p\to\infty$, this provides (A.1.6). Existence of a classical solution follows similarly, by extending L^p bounds to L^∞. □

Because the semigroup $(\mathsf{S}_\alpha(t))$ has no explicit form, it will be convenient in various contexts to work with implicit forms of f in terms of $\mathsf{T}(t)$ and $\tilde{\mathsf{T}}(t)$. They originate as fixed points of the source iteration maps

$$\Phi,\tilde\Phi: C^0([0,T];\mathcal{Y})\to C^0([0,T];\mathcal{Y}),\quad \Phi(u)=\hat f,\qquad\text{and}\qquad \tilde\Phi(u)=\tilde f$$

that maps u, respectively, to the mild solutions of

$$\begin{cases}\partial_t\hat f=\mathcal{T}\hat f+\mathcal{K}_\alpha(u)+h,\\ \hat f(t=0)=\phi,\end{cases}\qquad\text{and}\qquad\begin{cases}\partial_t\tilde f=\tilde{\mathcal{T}}\tilde f+\alpha\mathcal{L}(u)+h,\\ \tilde f(t=0)=\phi.\end{cases}$$

Lemma A.1.4 (Implicit forms of f, source iteration)**.** *In the setting of Corollary A.1.3, $f\in C^0([0,T];\mathcal{Y})$ is the mild solution to (A.1.2), if and only if it satisfies one of the two equivalent implicit forms*

$$f(t)=\tilde{\mathsf{T}}(t)\phi+\int_0^t\tilde{\mathsf{T}}(t-s)(h(s)+\alpha\mathcal{L}(f)(s))\,\mathrm{d}s,\qquad\textit{or} \tag{A.1.7}$$

$$f(t)=\mathsf{T}(t)\phi+\int_0^t\mathsf{T}(t-s)(h(s)+\mathcal{K}_\alpha(f)(s))\,\mathrm{d}s. \tag{A.1.8}$$

Moreover, f is the limit of any source iteration through Φ and $\tilde\Phi$ as defined above, i.e. $f=\lim_{n\to\infty}\Phi^n(u)=\lim_{n\to\infty}\tilde\Phi^n(u)$ for any $u\in\mathcal{Y}$.

Proof. $\Rightarrow$: This representation follows directly from semigroup theory and the fact that the strongly continuous semigroup $(\mathsf{S}_\alpha(t))_{t\in[0,T]}$ of the sum $\mathcal{A} = \tilde{\mathcal{T}} + \alpha\mathcal{L}$ of the generator $\tilde{\mathcal{T}}$ of another strongly continuous semigroup $(\tilde{\mathsf{T}}(t))_{t\in[0,T]}$ and a bounded linear operator $\alpha\mathcal{L}$ can be expressed as

$$\mathsf{S}_\alpha(t)u = \tilde{\mathsf{T}}(t)u + \int_0^t \tilde{\mathsf{T}}(t-s)\alpha\mathcal{L}(\mathsf{S}_\alpha(s)u)\,\mathrm{d}s$$

for $u \in \mathcal{Y}$ [EN01, Cor. III.1.7]. Inserting this into (A.1.5) and using the linearity of $\tilde{\mathsf{T}}$ and $\mathcal{L}$ and time independence of $\alpha\mathcal{L}$ to retrieve f in the argument of the operator $\mathcal{L}$ yields the representations (A.1.7). Formula (A.1.8) follows analogously by boundedness, linearity and time independence of the operator $\mathcal{K}_\alpha$.

$\Leftarrow$: Obviously f is a fixed point of $\Phi, \tilde{\Phi}$ by (A.1.7)–(A.1.8) and thus satisfies (iACP) in a mild sense.

By proving that $\tilde{\Phi}$ is a contraction in a suitable norm, Banach's fixed point theorem then guarantees that the fixed point is unique and every source iteration sequence $(\tilde{\Phi}^n(u))_n$ converges to f. Spell out

$$\tilde{\Phi} : u \mapsto \tilde{\mathsf{T}}(t)\phi + \int_0^t \tilde{\mathsf{T}}(t-s)(h(s) + \alpha\mathcal{L}(u)(s))\,\mathrm{d}s,$$

and define - in analogy to [Maj97] - the weighted norm $\|u\|_\lambda := \|e^{-\lambda t}u\|_{C^0([0,T];\mathcal{Y})}$ on $C^0([0,T];\mathcal{Y})$, and set $\lambda := 2\|\alpha\mathcal{L}\| = 2\alpha C_K|V|$. By norm equivalence, the completeness of $C^0([0,T];\mathcal{Y})$ w.r.t. $\|\cdot\|_\lambda$ is inherited from $\|\cdot\|_{C^0([0,T];\mathcal{Y})}$. Furthermore boundedness of $\|\tilde{\mathsf{T}}(t)\| \leq 1$ and $\|\alpha\mathcal{L}\| \leq \frac{\lambda}{2}$ on $\mathcal{Y}$ follows from uniform in p boundedness of both operators on $L^p(\mathbb{R}^d \times V)$ and provide the contraction property of $\tilde{\Phi}$, as

$$\begin{aligned}
\|\tilde{\Phi}(u) - \tilde{\Phi}(u')\|_\lambda &= \left\| \int_0^t \tilde{\mathsf{T}}(t-s)\alpha\mathcal{L}(u-u')(s)\,\mathrm{d}s \right\|_\lambda \\
&\leq \sup_t e^{-\lambda t} \int_0^t \alpha C_K|V| e^{\lambda s} e^{-\lambda s} \|(u-u')(s)\|_{\mathcal{Y}}\,\mathrm{d}s \\
&\leq \frac{1}{2} \sup_t \int_0^t \lambda e^{\lambda(s-t)}\,\mathrm{d}s\ \|u-u'\|_\lambda \\
&\leq \frac{1}{2}\|u-u'\|_\lambda.
\end{aligned}$$

The contraction property for Φ is shown analogously.

□

A.1.1.2. Properties of f

Besides boundedness in (A.1.6), f also preserves non-negativity and compactness of the spatial support of initial data ϕ and source term h. Intuitively, non-negativity of f makes sense by its interpretation as the particle density and the fact that particles are only transported, absorbed with a certain rate σ, or produced by a positive source h and may ($\alpha = 1$) or may not ($\alpha = 0$) reappear after a change of velocity. Because the evolution of f is governed by a kinetic equation, the speed of propagation in space is finite - to be

precise, it is 1 by choice of $V = \mathbb{S}^{d-1}$. This means, that a compactly in x supported initial condition evolves to a compactly in x supported density f, given that the source term h does not introduce any non-compactly supported contributions.

Lemma A.1.5. *Consider the setting of Corollary A.1.3.*

a) Consider non-negative initial data $\phi(x,v), h(t,x,v) \geq 0$ for all $t \in [0,t], x \in \mathbb{R}^d, v \in V$. Then the solution f to (A.1.2) is non-negative $f(t,x,v) \geq 0$ a.e.w. in $\mathbb{R}^d \times V$, for all $t \in [0,T]$.

b) Equation (A.1.2) has constant speed of propagation 1.

In particular, let h be zero and consider ϕ that is compactly supported in x a.e. in V, in the sense that there exists a compact subset S of $\mathbb{R}^d$ such that for a.a. $v \in V$ one has $\phi(x,v) = 0$ for all $x \in S^C$. Then one has for a.a. v that $f(t,x,v) = 0$ for all $t \in [0,T]$ and a.a. x with Euclidean distance $d(x,S) > t$, i.e. $f(t)$ has a compact essential support in x, that is contained in the closed ball $\bar{B}(S,t)$ around S with radius t, for every $t \in [0,T]$.

The proof follows the arguments of [Maj97] by source iteration.

Proof. By Lemma A.1.4, f is the pointwise a.e. limit of the source iteration $(f^{(n)})_n$ with $f^{(n+1)} = \tilde{\Phi}(f^{(n)})$ and starting value $f^{(0)} = 0$.

a) Because $\tilde{\mathsf{T}}(t)$ sustains non-negativity of ϕ and h, one has

$$f^{(1)}(t) - f^{(0)}(t) = f^{(1)}(t) = \tilde{\mathsf{T}}(t)\phi + \int_0^t \tilde{\mathsf{T}}(t-s)h(s)\,\mathrm{d}s \geq 0 \qquad \text{a.e.w. in } \mathbb{R}^d \times V.$$

Because $\mathcal{L}(g) \geq 0$ if $g \geq 0$ a.e.w. in $\mathbb{R}^d \times V$, one iteratively obtains

$$(f^{(n+1)} - f^{(n)})(t,x,v) = \alpha \int_0^t \tilde{\mathsf{T}}(t-s)\mathcal{L}(f^{(n)} - f^{(n-1)})(s)\,\mathrm{d}s \geq 0$$

which provides $f^{(N)} = \sum_{n=0}^{N}(f^{(n+1)} - f^{(n)}) \geq 0$. This translates to the $f = \lim_N f^{(N)}$ by point wise convergence a.e.w. in x, v.

b) An induction shows that $f^{(n)}(t,x,v) = 0$ for a.e. $v \in V$ and all x with distance $d\big(x,S\big) > t$. The assertion is clear for $f^{(0)} \equiv 0$ and follows for all $n \in \mathbb{N}$ from the recursive definition of

$$\begin{aligned} f^{(n+1)}(t,x,v) &= \tilde{\Phi}(f^{(n)})(t,x,v) \\ &= e^{-\int_0^t \sigma(x-vs,v)\,\mathrm{d}s}\phi(x-vt,v) + \int_0^t e^{-\int_0^{t-s}\sigma(x-v\tau),v)}\alpha\mathcal{L}(f^{(n)})(s,x-v(t-s),v)\,\mathrm{d}s. \end{aligned}$$

For $x \in \mathbb{R}^d$ with $d\big(x,S\big) > t$, then $d(x-vt,S) > 0$ and thus $\phi(x-vt,v) = 0$ for a.a. v. Due to $d\big(x-v(t-s),S\big) > s$, the induction hypothesis yields for a.a. v' that $f^{(n)}(s,x-v(t-s),v') = 0$, which shows $\mathcal{L}(f^{(n)})(s,x-v(t-s),v) = 0$ for all $s \leq t$. In summary, this gives for a.e. v that $f^{(n+1)}(t,x,v) = 0$ for these x, which concludes the induction. The assertion follows again by point-wise a.e.w. in x, v convergence of $f^{(n)}$ to f.

□

A.1.1.3. Regularity in K

Regularity of the solutions of PDEs to their coefficients is frequently encountered and sometimes even included in the definition of well-posedness. In this section, it shall be verified for the solution f_K to (Ch)–(iCh), in order to prove Lemma 3.4 on twice continuous differentiability. In fact, one can show that f_K is smooth in K by further iteration of the arguments in the proof. For the purposes of this work, however, C^2 is sufficient.

The desired regularity is established iteratively, starting with Lipschitz continuity of f.

Lemma A.1.6. *The solution f to (Ch)–(iCh) is Lipschitz continuous w.r.t. K*

$$\|f_{\tilde{K}} - f_K\|_{C([0,T];L^1\cap L^\infty(\mathbb{R}^d\times V))} \leq C\|\tilde{K} - K\|_\infty \qquad \text{for} \qquad \tilde{K}, K \in \mathcal{A}_K \tag{A.1.9}$$

for some $C = C(T, C_K, \|\phi\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)})$ independent of $K, \tilde{K}$.

Proof. Consider the difference $\bar{f} := f_{\tilde{K}} - f_K$ of two solutions to (Ch)–(iCh) corresponding to two distinct tumbling kernel values $\tilde{K}, K \in \mathcal{A}_K$. This is a solution to the difference of the PDEs and thus satisfies

$$\partial_t \bar{f} + v \cdot \nabla_x \bar{f} = \mathcal{K}_K(\bar{f}) + \mathcal{K}_{\tilde{K}-K}(f_{\tilde{K}}) \tag{A.1.10}$$

with vanishing initial condition. By regularity of $f_{\tilde{K}}$ in Proposition 3.3, the source $\mathcal{K}_{\tilde{K}-K}(f_{\tilde{K}})$ belongs to $L^1(0,T;L^1\cap L^\infty(\mathbb{R}^d\times V))$, which allows application of the solution bound in Corollary A.1.3 and boundedness of $f_{\tilde{K}}$ in (A.1.6) yields

$$\begin{aligned}\|\bar{f}(t)\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)} &\leq e^{C_K|V|t}\int_0^t \|\mathcal{K}_{\tilde{K}-K}(f_{\tilde{K}})(s)\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)}\,\mathrm{d}s \\ &\leq 2\|K_\star - K\|_\infty|V|\ t\ e^{2C_K|V|t}\|\phi\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)}.\end{aligned}$$

□

Directional derivatives can be obtained as limits of difference quotients and their Lipschitz continuity provides continuous differentiability. In the following, an admissible variation for a given $K \in \mathcal{A}_K$ describes an $\eta \in L^\infty(\mathbb{R}^d \times V \times V)$ such that $K + \varepsilon\eta \in \mathcal{A}_K$ for small enough $\varepsilon \neq 0$.

Lemma A.1.7. *The solution f to (Ch)–(iCh) is continuously differentiable w.r.t. K. The directional derivatives $\partial_\eta f_K \in C([0,T];L^1\cap L^\infty(\mathbb{R}^d\times V))$ for admissible variations $\eta \in L^\infty(\mathbb{R}^d\times V\times V)$ are solutions to the kinetic equations*

$$\begin{aligned}\partial_t\partial_\eta f + v\cdot\nabla_x\partial_\eta f &= \mathcal{K}_K(\partial_\eta f) + \mathcal{K}_\eta(f_K), \\ \partial_\eta f(t=0) &= 0,\end{aligned} \tag{A.1.11}$$

they are Lipschitz continuous in K and bounded by

$$\|\partial_\eta f_K(t)\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)} \leq 2|V|\|\eta\|_\infty e^{2|V|C_K t}\|\phi\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)}. \tag{A.1.12}$$

Proof. Let $\eta \in L^\infty(\mathbb{R}^d \times V \times V)$ be an admissible variation for the considered $K \in \mathcal{A}_K$. Then by (A.1.10), the difference quotient $\delta_\eta^\varepsilon f := \varepsilon^{-1}(f_{K+\varepsilon\eta} - f_K)$ satisfies

$$\partial_t \delta_\eta^\varepsilon f + v \cdot \nabla_x \delta_\eta^\varepsilon f = \mathcal{K}_K(\delta_\eta^\varepsilon f) + \mathcal{K}_\eta(f_{K+\varepsilon\eta})$$

with vanishing initial condition. As $\varepsilon \to 0$, the directional derivative $\partial_\eta f$ of f in direction η then satisfies (A.1.11). It is thus well defined in $C^0([0,T]; L^1 \cap L^\infty(\mathbb{R}^d \times V))$ and satisfies the estimate (A.1.6) by regularity of the source $\mathcal{K}_\eta(f_K) \in L^1([0,T]; L^1 \cap L^\infty(\mathbb{R}^d \times V))$ and Corollary A.1.3. Inserting the estimate (3.4) for f_K then yields formula (A.1.12). Taking the limit in the equation is a standard procedure. For sake of completeness, it is rigorously justified once in the following: Define the difference $\bar{f} := \delta_\eta^\varepsilon f - \partial_\eta f$, then its propagation is determined by the difference equation

$$\partial_t \bar{f} + v \cdot \nabla_x \bar{f} = \mathcal{K}_K(\bar{f}) + \mathcal{K}_\eta(f_{K+\varepsilon\eta} - f_K).$$

Then Corollary A.1.3 shows that

$$\begin{aligned}
\|\bar{f}(t)\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)} &\le e^{|V|C_K t} \int_0^t \|\mathcal{K}_\eta(f_{K+\varepsilon\eta} - f_K)(s)\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)}\,\mathrm{d}s \\
&\le e^{|V|C_K t} \int_0^t 2|V|\|\eta\|_\infty \|(f_{K+\varepsilon\eta} - f_K)(s)\|_{L^1\cap L^\infty(\mathbb{R}^d\times V)}\,\mathrm{d}s \\
&\le \varepsilon 2|V|\|\eta\|_\infty^2\; t\; e^{|V|C_K t} C,
\end{aligned}$$

where the Lipschitz continuity of f in K with Lipschitz constant C from (A.1.9) was applied in the last row. As $\varepsilon \to 0$, this shows that $\delta_\eta^\varepsilon f \to \partial_\eta f$, i.e. f is directionally differentiable in K.

Lipschitz continuity of the directional derivative is established in analogy to the proof of Lemma A.1.6. □

Lemma A.1.8. *The solution f to (Ch)–(iCh) is twice continuously differentiable w.r.t. K and the second order directional derivatives are Lipschitz continuous in K.*

Proof. Let $\eta, \zeta \in L^\infty(\mathbb{R}^d \times V \times V)$ and $K \in \mathcal{A}_K$ such that $K + \varepsilon(\eta + \zeta) \in \mathcal{A}_K$ for small enough $\varepsilon > 0$. Similarly as in the proof of Lemma A.1.7, one obtains from consideration of the difference quotient $\delta_\zeta^\varepsilon \partial_\eta f = \varepsilon^{-1}(\partial_\eta f_{K+\varepsilon\zeta} + \partial_\eta f_K)$, that the second order directional derivative $\partial^2_{\zeta,\eta} f \in C^0([0,T]; L^1 \cap L^\infty(\mathbb{R}^d \times V))$ is the unique mild solution to the equation

$$\partial_t \partial^2_{\zeta,\eta} f + v \cdot \nabla_x \partial^2_{\zeta,\eta} f = \mathcal{K}_K(\partial^2_{\zeta,\eta} f) + \mathcal{K}_\zeta(\partial_\eta f) + \mathcal{K}_\eta(\partial_\zeta f).$$

Lipschitz continuity w.r.t. K follows analogously by Lipschitz continuity of $\partial_\eta f$ and $\partial_\zeta f$. □

A.1.2. Adjoint Problem Existence

For the adjoint problem, consider the final value problem

$$\begin{aligned}
-\partial_t g - v \cdot \nabla_x g &= -\sigma g + \alpha \mathcal{L}^\star(g) + \mu, && \text{for } (t,x,v) \in [0,T] \times \mathbb{R}^d \times V, \\
g(t = T, x, v) &= \psi(x,v), && \text{for } (x,v) \in \mathbb{R}^d \times V,
\end{aligned} \tag{A.1.13}$$

where $\alpha \in \{0,1\}$, $\mathcal{L}^\star(g) := \int_V K(x,v',v)g(t,x,v')\,\mathrm{d}v'$ for some $K \in \mathcal{A}_K$ and $\mu \in L^1([0,T];\mathcal{X})$ and $\psi \in \mathcal{X}$ with Banach space $\mathcal{X} := L^1(\mathbb{R}^d; L^1 \cap L^\infty(V))$.

Semigroup theory provides existence of mild solutions, in analogy to the results in Subsection A.1.1.1.

Lemma A.1.9. *The final value problem (A.1.13) with final data $\psi \in \mathcal{X} := L^1(\mathbb{R}^d; L^1 \cap L^\infty(V))$ and source term $\mu \in L^1([0,T];\mathcal{X})$ attains a unique mild solution $g \in C^0([0,T];\mathcal{X})$ that is bounded by*

$$\|g(t)\|_{\mathcal{X}} \leq e^{\alpha C_K |V|(T-t)} \left(\|\psi\|_{\mathcal{X}} + \int_t^T \|\mu(s)\|_{\mathcal{X}}\,\mathrm{d}s \right). \tag{A.1.14}$$

Furthermore, g coincides with the unique solutions in $C^0([0,T];\mathcal{X})$ of the implicit equations

$$g(t) = \mathsf{T}^\star(T-t)\psi + \int_t^T \mathsf{T}^\star(s-t)(\mu(s) + \mathcal{K}_\alpha^\star(g(s)))\,\mathrm{d}s, \tag{A.1.15}$$

$$g(t) = \tilde{\mathsf{T}}^\star(T-t)\psi + \int_t^T \tilde{\mathsf{T}}^\star(s-t)(\mu(s) + \alpha\mathcal{L}^\star(g(s)))\,\mathrm{d}s, \tag{A.1.16}$$

with adjoint tumbling operator $\mathcal{K}_\alpha^\star := -\sigma + \alpha\mathcal{L}^\star$ and the semigroups $(\mathsf{T}^\star(t))_t$ and $(\tilde{\mathsf{T}}^\star(t))_t$ on $\mathcal{X}$ corresponding to backward transport and backward transport with decay given by

$$\mathsf{T}^\star(t) : u(x,v) \mapsto u(x+vt,v), \quad \text{and} \quad \tilde{\mathsf{T}}^\star(t) : u(x,v) \mapsto e^{-\int_0^t \sigma(x+v\tau,v)\,\mathrm{d}\tau} u(x+vt,v).$$

Proof. Through a transformation of variables $t \mapsto T-t, x \mapsto -x$, the final value problem (A.1.13) can be transformed into an initial value problem of the form of (A.1.2)

$$\begin{aligned} \partial_t \vec{g} + v \cdot \nabla \vec{g} &= -\vec{\sigma}\vec{g} + \alpha\vec{\mathcal{L}}^\star(\vec{g}) + \vec{\mu} \\ \vec{g}(t=0) &= \vec{\psi}, \end{aligned} \tag{A.1.17}$$

where the notation $\vec{h}(t,x,v) := h(T-t,-x,v)$ is used for quantities in the forward transport equation.[1] It then follows in analogy to Lemma A.1.2 and Corollary A.1.3, that the full operator $\vec{\mathcal{T}} + \alpha\vec{\mathcal{L}}^\star$ generates a strongly continuous semigroup on $L^1(\mathbb{R}^d; L^p(V))$ that is bounded in operator norm by $e^{\alpha\|\vec{\mathcal{L}}^\star\|t} = e^{\alpha C_K |V| t}$ for all p and thus (A.1.17) attains a unique mild solution $\vec{g} \in C([0,T]; L^1(\mathbb{R}^d; L^p(V)))$ that is uniformly in p bounded by

$$\|\vec{g}(t)\|_{L^1(\mathbb{R}^d;L^p(V)))} \leq e^{\alpha C_K |V| t} \left(\|\vec{\psi}\|_{\mathcal{X}} + \int_0^t \|\vec{\mu}(s)\|_{\mathcal{X}}\,\mathrm{d}s \right).$$

An analogous argumentation as in Lemma A.1.4 shows that $\vec{g} \in C([0,T];\mathcal{X})$ and it solves

$$\vec{g}(t) = \mathsf{T}(t)\vec{\psi} + \int_0^t \mathsf{T}(t-s)(\vec{\mu}(s) + \vec{\mathcal{K}}_\alpha^\star(\vec{g})(s))\,\mathrm{d}s, \quad \text{and}$$

$$\vec{g}(t) = \vec{\tilde{\mathsf{T}}}(t)\vec{\psi} + \int_0^t \vec{\tilde{\mathsf{T}}}(t-s)(\vec{\mu}(s) + \vec{\alpha}\mathcal{L}^\star(\vec{g})(s))\,\mathrm{d}s,$$

[1] In particular, it does not mean that the considered quantity is vector valued, as in other literature.

where $(\mathsf{T}(t))_{t>0}$ and $(\vec{\mathsf{T}}(t))_{t>0}$ are given by (A.1.4) and denote the strongly continuous semigroups on $\mathcal{X}$ that are generated by the transport and transport with decay operators $\mathcal{T}$ and $\vec{\mathcal{T}} := \mathcal{T} - \vec{\sigma}$, respectively. Vice versa, any solution to either of these equation in $C^0([0,T];\mathcal{X})$ coincides with the mild solution to (A.1.17) and is thus unique. Transforming back to original variables yields (A.1.15) and the bound (A.1.14). □

By transforming (A.1.13) into (A.1.17), it is clear that g shares certain properties with f, in particular the constant speed of propagation 1. Furthermore, g vanishes for times bigger than the upper bound of the temporal support of μ, if the final condition ψ vanishes.

Corollary A.1.10. *Let g be the solution of (A.1.13) with $\psi = 0$ and $\mu \in L^1([0,T];L^1(\mathbb{R}^d;\mathcal{X}))$ with $\mu(t) = 0$ for $t \in [b,T]$ for some $0 < b < T$. Then $g(t) = 0$ for $t > b$ and*

$$\|g(t)\|_{\mathcal{X}} \leq e^{\alpha C_K |V|(b-t)} \|\mu\|_{L^1([t,b];\mathcal{X})}.$$

Proof. By (A.1.14), it follows that $g(t)$ vanishes for $t > b$ and the bound on $g(t)$ for $t < b$ follows by setting $T = b$. □

A.2. Existence Results for the Stationary Schrödinger Equation via Elliptic Theory

In this section, existence of solutions to the forward and adjoint Schrödinger equation (8.9) and (8.13) is established through standard elliptic theory [Bre11, Eva22]. Some attention is required in the applications of standard results, given that the domain $X = (-1,1)^2$ is not C^1. In the following, consider the stationary PDE

$$-\Delta u + pu = f \quad \text{in } X = (-1,1)^2 \qquad \text{and} \qquad u = 0 \quad \text{on } \partial X, \tag{A.2.1}$$

with non negative potential $p \in C^\infty_+(X)$ and a source term f.

A.2.1. L^2 Data

If $f \in L^2(X)$, the weak formulation of (A.2.1) can be derived by testing with a test function $\phi \in H^1_0(X)$ and application of integration by parts, assuming sufficient regularity of the solution u:

$$\langle f, \phi \rangle_{L^2(X)} = \langle -\Delta u + pu, \phi \rangle_{L^2(X)} = \langle \nabla u, \nabla \phi \rangle_{L^2(X)} + \langle u, p\phi \rangle_{L^2(X)}. \tag{A.2.2}$$

Note that the boundary terms vanish due to $\phi = 0$ on ∂X. This formulation allows for weaker regularity of solutions. Setting the solution space to H^1_0, boundary conditions can be taken into account.

Definition A.2.1. Let $f \in L^2(X)$. A function $u \in H^1_0(X)$ is called weak solution to (A.2.1), if it satisfies (A.2.2) for all test functions $\phi \in H^1_0(X)$.

Existence of weak solutions to the forward Schrödinger equation (8.9) follows from standard elliptic theory [Bre11, Eva22]. Higher regularity of the source and the parameter then translate to the solution, turning weak solutions into a classical one, that satisfied (A.2.1) point wise.

Proposition A.2.2. *Let $p \in C^\infty_+(X)$ be non negative and $f \in C^\infty(X)$, then (A.2.1) attains a unique solution $u \in C^\infty(X)$. Moreover, if $f \geq 0$ is positive somewhere inside X, then $u > 0$ in X.*

A short proof, as can be found in [Bre11], is included for sake of completeness.

Proof. Define the bilinear form $B : H^1_0(X) \times H^1_0(X) \to \mathbb{R}$ according to the RHS of (A.2.2) by

$$B[u, \phi] = \langle \nabla u, \nabla \phi \rangle_{L^2(X)} + \langle u, p\phi \rangle_{L^2(X)} .$$

It is continuous $|B[u, \phi]| \leq \max(1, \|p\|_\infty) \|u\|_{H^1_0} \|\phi\|_{H^1_0}$ and coercive due to non negativity of p and Poincaré's inequality [Bre11, Cor.9.19] $\|u\|^2_{L^2} \leq C \|\nabla u\|^2_{L^2}$ for some $C > 0$, as

$$B[u, u] = \|\nabla u\|^2_{L^2} + \|\sqrt{p} u\|^2_{L_2} \geq \|\nabla u\|^2_{L^2} \geq \frac{1}{2} \left(\|\nabla u\|^2_{L^2} + \frac{1}{C} \|u\|^2_{L^2} \right) \geq \beta \|u\|_{H^1_0}$$

for $\beta = \min(1/2, 1/(2C)) > 0$. The Lax Milgram theorem [Bre11, Cor.5.8] hence provides existence of a unique solution $u \in H^1_0(X)$ to the weak formulation $B[u, \phi] = \langle f, \phi \rangle_{L^2(X)}$ of (A.2.1) in (A.2.2). Higher regularity $u \in C^\infty(X)$ then follows from Theorem 6.3.3 in [Eva22], and positivity for positive f from the strong maximum principle, compare [Eva22, Thm. 6.4.4]. □

A.2.2. Dirac Data

Regarding the adjoint Schrödinger equation (8.13), the singularity in the source term requires a new solution concept. Following [Pon12], we may introduce the following notion of solutions to the Schrödinger equation with measure data, in the spirit of the Definition A.2.1 of weak solutions:

Definition A.2.3. Denote by $\mathcal{M}(X)$ the set of finite Borel measures on X. A function $u \in L^1(X)$ will be called a solution to (A.2.1) with measure valued data $f \in \mathcal{M}(X)$, if it satisfies

$$-\int_X u \Delta \psi \, \mathrm{d}x + \int_X pu\psi \, \mathrm{d}x = \int_X \psi \, \mathrm{d}f(x) \tag{A.2.3}$$

for all continuous test functions $\psi \in C_c \cap W^{2,\infty}(\bar{X})$ that vanish on the boundary ∂X.

An adaptation of the proof of Proposition 3.2 in [Pon12] to the Schrödinger equation, involving the additional potential term, shows existence of solutions to (8.13).

Proposition A.2.4. *For every $\hat{x} \in X$, there exists a unique solution $g_{\hat{x}} \in L^1(X)$ to (8.13) in the sense of Definition A.2.3.*

Proof. Fix $\hat{x} \in X$. Through mollification, there exists a sequence $(f_n)_n \subset C_c^\infty(\bar{X})$ that converges weakly in the sense of measures to $\delta_{\hat{x}} \in \mathcal{M}(X)$, i.e.

$$\lim_{n\to\infty} \int_X \psi f_n \,\mathrm{d}x = \int_X \psi \,\mathrm{d}\delta_{\hat{x}}(x) \qquad \text{for all } \psi \in C_c(\bar{X}),$$

with $\lim_{n\to\infty} \|f_n\|_{L^1(X)} = \|\delta_{\hat{x}}\|_{\mathcal{M}(X)} = |\delta_{\hat{x}}|(X) = 1$, and thus $\|f_n\|_{L^1(X)}$ is bounded. Denoting by g_n the corresponding solution to (A.2.1) with source term $f_n \in C_c^\infty(\bar{X})$ then in analogy to Corollary 4.3 in [Pon12] one can show boundedness of g_n in $W_0^{1,q}(X)$ by $\|f_n\|_{L^1(X)}$ for some $q \in (1,2)$. The compact embedding of this Sobolev space in $L^1(X)$ by the Rellich Kondrachov theorem [Ada75, Thm.6.2] allows us to extract a weakly convergent subsequence $(g_{n_k})_k$, whose limit shall be denoted by $g \in L^1(X)$. Together with weak convergence of $f_n \rightharpoonup \delta_{\hat{x}}$ in the sense of measures, this ensures that the limit is a solution to (8.13) in the sense of Definition A.2.3: For all $\psi \in C_c \cap W^{2,\infty}(\bar{X})$ one has

$$\begin{aligned} -\int_X g\Delta\psi \,\mathrm{d}x + \int_X pg\psi \,\mathrm{d}x &= \lim_{k\to\infty} \left(-\int_X g_{n_k}\Delta\psi \,\mathrm{d}x + \int_X pg_{n_k}\psi \,\mathrm{d}x \right) \\ &= \lim_{k\to\infty} \left(\int_X f_{n_k}\psi \,\mathrm{d}x \right) \qquad\qquad = \int_X \psi \,\mathrm{d}\delta_{\hat{x}}(x). \end{aligned}$$

Uniqueness of g, and thus $g = g_{\hat{x}}$, then follows from uniqueness of the zero solution to (A.2.1) with RHS $f \equiv 0 \in C^\infty(X)$ according to Proposition A.2.2. □

B

Transformations on the Sphere

This section collects several transformations and their characteristics that were used during Chapter 4. In Appendix B.1, the stereographic projection is explained in more detail and its Jacobi determinant is derived. Then, in Appendix B.2, the behaviour of transformation $\mathcal{T}_a^w$ as applied in the proof of Lemma 4.8 is investigated.

B.1. Stereographic projection

The stereographic projection $\mathbb{P}_N$ is a well established tool from differential geometry [Spi75, Cox69] that projects the sphere $\mathbb{S}^{d-1}\backslash\{N\}$ in $\mathbb{R}^d$, with exception of the north pole $N = (0, ..., 1)^T$, onto $\mathbb{R}^{d-1}$ bijectively for $d = 2, 3$. Every point $v = (v_1, ..., v_d) \in \mathbb{S}^{d-1}\backslash\{N\}$ is projected onto the unique intersection of the line through N and v, with the plane at $v_d = 0$, as displayed in Figure B.1.1 in 2D.

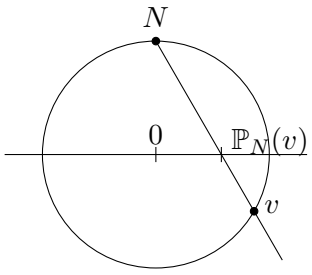


Figure B.1.1.: Stereographic projection in 2D.

An explicit formula for $\mathbb{P}_N$ in Cartesian coordinates reads

$$\mathbb{P}_N(v) = (\tfrac{v_1}{1-v_d}, ..., \tfrac{v_{d-1}}{1-v_d})^T.$$

The stereographic projection can also be defined for another point $w \in \mathbb{S}^{d-1}$ than the north pole by setting $\mathbb{P}_w(v) = \mathbb{P}_N(R_w v)$ for $v \in \mathbb{S}^{d-1}\backslash\{w\}$, where R_w is a fixed rotation that maps w onto N.

In Section 4.2, the stereographic projection is used to define closeness of points on the sphere. A change of variables via $\mathbb{P}_w$ requires its bijectivity and its absolute Jacobi determinant.

Lemma B.1.1 (Stereographic Projection)**.**

a) The stereographic projection $\mathbb{P}_w : \mathbb{S}^{d-1}\backslash\{w\} \to \mathbb{R}^{d-1}$ *for a fixed* $w \in \mathbb{S}^{d-1}$ *is smooth and bijective. Its image has norm* $\|\mathbb{P}_w(v)\|^2 = \frac{1+\langle v,w\rangle}{1-\langle v,w\rangle}$ *and its absolute Jacobi determinant is given by*

$$j_{\mathbb{P}_w}(v) = \frac{1}{(1-\langle v,w\rangle)^{d-1}}.$$

b) Its inverse $\mathbb{P}_w^{-1} : \mathbb{R}^{d-1} \to \mathbb{S}^{d-1}\backslash\{w\}$ *is smooth and reads* $\mathbb{P}_w^{-1}(y) = R_w^{-1} \cdot ((\mathbb{P}_N^{-1}(y))_i)_{i=1}^d$*, with*

$$(\mathbb{P}_N^{-1}(y))_i = \begin{cases} \frac{2y_i}{1+y_1^2+\ldots+y_{d-1}^2}, & i = 1,\ldots,d-1, \\ \frac{-1+y_1^2+\ldots+y_{d-1}^2}{1+y_1^2+\ldots+y_{d-1}^2}, & i = d \end{cases}.$$

For $\lambda \in \mathbb{R}$ *it satisfies* $\mathbb{P}_w^{-1}(\lambda y) = -w + \lambda v^\lambda(y)$ *for some* $v^\lambda(y) \in \mathbb{R}^d$ *that is bounded independently of* λ *by* $\|v^\lambda(y)\| \le 2\|y\|$*.*

The proof of this lemma is an easy calculation exercise. It is displayed for the sake of completeness.

Proof. Bijectivity follows from the existence of a well-defined inverse. Smoothness of $\mathbb{P}_w$ and its inverse is a result of smoothness of the entries, because the denominator cannot vanish, as

$$(Rv)_d = \langle Rv, N\rangle = \langle Rv, Rw\rangle = \langle v, w\rangle \neq 1 \tag{B.1.1}$$

for $v \in \mathbb{S}^{d-1}\backslash\{w\}$.

a) Using $R_w v \in \mathbb{S}^{d-1}$ and thus $\sum_{i=1}^d (R_w v)_i^2 = 1$, as R is a rotation matrix and thus an isometry, the formula for the norm of the image follows from (B.1.1)

$$\|\mathbb{P}_w(v)\|^2 = \sum_{i=1}^{d-1} \frac{(R_w v)_i^2}{(1-(R_w v)_d)^2} = \frac{1-(R_w v)_d^2}{(1-(R_w v)_d)^2} = \frac{1+\langle v,w\rangle}{1-\langle v,w\rangle}$$

To compute the absolute Jacobi determinant, we use polar or spherical coordinates.

- In 2D, a change to polar coordinates gives $\mathbb{P}_N(\phi) = \frac{\cos(\phi)}{1-\sin(\phi)}$ and therefore the polar coordinate Jacobi determinant

$$\partial_\phi \mathbb{P}_N(\phi) = \frac{-\sin(\phi)(1-\sin(\phi)) + \cos^2(\phi)}{(1-\sin(\phi))^2} = \frac{1}{1-\sin(\phi)}$$

and with the relation

$$j_{\mathbb{P}_N}(v)\,\mathrm{d}v = j_{\mathbb{P}_N}(v(\phi))\,\mathrm{d}\phi = \frac{1}{1-v_2(\phi)}\,\mathrm{d}\phi,$$

one obtains $j_{\mathbb{P}_N}(v) = \frac{1}{1-v_2}$.

- In 3D, the relation

$$j_{\mathbb{P}_N}(v)\,\mathrm{d}v = j_{\mathbb{P}_N}(v(\phi,\theta))\sin(\theta)\,\mathrm{d}(\phi,\theta) = |\det D_{\phi,\theta}[\mathbb{P}_N(v(\phi,\theta)]|\,\mathrm{d}(\phi,\theta)$$

describes the change of variables to spherical coordinates $v(\phi,\theta) = (\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta))^T$. With

$$\begin{aligned}\det D_{\phi,\theta}[\mathbb{P}_N(v(\phi,\theta)] &= \det D_{\phi,\theta}\left[\left(\tfrac{\sin(\theta)\cos(\phi)}{1-\cos(\theta)}, \tfrac{\sin(\theta)\sin(\phi)}{1-\cos(\theta)}\right)^T\right] \\ &= \det\begin{pmatrix} \frac{-\sin(\theta)\sin(\phi)}{1-\cos(\theta)} & \frac{-\cos(\phi)}{1-\cos(\theta)} \\ \frac{\cos(\phi)\sin(\theta)}{1-\cos(\theta)} & \frac{-\sin(\phi)}{1-\cos(\theta)}\end{pmatrix} = \frac{\sin(\theta)}{(1-\cos(\theta))^2} = \frac{\sin(\theta)}{(1-v_3(\phi,\theta))^2}\end{aligned}$$

one obtains $j_{\mathbb{P}_N}(v) = \frac{1}{(1-v_3)^2}$.

In summary, $j_{\mathbb{P}_N}(v) = \frac{1}{(1-v_d)^{d-1}}$ for $d \in \{2,3\}$. For the stereographic projection from a general $w \in \mathbb{S}^{d-1}$ one then has

$$\begin{aligned}j^{(d)}_{\mathbb{P}_w}(v) &= |\det(D[\mathbb{P}_N](R_w v)R_w)| = |\det(D[\mathbb{P}_N](R_w v))\det(R_w)| = j_{\mathbb{P}_N}(R_w v) \\ &= \tfrac{1}{(1-(R_w v)_d)^{d-1}},\end{aligned}$$

given that the determinant of a rotation matrix is ± 1. The given formula arises when inserting $(Rv)_d = \langle v, w\rangle$.

b) Using the explicit formula for $\mathbb{P}^{-1}_{-v_{\text{in}}}$, one sees that

$$\begin{aligned}\mathbb{P}^{-1}_w(\lambda y) &= R_w^{-1}\left(\left(\tfrac{2\lambda y_i}{1+\lambda^2\sum_{i=1}^{d-1} y_i^2}\right)_{i=1,\dots,d-1}, \tfrac{-1+\lambda^2\sum_{i=1}^{d-1} y_i^2}{1+\lambda^2\sum_{i=1}^{d-1} y_i^2}\right)^T \\ &= R_w^{-1}\left[-N + \left(\left(\tfrac{2\lambda y_i}{1+\lambda^2\sum_{i=1}^{d-1} y_i^2}\right)_{i=1,\dots,d-1}, \tfrac{2\lambda^2\sum_{i=1}^{d-1} y_i^2}{1+\lambda^2\sum_{i=1}^{d-1} y_i^2}\right)^T\right] = -w + \lambda v^\lambda(y),\end{aligned}$$

by linearity of the inverse R_w^{-1} of the rotation matrix R_w that maps $w \mapsto N$ and thus $R_w^{-1}(N) = w$. Moreover,

$$v^\lambda(y) := R_w^{-1}\left(\left(\tfrac{2y_i}{1+\lambda^2\sum_{i=1}^{d-1} y_i^2}\right)_{i=1,\dots,d-1}, \tfrac{2\lambda\sum_{i=1}^{d-1} y_i^2}{1+\lambda^2\sum_{i=1}^{d-1} y_i^2}\right)^T$$

is bounded by $2\|y\|$ independently of λ, since R_w^{-1} is an isometry and thus

$$\|v^\lambda(y)\|^2 = \sum_{i=1}^{d-1} \tfrac{4y_i^2}{(1+\lambda^2\sum_{i=1}^{d-1} y_i^2)^2} + \tfrac{4\lambda^2(\sum_{i=1}^{d-1} y_i^2)^2}{(1+\lambda^2\sum_{i=1}^{d-1} y_i^2)^2} = \tfrac{4\sum_{i=1}^{d-1} y_i^2}{1+\lambda^2\sum_{i=1}^{d-1} y_i^2} \le 4\|y\|^2.$$

□

B.2. On the transformation $\mathcal{T}_a^w$

In the $d \in \{2,3\}$-dimensional setting, let a constant $t' > 0$ as well as a $w \in \mathbb{S}^{d-1}$ and $a \in \mathbb{R}^d$ be given and consider the map $\mathcal{T}_a^w : [0,t'] \times \mathbb{S}^{d-1} \to \mathbb{R}^d, (s,v) \mapsto a - s(v-w)$. It first transforms the points by $\mathcal{S}^w : (s,v) \mapsto s(v-w)$ and afterwards applies a point reflection at the origin and a shift by a. Obviously, all points (s,w) as well as all points $(0,v)$ are mapped to a. Excluding these points from the domain, however, one obtains an injective map that maps inner points onto inner points. The behaviour of $\mathcal{T}_a^w$ is summarized by the following lemma, and can be understood when considering Figure B.2.1 that focuses on the first transformation $\mathcal{S}^w$.

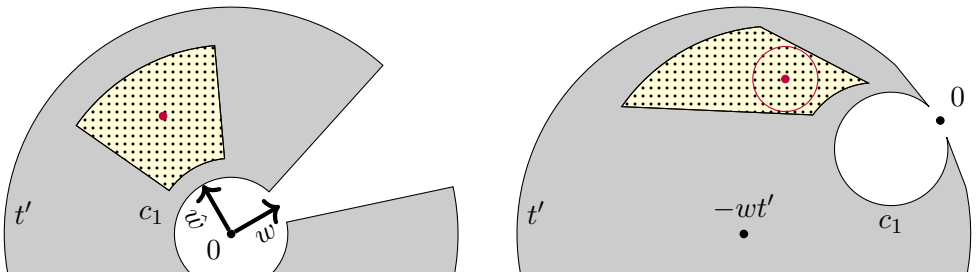
Lemma B.2.1. *Fix $t' > 0$ and $w \in \mathbb{S}^{d-1}$ and $a \in \mathbb{R}^d$.*

a) For all $c_1 \in (0,t')$ and $c_2 \in (0,1)$, the map $\mathcal{T}_a^w : U \to \mathcal{S}^w(U), (s,v) \mapsto a - s(v-w)$ defined on $U := [c_1,t'] \times \{v \in \mathbb{S}^{d-1} \mid \langle v,w\rangle \leq 1-c_2\}$ is bijective with continuous inverse

$$(\mathcal{T}_a^w)^{-1}(z) = (\zeta,\omega)(z) = \left(\frac{\|a-z\|^2}{2|\langle a-z,w\rangle|}, w + \frac{a-z}{\zeta(z)} \right), \quad z \in \mathcal{T}_a^w(U). \tag{B.2.1}$$

and absolute Jacobi determinant $j_{\mathcal{T}_a^w}(s,v) = s^{d-1}(1-\langle v,w\rangle)$.

b) Furthermore, let $(\hat{s} = \lambda t', \hat{w}) \in (0,t') \times \mathbb{S}^{d-1}\backslash\{w\}$ be arbitrary with $1 > \lambda > 0$ and let $a = \hat{s}(\hat{w} - w)$. If $c_1 = \hat{c}_1 t'$ for a $\hat{c}_1 < \lambda$ and $c_2 < \lambda(1 - \langle \hat{w}, w\rangle)$, then there exists a $\mu = \mu(w,\hat{w},\lambda,\hat{c}_1,c_2)$ that does not depend on t' such that the ball $B(0,\mu t')$ around the origin with radius of order t' is contained in $\mathcal{T}_a^w(U)$.



(a) Sliced annulus A. (b) Perturbation of U by $\mathcal{S}^w$.

Figure B.2.1.: Transformation $\mathcal{S}^w$ in 2D:
(a) For any point $\hat{s}\hat{w}$ (red) in the interior of the sliced annulus $A = \{sv \mid (s,v) \in U\}$ (gray), a neighbourhood in A in form of a slice of the annulus (yellow, dotted) can be constructed as the area between the arches of two circles.
(b) In the image of U by $\mathcal{S}^w$ (gray), each point in $vs \in A$ is translated by $-ws$. The image of the yellow dotted slice of A (again yellow dotted) is bounded by the same arches of the circles, whose centres are shifted in direction $-w$ by the amount of the radius such that they touch 0 at angle w. It contains the image $\mathcal{S}^w(\hat{s},\hat{w})$ (red) as an interior point, as well as a ball around $\hat{a}$ whose boundary is depicted by the red circle. The radius of the circle can be chosen as a fixed fraction $\mu t'$.

Proof. a) The inverse (B.2.1) is well defined, since for every $z = a - s(v - w) \in \mathcal{T}_a^w(U)$, the denominators are positive due to

$$|\langle a - z, w\rangle| = s(1 - \langle v, w\rangle) \geq c_1(1 - (1 - c_2)) = c_1 c_2 > 0, \quad \text{and}$$
$$\|a - z\|^2 = s^2(2 - 2\langle v, w\rangle) \geq 2c_1^2 c_2 > 0.$$

This also assures continuity of ζ, ω. Bijectivity thus follows from the existence of an inverse. In order to compute the absolute Jacobi determinant $j_{\mathcal{T}_a^w}$, in 3D a change to spherical coordinates ϕ, θ with $v(\phi, \theta) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)^T$ yields

$$j^{(3)}_{\mathcal{T}_a^w}(s, v)\,\mathrm{d}s\,\mathrm{d}v = j^{(3)}_{\mathcal{T}_a^w}(s, v(\phi,\theta))\sin(\theta)\,\mathrm{d}s\,\mathrm{d}(\phi,\theta) = |\det D_{s,\phi,\theta}[\mathcal{T}_a^w(s, v(\phi,\theta))]|\,\mathrm{d}s\,\mathrm{d}(\phi,\theta),$$

where the Jacobi matrix of $\mathcal{T}_a^w(s, v(\phi,\theta))$ w.r.t. s, ϕ, θ is given by

$$D_{s,\phi,\theta}[\mathcal{T}_a^w(s, v(\phi,\theta))] = \begin{pmatrix} w_1 - \sin\theta\cos\phi & s\sin\theta\sin\phi & -s\cos\theta\cos\phi \\ w_2 - \sin\theta\sin\phi & -s\sin\theta\cos\phi & -s\cos\theta\sin\phi \\ w_3 - \cos\theta & 0 & s\sin\theta \end{pmatrix}$$

and has determinant

$$\det D_{s,\phi,\theta}[\mathcal{T}_a^w(s, v(\phi,\theta))] = -\sin\theta s^2\langle v(\phi,\theta) - w, v(\phi,\theta)\rangle = -\sin\theta s^2(1 - \langle w, v(\phi,\theta)\rangle).$$

In 2D, polar coordinates $v(\phi) = (\cos\phi, \sin\phi)$ similarly give

$$\begin{aligned} j^{(2)}_{\mathcal{T}_a^w}(s, v)\,\mathrm{d}s\,\mathrm{d}v = j^{(2)}_{\mathcal{T}_a^w}(s, v(\phi))\,\mathrm{d}s\,\mathrm{d}\phi &= |\det D_{s,\phi}[\mathcal{T}_a^w(s, v(\phi))]|\,\mathrm{d}s\,\mathrm{d}\phi \\ &= \left|\det \begin{pmatrix} w_1 - \cos\phi & s\sin\phi \\ w_2 - \sin\phi & -s\cos\phi \end{pmatrix}\right| \mathrm{d}s\,\mathrm{d}\phi = s\langle v(\phi) - w, v(\phi)\rangle\,\mathrm{d}s\,\mathrm{d}\phi \\ &= s(1 - \langle w, v(\phi)\rangle)\,\mathrm{d}s\,\mathrm{d}\phi. \end{aligned}$$

b) To prove the last assertion, let $\hat{s} = \lambda t'$, $\hat{w}$, a, c_1 and c_2 be as in the Lemma. Consider an element $z \in B(0, \mu t')$ with $\mu := \min(\delta\|a/t'\|, \delta|\langle a/t', w\rangle|, \lambda(1 - \langle\hat{w}, w\rangle) - c_2)$ for some $1 > \delta > 0$ such that

$$\lambda\frac{(1-\delta)^2}{1+\delta} \geq \hat{c}_1, \qquad \lambda\frac{(1+\delta)^2}{1-\delta} \leq 1.$$

Such δ exists due to the monotone convergence of $\frac{(1-\delta)^2}{1+\delta}, \frac{(1+\delta)^2}{1-\delta}$ to 1 and the restriction on $\hat{c}_1$. The definition of a shows that a/t' is in fact independent of t' and therefore, so is μ. Our aim is now to show that the preimage of z by ζ, ω is contained in U. Using $\frac{\|a\|^2}{2|\langle a,w\rangle|} = \lambda t'$ and the (reverse) triangle inequality as well as the Cauchy-Schwartz inequality, one has

$$\begin{aligned} \zeta(z) &= \frac{\|a - z\|^2}{2|\langle a - z, w\rangle|} \geq \frac{(\|a\| - \|z\|)^2}{2(|\langle a, w\rangle| + |\langle z, w\rangle|)} \geq \frac{(\|a\| - \mu t')^2}{2(|\langle a, w\rangle| + \mu t')} \\ &\geq \frac{\|a\|^2}{2|\langle a, w\rangle|}\frac{(1-\delta)^2}{1+\delta} = \lambda t'\frac{(1-\delta)^2}{1+\delta} \geq c_1, \\ \zeta(z) &\leq \frac{(\|a\| + \|z\|)^2}{2(|\langle a, w\rangle| - |\langle z, w\rangle|)} \leq \frac{(\|a\| + \mu t')^2}{2(|\langle a, w\rangle| - \mu t')} \leq \frac{\|a\|^2}{2|\langle a, w\rangle|}\frac{(1+\delta)^2}{1-\delta} \leq t'. \end{aligned}$$

Moreover, by definition of ζ, ω in (B.2.1), it holds that

$$\|\omega(z)\|^2 = \left\| w + 2|\langle w, \tfrac{a-z}{\|a-z\|}\rangle| \tfrac{a-z}{\|a-z\|} \right\|^2 = 1 + 4|\langle w, \tfrac{a-z}{\|a-z\|}\rangle|^2 \left(\operatorname{sgn}(\langle w, \tfrac{a-z}{\|a-z\|}\rangle) + 1\right) = 1,$$

where $\operatorname{sgn}(x)$ denotes the sign of a quantity x. It is negative for the considered scalar product, since

$$\langle a - z, w\rangle = \langle a, w\rangle - \langle z, w\rangle \leq t' \hat{\lambda} \langle \hat{w} - w, w\rangle + \mu t' = t'(\mu - \hat{\lambda}(1 - \langle \hat{w}, w\rangle)) \leq -t' c_2 < 0$$

by the Cauchy-Schwartz inequality and the definition of μ. Finally, the above as well as the bound $\zeta(z) \leq t'$ show

$$\langle \omega(z), w\rangle = \left\langle w + \tfrac{a-z}{\zeta(z)}, w \right\rangle = 1 + \tfrac{1}{\zeta(z)} \langle w, a - z\rangle \leq 1 + \frac{-c_2 t'}{t'} = 1 - c_2.$$

This shows that $(\zeta(z), \omega(z)) \in U$.

□

C

Numerical Schemes for PDE Discretization

This chapter introduces numerical schemes to discretize the PDEs that we encountered during the main part. For the kinetic equations, a simple finite difference scheme will be laid out in Appendix C.1, whereas Appendix C.2 displays the finite element schemes that were used for the elliptic stationary Schrödinger equation.

C.1. Finite Difference Schemes for Transport Equations

This section starts with a brief overview over the central concepts of accuracy, stability and convergence for finite difference schemes. After motivating the derivation of the Lax Wendroff scheme, the schemes for discretizing the forward and adjoint kinetic chemotaxis equations (Ch) and (6.2) are derived. For more detailed introductions to numerics of PDEs, the reader is referred to literature, for instance [LL92, Str04].

C.1.1. Review on Properties of Numerical Schemes.

Fundamental properties of numerical schemes are their capability to approximate the original PDE, computability in the sense that errors do not propagate in an uncontrolled fashion and make the solution explode, as well as the convergence of their solutions to the PDE solution. In the following, denote the PDE by

$$Pu = 0 \text{ on } (0,T] \times X, \tag{C.1.1}$$

for some spatial domain $X \subset \mathbb{R}^d$, and let it be complemented by suitable initial and boundary conditions. Finite difference schemes rely on a pointewise discretization of the domain $[0,T] \times X$.

Space-time Discretization. After grid sizes $\kappa > 0, \tau > 0$ for spatial and temporal domain have been fixed, the computational domain $[0,T] \times X$ of a space and time dependent PDE

(C.1.1) can be discretized to $\mathtt{T} \times \mathtt{X}$, where $\mathtt{X} := \{X \ni \mathtt{x}_m = \kappa \sum_{i=1}^{d} m_i e_i \mid m = (m_i)_i \in \mathbb{Z}^d\}$ denotes a set of equidistant points in space, described by the linear combinations of unit vectors $e_i \in \mathbb{R}^d$ with entries $(e_i)_j = \mathbb{1}_i(j)$, and the temporal equidistant grid $\mathtt{T} := \{\mathtt{t}^n = n\tau \mid n \in \{0, 1, ..., \lfloor \frac{T}{\tau} \rfloor = N\}\}$.

One Step Schemes. In the following, denote by $\mathtt{P}_{\tau,\kappa}$ the numerical one-step update scheme $\mathtt{u}(\mathtt{t}^{n+1}) = \mathtt{P}_{\kappa,\tau}\mathtt{u}(\mathtt{t}^n)$ defined on the space-time grid $\mathtt{T} \times \mathtt{X}$ with time step size $\tau > 0$ and spatial step size $\kappa > 0$, corresponding to the PDE (C.1.1) with solution u.

Accuracy. The *order of accuracy* of a numerical scheme measures how well it approximates the temporal update through the evolution equation, in relation to the time step size τ.

Definition C.1.1 ([LL92, Str04])**.** Assume that u is smooth and define the local truncation error as $E_{\tau,\kappa}(\mathtt{t}, \mathtt{x}) := \frac{u(\mathtt{t}+\tau,\mathtt{x}) - \mathtt{P}_{\tau,\kappa}u(\mathtt{t},\mathtt{x})}{\tau}$. Then the numerical scheme is called consistent, if $\|E_{\kappa,\tau}(\mathtt{t})\| \to 0$ for all $\mathtt{t} \in \mathtt{T}$ in a suitable norm (typically the discrete $L^1(\mathtt{X})$ or $L^2(\mathtt{X})$ norm) as $\tau, \kappa \to 0$, and accurate of order p in time and q in space, if for any smooth solution $u(t, x)$, corresponding to a compactly supported initial condition u_0, one has $\|E_{\kappa,\tau}(\mathtt{t})\| = \mathcal{O}(\tau^p) + \mathcal{O}(\kappa^q)$. Accurate schemes with orders $p, q \geq 1$ are consistent.

Stability. Numerical stability ensures that errors, introduces for instance by rounding values to computer precision, do not accumulate over time in an uncontrolled fashion which might eventually dominate the true behaviour of the solution. Through bounding the operator norm of $\mathtt{P}^n_{\tau,\kappa}$, one can ensured that errors, that have been introduced in the initial condition and propagated by the scheme up to time $\mathtt{t}^n$, stay bounded and the computation does not diverge.

Definition C.1.2 ([LL92, Str04])**.** A numerical time update scheme $\mathtt{P}_{\tau,\kappa}$ is stable in a stability region Λ, if for every time $T > 0$, there exists a constant C_T such $\|\mathtt{P}^n_{\tau,\kappa}\| \leq C_T$ for all $n\tau \leq T$ with $\tau, \kappa \in \Lambda$, where $\|\cdot\|$ again refers to a suitable operator norm.

Convergence. Ideally, the numerical solution should converge to the true solution point wise, as the grid size becomes arbitrarily small.

Definition C.1.3 ([Str04])**.** A numerical scheme $\mathtt{u}(\mathtt{t}^{n+1}, \mathtt{x}) = \mathtt{P}_{\tau,\kappa}\mathtt{u}(\mathtt{t}^n, \mathtt{x})$ for the PDE (C.1.1) is convergent, if the numerical solution $\mathtt{u}(\mathtt{t}^n, \mathtt{x}_m) \to u(t, x)$ for $(\mathtt{t}^n, \mathtt{x}_m) \to (t, x)$ and $\mathtt{u}_0(\mathtt{x}_m) \to u_0(x)$ as $\tau, \kappa \to 0$.

For finite difference schemes that are linear in the previous time solution, the above properties are linked.

Theorem C.1.4 (Lax-Richtmyer Equivalence Theorem [LR05])**.** *A consistent and linear finite difference scheme corresponding to a well-posed initial value problem is convergent if and only if it is stable.*

C.1.2. The Lax Wendroff Scheme

The Lax Wendroff scheme [LW64] is a second order finite difference scheme for the advection equation in $\mathbb{R}^d$ with constant velocity v

$$\partial_t u + v \cdot \nabla_x u = 0 \qquad \text{with initial data} \qquad u(t=0) = u_0. \tag{C.1.2}$$

It updates point values of the solution on the space grid, based on the Taylor approximation

$$u(t+\tau, x) = u(t,x) + \tau \partial_t u(t,x) + \frac{\tau^2}{2} \partial_{tt} u(t,x) + \mathcal{O}(\tau^3) \tag{C.1.3}$$

for sufficiently regular u. Using (C.1.2), the time derivatives can be exchanged for spatial derivatives

$$\partial_t u(t,x) = -v \cdot \nabla_x u = -\sum_{i=1}^d v_i \partial_{x_i} u, \quad \text{and} \quad \partial_{tt} u(t,x) = (-v \cdot \nabla_x)^2 u = \sum_{i,j=1}^d v_i v_j \partial_{x_i x_j} u.$$

The spatial derivatives can then be approximated by centered finite differences according to the Taylor expansion

$$\begin{aligned}
\partial_{x_i} u(t,x) &= \tfrac{u(t,x+\kappa e_i) - u(t,x-\kappa e_i)}{2\kappa} + \mathcal{O}(\kappa^2) =: \mathtt{d}^\kappa_{x_i} u(t,x) + \mathcal{O}(\kappa^2), \\
\partial_{x_i x_i} u(t,x) &= \tfrac{u(t,x+\kappa e_i) - 2u(t,x) + u(t,x-\kappa e_i)}{\kappa^2} + \mathcal{O}(\kappa^2) =: \mathtt{d}^\kappa_{x_i x_i} u(t,x) + \mathcal{O}(\kappa^2), \quad \text{for } i \in \{1,\dots,d\}, \\
\partial_{x_i x_j} u(t,x) &= \tfrac{u(t,x+\kappa(e_i+e_j)) - u(t,x+\kappa(e_i-e_j)) - u(t,x+\kappa(e_j-e_i)) + u(t,x-\kappa(e_i+e_j))}{(2\kappa)^2} + \mathcal{O}(\kappa^2) \\
&=: \mathtt{d}^\kappa_{x_i x_j} u(t,x) + \mathcal{O}(\kappa^2), \qquad \text{for } i \neq j.
\end{aligned}$$

If one considers points $x = \mathtt{x} \in \mathtt{X}$ and $t = \mathtt{t}^n \in \mathtt{T}$ on the grid and neglects higher order terms, then the computations involve only values of u on neighbouring grid points at the previous time $\mathtt{t}^n$ to update to $u(\mathtt{t}^{n+1}, \mathtt{x})$. Inserting the approximations into (C.1.3) then yields the Lax Wendroff time update formula for the point values on the grid

$$\mathtt{u}(\mathtt{t}^{n+1}, \mathtt{x}) = \mathtt{u}(\mathtt{t}^n, \mathtt{x}) - \tau \sum_{i=1}^d v_i \mathtt{d}^\kappa_{x_i} \mathtt{u}(\mathtt{t}^n, \mathtt{x}) + \frac{\tau^2}{2} \sum_{i,j=1}^d v_i v_j \mathtt{d}^\kappa_{x_i x_j} \mathtt{u}(\mathtt{t}^n, \mathtt{x}) =: \mathtt{P}^{\mathrm{LW}}_{\kappa,\tau;v} \mathtt{u}(\mathtt{t}^n, \mathtt{x}), \tag{C.1.4}$$

where $\mathtt{u}$ denotes the numerical solution and $\mathtt{P}^{\mathrm{LW}}_{\kappa,\tau;v} : \mathtt{u}(\mathtt{t}^n) \mapsto \mathtt{u}(\mathtt{t}^{n+1})$ the one step finite difference update operator with velocity v. The scheme can be used to propagate an initial condition $\mathtt{u}_0(\mathtt{x}) = u_0(\mathtt{x})$ through the time grid to obtain the solution on every space time grid point in $\mathtt{T} \times \mathtt{X}$.

Properties of the Lax Wendroff Scheme. The above derivations show that the local truncaiton error is or order $\tau^{-1}(u(\mathtt{t}+\tau) - \mathtt{P}^{\mathrm{LW}}_{\tau,\kappa;v} u(\mathtt{t})) = \mathcal{O}(\kappa^2) + \mathcal{O}(\tau^2)$ in $L^2(\mathtt{X})$, because the order of approximation of the spatial derivatives by finite differences holds uniformly in x for sufficiently smooth and compactly supported u by Taylor's theorem. An obvious extension of the original proof by Lax and Wendroff [LW64] in two space dimensions provides stability of scheme (C.1.4) in general space dimension d and the Lax-Richtmyer theorem provides convergence.

Lemma C.1.5 ([LW64]). *The Lax Wendroff scheme* (C.1.4) *for the advection equation* (C.1.2) *is consistent, second order accurate in space and time, and stable in the discrete* $L^2(\mathtt{X})$ *norm if for* $\frac{\tau}{\kappa}\max_i |v_i| \leq d^{-3/2}$ *in which case* $\|\mathrm{P}^{\mathrm{LW}}_{\tau,\kappa;v}\| \leq 1$. *Fixing the ratio* $\tau = \lambda\kappa$ *that satisfies the stability condition, the scheme is convergent for* $\kappa \to 0$.

C.1.3. Finite Difference Schemes for the Forward and Adjoint Chemotaxis Equations

In order to apply a Lax Wendroff type scheme to numerically compute solutions to the forward and adjoint equations, a discretized velocity space is considered that turns the kinetic PDEs into a system of coupled transport equations. Note that no discretization is necessary in spatial dimension $d = 1$, with velocity space $V = \{\pm 1\}$, where the forward and adjoint models (Ch), (6.2) can be regarded as a two-species model.

In higher dimensions $d = 2$ the choice of the admissible set $\mathcal{A}_K^{\mathrm{pwc}}$ in (5.28) proposes the use of the discrete ordinate method (S_N) to discretize the velocity space $V = \mathbb{S}^1$.

Discrete Ordinates Velocity Discretization for $d = 2$. The discrete ordinate method is based on a quadrature rule of the velocity integral, and the solution is only computed at the quadrature velocities, and only those values of the tumbling kernel at the quadrature velocities enter the further computation.

The method thus nicely aligns with the piecewise constant in velocity tumbling kernels prescribed by $\mathcal{A}_K^{\mathrm{pwc}}$: by choosing a simple quadrature rule based on an equidistant step function representation of the integral in polar coordinates, the quadrature points can be set to $\mathtt{V} = \{\mathtt{v}_j = (\cos(2\pi j(N_v), \sin(2\pi j/N_v)) \mid j = 0, ..., N_v - 1\}$, the middle points of the segments S_j defined by the admissible set $\mathcal{A}_K^{\mathrm{pwc}}$, and the coefficient values $K_r^{j,i}$ correspond to the values of K at the quadrature velocities $\mathtt{v}_j, \mathtt{v}_i$. Weights are uniformly set to $|V|/N_v$.

The method then exchanges every integral in v by the quadrature rule, and computes the solution only in the quadrature points on $\mathtt{V}$. This transforms (Ch), and analogously (6.2), into a system of transport equations for the densities corresponding to the quadrature velocities $\mathtt{v}$ that are coupled through the discretized tumbling operator on the right hand side, and can be regarded as a multi-species model:

$$\begin{aligned}\partial_t f(t,x,\mathtt{v}) + \mathtt{v}\cdot\nabla_x f(t,x,\mathtt{v}) =& \frac{|V|}{N_v}\sum_{\mathtt{v}'\in\mathtt{V}} K(x,\mathtt{v},\mathtt{v}')f(t,x,\mathtt{v}') - K(x,\mathtt{v}',\mathtt{v})f(t,x,\mathtt{v}) \qquad \text{(C.1.5)}\\ =:& \mathcal{K}(f(t,x)), \qquad \text{and}\end{aligned}$$

$$\begin{aligned}-\partial_t g(t,x,\mathtt{v}) - \mathtt{v}\cdot\nabla_x g(t,x,\mathtt{v}) =& \frac{|V|}{N_v}\sum_{\mathtt{v}'\in\mathtt{V}} K(x,\mathtt{v}',\mathtt{v})(g(t,x,\mathtt{v}') - g(t,x,\mathtt{v})) \qquad \text{(C.1.6)}\\ &- \frac{1}{L}\sum_{l=0}^{L}(\mathtt{M}_l(f) - y_l)\mu_l(t,x),\\ =:& \mathcal{K}^\star(g(t,x)) - \frac{1}{L}\sum_{l=0}^{L}(\mathtt{M}_l(f) - y_l)\mu_l(t,x),\end{aligned}$$

where $\mathcal{K}, \mathcal{K}^\star$ and $\mathtt{M}_l(f) := \int_0^T \int_{\mathbb{R}^d} \frac{|V|}{N_v} \sum_{\mathtt{v}' \in \mathtt{V}} f(t,x,\mathtt{v}')\mu_l(t,x)\,\mathrm{d}x\,\mathrm{d}t$ describe the discrete velocity forward and adjoint tumbling operators and measurement operator, respectively. Initial and final data $f(t=0,x,\mathtt{v}) = \phi(x,\mathtt{v})$ and $g(t=T,x,\mathtt{v}) = 0$ are defined point wise in velocity.

Remark C.1.6. An alternative approach to discretize the velocity space is given by the spherical harmonics (P_N) method that works with an expansion of the function into finitely many basis functions on the sphere. This non local discretization seems less suited for the purpose of this work, given that a spherical harmonics expansion form of the locally discretized tumbling kernel $K \in \mathcal{A}_K^{\mathrm{pwc}}$ would have to be constructed for the computation, introducing an additional level of error. For more details on the P_N and S_N methods, the reader is referred to [Cha50, LM84, Pom73].

Finite Difference Schemes for the Forward and Adjoint Equation. The Lax Wendroff scheme can be extended to deal with the systems for velocity discretized forward and adjoint equations (C.1.5) and (C.1.6). We propose a scheme that uses a Lax Wendroff discretization for the transport part and, for simplicity, treats the discrete tumbling operators and source terms explicitly:

$$\mathtt{f}(\mathtt{t}^{n+1},\mathtt{x},\mathtt{v}) = \mathtt{P}^{\mathrm{LW}}_{\tau,\kappa;\mathtt{v}}\mathtt{f}(\mathtt{t}^n,\mathtt{x},\mathtt{v}) + \tau\mathcal{K}(\mathtt{f})(\mathtt{t}^n,\mathtt{x},\mathtt{v}) =: \mathtt{P}^{\mathrm{chem}}_{\tau,\kappa}\mathtt{f}(\mathtt{t}^n,\mathtt{x},\mathtt{v}), \tag{C.1.7}$$

$$\mathtt{g}(\mathtt{t}^{n-1},\mathtt{x},\mathtt{v}) = \mathtt{P}^{\mathrm{LW}}_{\tau,\kappa;-\mathtt{v}}\mathtt{g}(\mathtt{t}^n,\mathtt{x},\mathtt{v}) + \tau\mathcal{K}^\star(\mathtt{g})(\mathtt{t}^n,\mathtt{x},\mathtt{v}) + \tau\mathtt{m}(\mathtt{t}^n,\mathtt{x},\mathtt{v}) =: \mathtt{P}^{\mathrm{adj}}_{\tau,\kappa}\mathtt{g}(\mathtt{t}^n,\mathtt{x},\mathtt{v}), \tag{C.1.8}$$

for all $\mathtt{x} \in \mathtt{X}$, $\mathtt{v} \in \mathtt{V}$ and $n < N$ in the forward scheme (C.1.7) and $n > 0$ in the backward adjoint scheme (C.1.8). The discretized source term reads $\mathtt{m}(\mathtt{t}^n,\mathtt{x},\mathtt{v}) = \frac{1}{L}\sum_{l=1} L(\mathtt{M}_l(\mathtt{f}) - y_l)\mu(\mathtt{t}^n,\mathtt{x})$ where $\mathtt{M}_l(\mathtt{f}) = \frac{|V|}{\tau\kappa N_v}\sum_{n,m,j}\mathtt{f}(\mathtt{t}^n,\mathtt{x}_m,\mathtt{v}_j)\mu_l(\mathtt{t}^n,\mathtt{x}_m)$ now denotes the fully discrete analogon to the measurement operator M_l. Boundary conditions are neglected, as computations will involve only compactly supported initial data, and the computational domain can be chosen large enough such that its boundary will not be reached, according to Lemma A.1.5. This yields consistent, stable and convergent schemes.

Proposition C.1.7. *Let $K \in \mathcal{A}_K^{\mathrm{fin}}$. Then the proposed schemes (C.1.7), (C.1.8) to numerically approximate the velocity discrete forward and adjoint chemotaxis equation (C.1.5) and (C.1.6) are accurate of order 1 in time and 2 in space, and therefore consistent in the discrete $L^2(\mathtt{X}\times\mathtt{V})$ norm. Furthermore, they are stable for $\tau/\kappa \le d^{-3/2}$. If a ratio of τ and κ that satisfies the constraint is fixed, then they are convergent as $\kappa \to 0$.*

Proof. The proof only refers to the forward scheme (C.1.7). The result for the adjoint scheme (C.1.8) follows analogously.

Accuracy. For smooth f the Taylor theorem provides

$$\begin{aligned}
&\frac{f(\mathtt{t}+\tau,\mathtt{x},\mathtt{v}) - \mathtt{P}^{\mathrm{LW}}_{\tau,\kappa} f(\mathtt{t},\mathtt{x},\mathtt{v}) - \tau\mathcal{K}(f)(\mathtt{t},\mathtt{x},\mathtt{v})}{\tau} \\
&= \frac{\tau}{2}\left(\partial_{tt} f(\mathtt{t},\mathtt{x},\mathtt{v}) - \sum_{i,j=1}^{d} \mathtt{v}_i\mathtt{v}_j \mathtt{d}^{\kappa}_{x_i x_j} f(\mathtt{t},\mathtt{x},\mathtt{v})\right) + \mathcal{O}(\tau^2) + \mathcal{O}(\kappa^2) \\
&= \frac{\tau}{2}\left(-v\cdot\nabla_x\mathcal{K}(f) + \mathcal{K}(-v\nabla_x f + \mathcal{K}(f))\right) + \mathcal{O}(\tau^2) + \mathcal{O}(\kappa^2) \qquad = \mathcal{O}(\tau) + \mathcal{O}(\kappa^2),
\end{aligned}$$

and thus consistency of the scheme. In the second to last line, a twofold application of (Ch) showed

$$\partial_{tt} f = -v \cdot \nabla_x (\partial_t f) + \mathcal{K}(\partial_t f) = (v \cdot \nabla_x)^2 f - v \cdot \nabla_x \mathcal{K}(f) + \mathcal{K}(-v \cdot \nabla_x f + \mathcal{K}(f)).$$

Stability is given, because the scheme is derived from the Lax Wendroff scheme under addition of a bounded in norm linear operator $\mathcal{K}$ with $L^2(\mathtt{X} \times \mathtt{V})$ operator norm $\|\mathcal{K}\| \leq 2C_K|V|$[Str04, Corollary 2.2.2]. In particular,

$$\|(\mathtt{P}^{\mathrm{chem}}_{\tau,\kappa})^n\| \leq \|\mathtt{P}^{\mathrm{chem}}_{\tau,\kappa}\|^n \leq (1 + 2C_K|V|\tau)^n \leq e^{2C_K|V|\tau n} \leq e^{2C_K|V|T}.$$

Convergence is provided by the Lax-Richtmyer theorem. □

One order of accuracy in time is lost in comparison to the Lax Wendroff scheme due to the first order in time explicit treatment of the tumbling operator, as can be observed by the fact that the tumbling part in the second order in time derivative is not taken care of.

C.2. Finite Element Scheme for the Stationary Schrödinger Equation.

In the following, a Ritz-Galerkin method with nodal basis will be applied to discretize the forward and adjoint Schrödinger equation (8.9) and (8.13). For an exhaustive introduction to finite element methods, the reader is referred to standard literature, such as [Cia02, Red93, Log11, SF97].

Weak formulation. The finite element approach is based on the weak formulations of the equations that, in analogy to (A.2.2) and (A.2.3), read

$$\langle \nabla u_p, \nabla \phi \rangle_{L^2(X)} + \langle p u_p, \phi \rangle_{L^2(X)} = \langle \gamma, \phi \rangle_{L^2(X)}\,, \tag{C.2.1}$$

$$-\langle g_{\hat{x}}, \Delta \psi \rangle_{L^1(X),L^\infty(X)} + \langle p g_{\hat{x}}, \psi \rangle_{L^1(X),L^\infty(X)} = -\psi(\hat{x}) \tag{C.2.2}$$

for test functions $\phi \in H^1_0(X)$ and $\psi \in C_c \cap W^{2,\infty}(X)$, where $\langle \cdot, \cdot \rangle_{L^1(X),L^\infty(X)}$ denotes the $L^1(X), L^\infty(X)$ duality bracket.

Ritz Approximation. After choosing a finite set of basis elements $\{\phi_n\}_{n=1}^N$ of the solution spaces $H^1_0(X)$ and $L^1(X)$, the method then searches for the solution to the equation in the finite dimensional subspace V_N generated by this basis. This is achieved by expanding the solutions $\mathtt{u}_p(x) = \sum_n \mathtt{u}^{(n)} \phi_n(x)$, $\mathtt{g}_{\hat{x}}(x) = \sum_n \mathtt{g}^{(n)} \phi_n(x)$ and testing (C.2.1) and (C.2.2) with $\phi = \psi = \phi_m$, respectively, which yields the linear system

$$S\vec{u} = \vec{F},$$
$$S\vec{g} = \vec{G},$$

where $\vec{\mathtt{u}} = (\mathtt{u}^{(n)})_{n=1}^N$ and $\vec{\mathtt{g}} = (\mathtt{g}^{(n)})_{n=1}^N$ collect the coefficients of $\mathtt{u}_p$ and $\mathtt{g}_{\hat{x}}$, the stiffness matrix $S = (S_{mn})_{m,n=1}^N$ is given by $S_{mn} = \langle \nabla\phi_m, \nabla\phi_n\rangle + \langle p\phi_m, \phi_n\rangle$ - assuming sufficient smoothness of the basis elements - and the loads are $\vec{F} = (\langle \gamma, \phi_m\rangle)_{m=1}^N$ and $\vec{G} = (-\phi_m(\hat{x}))_{m=1}^N$. The linear systems can be solved by inverting S.

Accuracy. Cea's lemma [Cia02, Thm.2.4.1] bounds the approximation error of the finite element solutions $\mathtt{u}_p, \mathtt{g}_{\hat{x}}$ at order of the projection error of $u \in H_0^1(X), g_x \in L^1(X)$ over V_N, the subspace generated by the chosen basis, stating that accuracy of the scheme is solely determined by the choice of the basis.

Choice of Basis. To ensure compatibility of the forward and adjoint computation in the "first optimize, then discretize" setting, we use the same basis will be used for $\mathtt{u}_p$ and $\mathtt{g}_{\hat{x}}$.

A nodal basis will be deployed. It is based on a discretization of the computational space by an equidistant Cartesian grid $\mathtt{X} = \{x_n\}_{n=1}^{(N_x+1)^2}$ with $N_x \in \mathbb{N}$ cells in every direction and a criss triangulation that connects the lower left vertex with the upper right one in every grid cell. The nodal basis functions ϕ_n for interior points $x_n \in \mathtt{X} \setminus \partial X$ are then defined as the unique piece wise linear functions that attain values 1 in x_n and 0 in x_m for $m \neq n$, i.e. $\phi_n(x_m) = \mathbb{1}_{\{n\}}(m)$. Because boundary points are excluded due to the vanishing boundary condition, $N = (N_x - 1)^2$ basis functions are considered, and w.l.o.g. they are numbered from 1 to N.

Computational aspects. Because computing the gradient $\nabla_p u_p(\hat{x})$ requires repeated solves of the adjoint equation, it makes sense to prepared the stiffness matrix beforehand and store it. Analytic forms of the nodal basis elements can be leveraged.

If p changes, for instance due to updating through an adaptive version of the design sampling method, certain parts of the p dependent stiffness matrix can still be precomputed: By the prescribed form of $p \in \mathcal{A}_p$ in (8.11), the scalar products in $S_{mn} = \langle \nabla\phi_m, \nabla\phi_n\rangle + \sum_q p_q \langle b_q\phi_m, \phi_n\rangle$ can be precomputed and the new stiffness matrix can be reassembled through cheap matrix summation.

Glossary of Abbreviations

a.a. almost all. 40, 61, 161, 163

a.e. almost every. 40, 163

a.e.w. almost everywhere. 19, 38, 40, 163

CBS Consensus Based Sampler. 13–15, 136, 138, 141–144, 146–149

E.coli Escherichia coli. 36, 37, 43, 46, 48

e.g. for example. 5, 12, 17, 27, 28, 35, 37, 44, 46, 49, 53, 70, 84, 137, 160

e.t.c. et cetera. 36

EKS Ensemble Kalman Sampler. 13, 136, 138, 141–144, 146–149

FD finite difference. 25

FEM finite element. 25

i.e. that is. 2, 6–8, 10, 11, 13–17, 19–22, 29, 30, 40, 42, 47, 51, 60, 74, 79, 80, 82, 85, 87, 88, 92, 93, 105, 106, 116, 125, 131–133, 137, 161, 163, 165, 169, 183

i.i.d. independent and identically distributed. 13–15, 30, 41, 127, 129, 133, 134

IVP initial value problem. 40

LHS left-hand side. 38

MAP maximum a-posteriori point. 11, 15, 17, 30, 148

MCMC Markov Chain Monte Carlo. 12, 24, 25, 135, 136

ODE ordinary differential equation. 1, 18, 22–26, 54

PDE partial differential equation. 1, 2, 6, 7, 12, 20, 24–27, 31, 35, 38, 39, 42, 43, 50, 53, 55, 82, 103, 131, 135, 137, 164, 167, 177, 178, 180

PKS Patlak-Keller-Segel. 43, 44, 48, 49, 120

RHS right-hand side. 15, 38, 39, 53, 129, 132, 134, 168, 169

RNLA Randomized Numerical Linear Algebra. 125, 126, 128, 147, 154

SDE stochastic differential equation. 13, 14, 28, 41–43, 48

w.l.o.g. without loss of generality. 89, 183

w.r.t. with respect to. 1, 3, 7, 8, 11–13, 18–21, 28, 50, 52, 70, 72–75, 79, 81, 91, 93–95, 104, 110, 112, 120, 121, 131–133, 141, 143, 144, 147, 148, 162, 164, 165, 175

Bibliography

[AAGO21] Kamal Aghazade, Hossein S Aghamiry, Ali Gholami, and Stéphane Operto. Randomized source sketching for full waveform inversion. IEEE Transactions on Geoscience and Remote Sensing, 60:1–12, 2021.

[ABT18] Richard C Aster, Brian Borchers, and Clifford H Thurber. Parameter estimation and inverse problems. Elsevier, 2018.

[AC22] Ahmed Attia and Emil Constantinescu. Optimal experimental design for inverse problems in the presence of observation correlations. SIAM Journal on Scientific Computing, 44(4):A2808–A2842, 2022.

[ACD24] Giacomo Albi, Elisa Calzola, and Giacomo Dimarco. A data-driven kinetic model for opinion dynamics with social network contacts. European Journal of Applied Mathematics, pages 1–27, 2024.

[ACDVM20] Floriane Anstett-Collin, Lilianne Denis-Vidal, and Gilles Millérioux. A priori identifiability: An overview on definitions and approaches. Annual Reviews in Control, 50:139–149, 2020.

[Ada75] Robert A Adams. Sobolev spaces. Elsevier, 1975.

[Ada03] K Zarb Adami. Variational methods in bayesian deconvolution. PHYSTAT 2003, SLAC, pages 143–147, 2003.

[Adl75] Julius Adler. Chemotaxis in bacteria. Annual review of biochemistry, 44(1):341–356, 1975.

[AF12] Thomas Apel and Thomas G. Flaig. Crank–nicolson schemes for optimal control problems with evolution equations. SIAM Journal on Numerical Analysis, 50(3):1484–1512, 2012. doi:10.1137/100819333.

[Ale21] Alen Alexanderian. Optimal experimental design for infinite-dimensional bayesian inverse problems governed by pdes: A review. Inverse Problems, 37, 01 2021. doi:10.1088/1361-6420/abe10c.

[Alt80] W. Alt. Biased random walk models for chemotaxis and related diffusion approximations. Journal of Mathematical Biology, 9:147–177, 1980. doi:10.1007/BF00275919.

[APSG16] Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. A fast and scalable method for a-optimal design of experiments for infinite-dimensional bayesian nonlinear inverse problems. SIAM Journal on Scientific Computing, 38(1):A243–A272, 2016.

[AS09] Simon R Arridge and John C Schotland. Optical tomography: forward and inverse problems. Inverse Problems, 25(12):123010, dec 2009. doi:10.1088/0266-5611/25/12/123010.

[AT21] Gurusamy Arumugam and Jagmohan Tyagi. Keller-segel chemotaxis models: A review. Acta Applicandae Mathematicae, 171, 02 2021. doi:10.1007/s10440-020-00374-2.

[BÅ70] Ror Bellman and Karl Johan Åström. On structural identifiability. Mathematical biosciences, 7(3-4):329–339, 1970.

[Bal09] Guillaume Bal. Inverse transport theory and applications. Inverse Problems, 25(5):053001, mar 2009. doi:10.1088/0266-5611/25/5/053001.

[Bal12] Guillaume Bal. Introduction to inverse problems. Lecture Notes-Department of Applied Physics and Applied Mathematics, Columbia University, New York, 2012.

[BB72] Howard C Berg and Douglas A Brown. Chemotaxis in escherichia coli analysed by three-dimensional tracking. Nature, 239(5374):500–504, 1972. doi:10.1038/239500a0.

[BBDM21] Mario Bertero, Patrizia Boccacci, and Christine De Mol. Introduction to inverse problems in imaging. CRC press, 2021.

[BBKS00] Irene Bauer, Hans Georg Bock, Stefan Körkel, and Johannes P Schlöder. Numerical methods for optimum experimental design in dae systems. Journal of Computational and Applied mathematics, 120(1-2):1–25, 2000.

[BBN22] Elishan C Braun, Gabriella Bretti, and Roberto Natalini. Parameter estimation techniques for a chemotaxis model inspired by cancer-on-chip (coc) experiments. International Journal of Non-Linear Mechanics, 140:103895, 2022.

[BC09] Nikolaos Bournaveas and Vincent Calvez. Critical mass phenomenon for a chemotaxis kinetic model with spherically symmetric initial data. Annales de l'Institut Henri Poincare (C) Non Linear Analysis, 26:1871–1895, 09 2009. doi:10.1016/j.anihpc.2009.02.001.

[BC10] Nikolaos Bournaveas and Vincent Calvez. A review of recent existence and blow-up results for kinetic models of chemotaxis. Can. Appl. Math. Q, 18(3):253–265, 2010.

[BCAB08] Eva Balsa-Canto, Antonio A Alonso, and Julio R Banga. Computational procedures for optimal experimental design in biological systems. IET systems biology, 2(4):163–172, 2008.

[BCGP08] Nikolaos Bournaveas, Vincent Calvez, Susana Gutiérrez, and Benoît Perthame. Global existence for a kinetic model of chemotaxis via dispersion and strichartz estimates. Communications in Partial Differential Equations, 33(1):79–95, 2008.

[BDM+22] J Bouvard, Carine Douarche, P Mergaert, H Auradou, and Frederic Moisy. Direct measurement of the aerotactic response in a bacterial suspension. Physical Review E, 106(3):034404, 2022.

[BDP06] Adrien Blanchet, Jean Dolbeault, and Benoit Perthame. Two-dimensional keller-segel model: Optimal critical mass and qualitative properties of the solutions, in "electron. Electronic Journal of Differential Equations, 44, 04 2006.

[Bec87] M Bruce Beck. Water quality modeling: a review of the analysis of uncertainty. Water resources research, 23(8):1393–1442, 1987.

[Ber75] Howard C Berg. Chemotaxis in bacteria. Annual review of biophysics and bioengineering, 4(1):119–136, 1975.

[Ber04] Howard C Berg. E. coli in Motion. Springer, 2004.

[BFTea20] J. Beal, N.G. Farny, Haddock-Angelli T., and et al. Robust estimation of bacterial cell count from optical density. Communications Biology, 3, 2020. doi:10.1038/s42003-020-01127-5.

[BHO+24] Helen Byrne, Heather Harrington, Alexey Ovchinnikov, Gleb Pogudin, Hamid Rahkooy, and Pedro Soto. Algebraic identifiability of partial differential equation models. arXiv preprint arXiv:2402.04241, 2024.

[BHT23] Nicolas Boullé, Diana Halikias, and Alex Townsend. Elliptic pde learning is provably data-efficient. Proceedings of the National Academy of Sciences, 120(39):e2303904120, 2023.

[Bil20] Sándor Bilicz. Sensitivity analysis of inverse problems in em nondestructive testing. IET Science, Measurement & Technology, 14(5):543–551, 2020.

[BJ08] Guillaume Bal and Alexandre Jollivet. Stability estimates in stationary inverse transport. Inverse Problems and Imaging, 2(4):427–454, 2008.

[BJ09] Guillaume Bal and Alexandre Jollivet. Time-dependent angularly averaged inverse transport. Inverse Problems, 25(7):075010, jun 2009. doi:10.1088/0266-5611/25/7/075010.

[BJ10] Guillaume Bal and Alexandre Jollivet. Stability for time-dependent inverse transport. SIAM journal on mathematical analysis, 42(2):679–700, 2010.

[BJ18] Guillaume Bal and Alexandre Jollivet. Generalized stability estimates in inverse transport theory. Inverse Problems and Imaging, 12(1):59–90, 2018. doi:10.3934/ipi.2018003.

[BJJ10] Guillaume Bal, Alexandre Jollivet, and Vincent Jugnon. Inverse transport theory of photoacoustics. Inverse Problems, 26(2):025011, 2010.

[BJLM11] Guillaume Bal, Alexandre Jollivet, Ian Langmore, and François Monard. Angular average of time-harmonic transport solutions. Communications in Partial Differential Equations, 36(6):1044–1070, 2011.

[BLM08] Guillaume Bal, Ian Langmore, and François Monard. Inverse transport with isotropic sources and angularly averaged measurements. Inverse Probl. Imaging, 2(1):23–42, 2008.

[BM02] Martin Burger and Wolfram Mühlhuber. Iterative regularization of parameter identification problems by sequential quadratic programming methods. Inverse Problems, 18(4):943, may 2002. doi:10.1088/0266-5611/18/4/301.

[BM12] Guillaume Bal and François Monard. Inverse transport with isotropic time-harmonic sources. SIAM Journal on Mathematical Analysis, 44(1):134–161, 2012. doi:10.1137/11083397X.

[BMDB09] Joost B Beltman, Athanasius FM Marée, and Rob J De Boer. Analysing immune cell migration. Nature Reviews Immunology, 9(11):789–798, 2009.

[Bol72] Ludwig Boltzmann. Weitere studien über das wärmegleichgewicht unter gasmolekülen, volume 66. Aus der kk Hot-und Staatsdruckerei, 1872.

[Bre11] H Brezis. Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer Science & Business Media, 2011.

[BRH13] Nawaf Bou-Rabee and Martin Hairer. Nonasymptotic mixing of the MALA algorithm. IMA Journal of Numerical Analysis, 33(1):80–110, 2013.

[BRK01] Roland Brun, Peter Reichert, and Hans R Künsch. Practical identifiability analysis of large environmental simulation models. Water Resources Research, 37(4):1015–1030, 2001.

[BRW24] Leon Bungert, Tim Roith, and Philipp Wacker. Polarized consensus-based dynamics for optimization and sampling. Mathematical Programming, pages 1–31, 2024.

[BS09] Alexandros Beskos and Andrew Stuart. MCMC methods for sampling function space. In Invited Lectures, Sixth International Congress on Industrial

and Applied Mathematics, ICIAM07, Editors Rolf Jeltsch and Gerhard Wanner, pages 337–364, 2009.

[BSAD07] Giuseppina Bellu, Maria Pia Saccomani, Stefania Audoly, and Leontina D'Angiò. Daisy: A new software tool to test global identifiability of biological and physiological systems. Computer methods and programs in biomedicine, 88(1):52–61, 2007.

[BSE+09] Samuel Bandara, Johannes P Schlöder, Roland Eils, Hans Georg Bock, and Tobias Meyer. Optimal experimental design for parameter estimation of a cell signaling model. PLoS computational biology, 5(11):e1000558, 2009.

[BT23] Nicolas Boullé and Alex Townsend. Learning elliptic partial differential equations with randomized linear algebra. Found. Comput. Math., 23(2):709–739, 2023. doi:10.1007/s10208-022-09556-w.

[BTFB24] Alexander P Browning, Maria Taşcă, Carles Falcó, and Ruth E Baker. Structural identifiability analysis of linear reaction–advection–diffusion processes in mathematical biology. Proceedings of the Royal Society A, 480(2286):20230911, 2024.

[BTGMS13] Tan Bui-Thanh, Omar Ghattas, James Martin, and Georg Stadler. A computational framework for infinite-dimensional bayesian inverse problems part i: The linearized case, with application to global seismic inversion. SIAM Journal on Scientific Computing, 35(6):A2494–A2523, 2013.

[BTLZNn22] Tan Bui-Thanh, Qin Li, and Leonardo Zepeda-Núñez. Bridging and improving theoretical and computational electrical impedance tomography via data completion. SIAM Journal on Scientific Computing, 44(3):B668–B693, 2022. doi:10.1137/21M141703X.

[BV04] Stephen Boyd and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.

[BVLE98] Gerrit Burgers, Peter Jan Van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble kalman filter. Monthly weather review, 126(6):1719–1724, 1998.

[BW16] Xueli Bai and Michael Winkler. Equilibration in a fully parabolic two-species chemotaxis system with competitive kinetics. Indiana University Mathematics Journal, pages 553–583, 2016.

[BWB+20] Alexander P Browning, David J Warne, Kevin Burrage, Ruth E Baker, and Matthew J Simpson. Identifiability analysis for stochastic differential equation models in systems biology. Journal of the Royal Society Interface, 17(173):20200652, 2020.

[Cal80] AP Calderón. On an inverse boundary value problem. Seminar on numerical analysis and its applications to continuum physics. Soc. Brasileira de Mathematica, pages 65–73, 1980.

[CBBC11a] Oana Chiş, Julio R Banga, and Eva Balsa-Canto. Genssi: a software toolbox for structural identifiability analysis of biological models. Bioinformatics, 27(18):2610–2611, 2011.

[CBBC11b] Oana-Teodora Chis, Julio R Banga, and Eva Balsa-Canto. Structural identifiability of systems biology models: a critical comparison of methods. PloS one, 6(11):e27755, 2011.

[CCBJ18] Xiang Cheng, Niladri S Chatterji, Peter L Bartlett, and Michael I Jordan. Underdamped langevin mcmc: A non-asymptotic analysis. In Conference on learning theory, pages 300–323. PMLR, 2018.

[CDM+24] Maria-Veronica Ciocanel, Lee Ding, Lucas Mastromatteo, Sarah Reichheld, Sarah Cabral, Kimberly Mowry, and Björn Sandstede. Parameter identifiability in pde models of fluorescence recovery after photobleaching. Bulletin of Mathematical Biology, 86(4):36, 2024.

[CDr80] Claudio Cobelli and Joseph J Distefano 3rd. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 239(1):R7–R24, 1980.

[CEL+15] Wayne Croft, Charles M Elliott, Graham Ladds, Björn Stinner, Chandrasekhar Venkataraman, and Cathryn Weston. Parameter identification problems in the modelling of cell motility. Journal of mathematical biology, 71:399–436, 2015.

[Cer12] C. Cercignani. The Boltzmann Equation and Its Applications. Applied Mathematical Sciences. Springer New York, 2012.

[CFG14] Tianqi Chen, Emily Fox, and Carlos Guestrin. Stochastic gradient hamiltonian monte carlo. In International conference on machine learning, pages 1683–1691. PMLR, 2014.

[CFRT09] J. Carrillo, M Fornasier, Jesús Rosado, and Giuseppe Toscani. Asymptotic flocking dynamics for the kinetic cucker–smale model. SIAM Journal on Mathematical Analysis, 42, 05 2009. doi:10.1137/090757290.

[CG08] Jin-De Chang and Bao-Zhu Guo. Application of ingham-beurling-type theorems to coefficient identifiability of vibrating systems: Finite time identifiability. Differential and Integral Equations, 21(11-12):1037–1054, 2008.

[CH12] Matthias Chung and Eldad Haber. Experimental design for biological systems. SIAM Journal on Control and Optimization, 50(1):471–489, 2012.

[Cha50] Subrahmanyan Chandrasekhar. Radiative transfer. Oxford, 1950.

[CHSV22] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes. Consensus-based sampling. Studies in Applied Mathematics, 148(3):1069–1140, 2022. doi:10.1111/sapm.12470.

[Cia02] Philippe G Ciarlet. The finite element method for elliptic problems. SIAM, 2002.

[CJ21] Ke Chen and Ruhui Jin. Tensor-structured sketching for constrained least squares. SIAM Journal on Matrix Analysis and Applications, 42(4):1703–1731, 2021.

[CKK97] C-K Cho, Y Kwon, and S Kang. Estimation of soil water distribution: Identifiability and observation design. Computers & Mathematics with Applications, 34(12):105–120, 1997.

[CKLMNO19] A. Chertock, A. Kurganov, M. Lukáčová-Medvid'ová, and Ş. Nur Özcan. An asymptotic preserving scheme for kinetic chemotaxis models in two space dimensions. Kinetic & Related Models, 12(1):195–216, 2019.

[CLL18] Ke Chen, Qin Li, and Jian-Guo Liu. Online learning in optical tomography: a stochastic approach. Inverse Problems, 34(7):075010, may 2018. doi:10.1088/1361-6420/aac220.

[CLL20] Francis J Chung, Ru-Yu Lai, and Qin Li. On diffusive scaling in acousto-optic imaging. Inverse Problems, 36(8):085011, 2020.

[CLNW20] Ke Chen, Qin Li, Kit Newton, and Stephen J Wright. Structured random sketching for pde inverse problems. SIAM Journal on Matrix Analysis and Applications, 41(4):1742–1770, 2020.

[CLP24] Weiqi Chu, Qin Li, and Mason A Porter. Inference of interaction kernels in mean-field models of opinion dynamics. SIAM Journal on Applied Mathematics, 84(3):1096–1115, 2024.

[CLW18a] Ke Chen, Qin Li, and Li Wang. Stability of inverse transport equation in diffusion scaling and fokker–planck limit. SIAM Journal on Applied Mathematics, 78(5):2626–2647, 2018. doi:10.1137/17M1157969.

[CLW18b] Ke Chen, Qin Li, and Li Wang. Stability of stationary inverse transport equation in diffusion scaling. Inverse Problems, 34(2):025004, 2018.

[CM81] V Capasso and Lucia Maddalena. Convergence to equilibrium states for a reaction-diffusion system modelling the spatial spread of a class of bacterial and viral diseases. Journal of Mathematical Biology, 13:173–184, 1981.

[CMPS04] F. Chalub, P. Markowich, B. Perthame, and C. Schmeiser. Kinetic models for chemotaxis and their drift-diffusion limits. Monatsh. Math., 142:123–141, 2004. doi:10.1007/s00605-004-0234-7.

[Cox69] H. S. M. (Harold Scott Macdonald) Coxeter. Introduction to Geometry. Wiley, 2nd edition, 1969.

[CPA81] M. Courdesses, M. Polis, and M. Amouroux. On identifiability of parameters in a class of parabolic distributed systems. IEEE Transactions on Automatic Control, 26(2):474–477, 1981. doi:10.1109/TAC.1981.1102609.

[CR80] L. Carotenuto and G. Raiconi. Identifiability and identification of a galerkin approximation for a class of distributed parameter systems†. International Journal of Systems Science, 11(9):1035–1049, 1980. doi:10.1080/00207728008967072.

[CR06] FACC Chalub and José Francisco Rodrigues. A class of kinetic models for chemotaxis with threshold to prevent overcrowding. Portugaliae Mathematica, 63(2):227, 2006. URL: http://eudml.org/doc/53089.

[CRY19] Frédérique Clément, Frédérique Robin, and Romain Yvinec. Analysis and calibration of a linear model for structured cell populations with unidirectional motion: Application to the morphogenesis of ovarian follicles. SIAM Journal on Applied Mathematics, 79(1):207–229, 2019. doi:10.1137/17M1161336.

[CS96a] Mourad Choulli and Plamen Stefanov. Inverse scattering and inverse boundary value problems for the linear boltzmann equation. Communications in Partial Differential Equations, 21(5-6):763–785, 1996.

[CS96b] Mourad Choulli and Plamen Stefanov. Reconstruction of the coefficients of the stationary transport equation from boundary measurements. Inverse Problems, 12(5):L19–L23, oct 1996. doi:10.1088/0266-5611/12/5/001.

[CS99] Mourad Choulli and Plamen Stefanov. An inverse boundary value problem for the stationary transport equation. Osaka Journal of Mathematics, 1999.

[CV95] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. Statistical science, pages 273–304, 1995.

[CV02] Didier Chauveau and Pierre Vandekerkhove. Improving convergence of the hastings-metropolis algorithm with an adaptive proposal. Scandinavian Journal of Statistics, 29(1):13–29, 2002. URL: http://www.jstor.org/stable/4616696.

[CZ63] KM Case and Paul Frederick Zweifel. Existence and uniqueness theorems for the neutron transport equation. Journal of Mathematical Physics, 4(11):1376–1385, 1963.

[DAB03] Rita M.C. De Almeida and Israel Jacob Rabin Baumvol. Reaction–diffusion in high-k dielectrics on si. Surface Science Reports, 49(1-3):1–114, 2003.

[Das96] Anirban DasGupta. 29 review of optimal bayes designs. Handbook of Statistics, 13:1099–1147, 1996.

[DBP+12] Simona Dobre, Thierry Bastogne, Christophe Profeta, Muriel Barberi-Heyob, and Alain Richard. Limits of variance-based sensitivity analysis for non-identifiability testing in high dimensional dynamic models. Automatica, 48(11):2740–2749, 2012.

[DCWY19] Raaz Dwivedi, Yuansi Chen, Martin J Wainwright, and Bin Yu. Log-concave sampling: Metropolis-hastings algorithms are fast. Journal of Machine Learning Research, 20(183):1–42, 2019.

[DET24] Marie Doumic, Miguel Escobedo, and Magali Tournus. An inverse problem: recovering the fragmentation kernel from the short-time behaviour of the fragmentation equation. Annales Henri Lebesgue, 7:621–671, 2024. doi:10.5802/ahl.207.

[DK19] Arnak S Dalalyan and Avetik Karagulyan. User-friendly guarantees for the langevin monte carlo with inaccurate gradient. Stochastic Processes and their Applications, 129(12):5278–5311, 2019.

[DL21] Zhiyan Ding and Qin Li. Ensemble kalman sampler: Mean-field limit and convergence analysis. SIAM Journal on Mathematical Analysis, 53(2):1546–1578, 2021. doi:10.1137/20M1339507.

[DNS23] Andrew Duncan, Nikolas Nüsken, and Lukasz Szpruch. On the geometry of stein variational gradient descent. Journal of Machine Learning Research, 24(56):1–39, 2023.

[DP04] Jean Dolbeault and Benoît Perthame. Optimal critical mass in the two dimensional keller–segel model in r2. Comptes Rendus Mathematique, 339(9):611–616, 2004. doi:10.1016/j.crma.2004.08.011.

[DP14] Giacomo Dimarco and Lorenzo Pareschi. Numerical methods for kinetic equations. Acta Numerica, 23:369–520, 2014.

[DP19] Sergey Dolgov and John W Pearson. Preconditioners and tensor product solvers for optimal control problems from chemotaxis. SIAM Journal on Scientific Computing, 41(6):B1228–B1253, 2019.

[DS58] B. Davison and J.B. Sykes. Neutron Transport Theory. International series of monographs on physics. Clarendon Press ; [Oxford University Press], 1958.

[DS17] Masoumeh Dashti and Andrew M. Stuart. The Bayesian Approach to Inverse Problems, pages 311–428. Springer International Publishing, Cham, 2017. doi:10.1007/978-3-319-12385-1_7.

[DT11] Judit Danis and Tamás Turányi. Sensitivity analysis of bacterial chemotaxis models. Procedia Computer Science, 7:233–234, 2011.

[DuC13] Paul DuChateau. An adjoint method for proving identifiability of coefficients in parabolic equations. Journal of Inverse and Ill-Posed Problems, 21(5):639–663, 2013.

[Dun19] Matthew M Dunlop. Multiplicative noise in bayesian inverse problems: Well-posedness and consistency of map estimators. arXiv preprint arXiv:1910.14632, 2019.

[DVJB00] Lilianne Denis-Vidal and Ghislaine Joly-Blanchard. An easy to check criterion for (un) indentifiability of uncontrolled systems and its applications. IEEE Transactions on Automatic Control, 45(4):768–771, 2000.

[EGB+16] Casimir Emako, Charlène Gayrard, Axel Buguin, Luís Neves de Almeida, and Nicolas Vauchelet. Traveling pulses for a two-species chemotaxis model. PLOS Computational Biology, 12(4):1–22, 04 2016. doi:10.1371/journal.pcbi.1004843.

[EHPS24] Herbert Egger, Kathrin Hellmuth, Nora Philippi, and Matthias Schlottbom. A kinetic chemotaxis model and its diffusion limit in slab geometry. arXiv preprint arXiv:2408.17243, 2024.

[EMJ10] Derek L Englert, Michael D Manson, and Arul Jayaraman. Investigation of bacterial chemotaxis in flow-based microfluidic devices. Nature protocols, 5(5):864–872, 2010.

[EN01] Klaus-Jochen Engel and Rainer Nagel. One-parameter semigroups for linear evolution equations. Semigroup Forum, 63:278–280, 06 2001. doi:10.1007/s002330010042.

[EO04] Radek Erban and Hans Othmer. From individual to collective behavior in bacterial chemotaxis. SIAM Journal of Applied Mathematics, 65:361–391, 01 2004. doi:10.1137/S0036139903433232.

[EPS15] Herbert Egger, Jan-Frederik Pietschmann, and Matthias Schlottbom. Identification of chemotaxis models with volume-filling. SIAM Journal on Applied Mathematics, 75(2):275–288, 2015. doi:10.1137/140967222.

[ES13] Herbert Egger and Matthias Schlottbom. Numerical methods for parameter identification in stationary radiative transfer. Computational Optimization and Applications, page 67–83, 11 2013. doi:10.1007/s10589-014-9657-9.

[ESV09] Elio E Espejo, Angela Stevens, and Juan JL Velázquez. Simultaneous finite time blow-up in a two-species model for chemotaxis. 2009.

[Eva22] Lawrence C Evans. Partial differential equations, volume 19. American Mathematical Society, 2022.

[EVVL22] Geir Evensen, Femke C Vossepoel, and Peter Jan Van Leeuwen. Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem. Springer Nature, 2022.

[FAJ05] Patrick Flaherty, Adam Arkin, and Michael Jordan. Robust design of biological experiments. Advances in neural information processing systems, 18, 2005.

[FCCB24] Carles Falcó, Daniel J Cohen, José A Carrillo, and Ruth E Baker. Mechanical constraints and cell cycle regulation in models of collective cell migration. arXiv preprint arXiv:2401.08805, 2024.

[FGBC22] Nicola Forti, Lin Gao, Giorgio Battistelli, and Luigi Chisci. Unknown source in spatially distributed systems: Identifiability analysis and estimation. Automatica, 136:110025, 2022. doi:10.1016/j.automatica.2021.110025.

[Fis61] Franklin M Fisher. Identifiability criteria in nonlinear systems. Econometrica: Journal of the Econometric Society, pages 574–590, 1961.

[FL91] Roseanne M. Ford and Douglas A. Lauffenburger. Measurement of bacterial random motility and chemotaxis coefficients: II. Application of single-cell-based mathematical model. Biotechnology and Bioengineering, 37(7):661–672, 1991. doi:10.1002/bit.260370708.

[FLP05] Francis Filbet, Philippe Laurençot, and Benoit Perthame. Derivation of hyperbolic models for chemosensitive movement. journal of mathematical biology 50:189-207. Journal of mathematical biology, 50:189–207, 03 2005. doi:10.1007/s00285-004-0286-2.

[FM03] K Renee Fister and C Maeve McCarthy. Optimal control of a chemotaxis system. Quarterly of Applied Mathematics, pages 193–211, 2003.

[FM08] K. Renee Fister and Maeve L. McCarthy. Identification of a chemotactic sensitivity in a coupled system. Mathematical Medicine and Biology: A Journal of the IMA, 25(3):215–232, 07 2008. doi:10.1093/imammb/dqn015.

[Fol99] Gerald B Folland. Real analysis: modern techniques and their applications, volume 40. John Wiley & Sons, 1999.

[FSU19] Joel Feldman, Mikko Salo, and Gunther Uhlmann. The calderón problem—an introduction to inverse problems. Preliminary notes on the book in preparation, 30, 2019.

[FY13] Francis Filbet and Chang Yang. Numerical simulations of kinetic models for chemotaxis. SIAM Journal on Scientific Computing, 36, 03 2013. doi:10.1137/130910208.

[GAM+15] Andrea Giometto, Florian Altermatt, Amos Maritan, Roman Stocker, and Andrea Rinaldo. Generalized receptor law governs phototaxis in the phytoplankton euglena gracilis. Proceedings of the National Academy of Sciences, 112(22):7045–7050, 2015. doi:10.1073/pnas.1422922112.

[GDI85] KR Godfrey and JJ DiStefano III. Identifiability of model parameter. IFAC Proceedings Volumes, 18(5):89–114, 1985.

[GH07] Semion Gutman and Junhong Ha. Identifiability of piecewise constant conductivity in a heat conduction process. SIAM journal on control and optimization, 46(2):694–713, 2007.

[GH21] Yishu Gong and Siming He. On the 8π-critical-mass threshold of a patlak–keller–segel–navier–stokes system. SIAM Journal on Mathematical Analysis, 53(3):2925–2956, 2021.

[GIHLS20] Alfredo Garbuno-Inigo, Franca Hoffmann, Wuchen Li, and Andrew M. Stuart. Interacting langevin diffusions: Gradient structure and ensemble kalman sampler. SIAM Journal on Applied Dynamical Systems, 19(1):412–441, 2020. doi:10.1137/19M1251655.

[Giu91] Mauro Giudici. Identifiability of distributed physical parameters in diffusive-like systems. Inverse problems, 7(2):231, 1991.

[GJML+19] Joseph HA Guillaume, John D Jakeman, Stefano Marsili-Libelli, Michael Asher, Philip Brunner, Barry Croke, Mary C Hill, Anthony J Jakeman, Karel J Keesman, Saman Razavi, et al. Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. Environmental Modelling & Software, 119:418–432, 2019.

[GLL24] Yuan Gao, Quanjun Lang, and Fei Lu. Self-test loss functions for learning weak-form operators and gradient flows, 2024. URL: https://arxiv.org/abs/2412.03506, arXiv:2412.03506.

[Gol51] Sidney Goldstein. On diffusion by discontinuous movements, and on the telegraph equation. The Quarterly Journal of Mechanics and Applied Mathematics, 4(2):129–156, 1951.

[GQMT22] Hayriye Gulbudak, Zhuolin Qu, Fabio Milner, and Necibe Tuncer. Sensitivity analysis in an immuno-epidemiological vector-host model. Bulletin of mathematical biology, 84(2):27, 2022.

[GRI21] Victor Gallego and David Rios Insua. Variationally inferred sampling through a refined bound. Entropy, 23(1):123, 2021.

[Gro93] Charles Groetsch. Inverse Problems in the Mathematical Sciences. Vieweg+Teubner Verlag, 01 1993. doi:10.1007/978-3-322-99202-4.

[GT21] Marianne Grognot and Katja M Taute. A multiscale 3d chemotaxis assay reveals bacterial navigation mechanisms. Communications biology, 4(1):669, 2021.

[Gun02] Max D. Gunzburger. Perspectives in Flow Control and Optimization. Society for Industrial and Applied Mathematics, 2002. doi:10.1137/1.9780898718720.

[GVB17] Attila Gábor, Alejandro F Villaverde, and Julio R Banga. Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. BMC systems biology, 11:1–16, 2017.

[Had02] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. Princeton university bulletin, pages 49–52, 1902.

[HAO00] Eldad Haber, Uri M Ascher, and Doug Oldenburg. On optimization techniques for solving nonlinear inverse problems. Inverse Problems, 16(5):1263, oct 2000. doi:10.1088/0266-5611/16/5/309.

[Has70] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. Biometrika, 57(1):97–109, 1970. URL: http://www.jstor.org/stable/2334940.

[HB21] Zhishen Huang and Stephen Becker. Spectral estimation from simulations via sketching. Journal of Computational Physics, 447:110686, 2021.

[HBY10] Fei He, Martin Brown, and Hong Yue. Maximin and bayesian robust experimental design for measurement set selection in modelling biochemical regulatory systems. International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal, 20(9):1059–1078, 2010.

[Hen72] Jørgen Henrichsen. Bacterial surface translocation: a survey and a classification. Bacteriological reviews, 36(4):478–503, 1972.

[HG13] Junhong Ha and Semion Gutman. Identifiability for linearized sine-gordon equation. Mathematical Modelling of Natural Phenomena, 8, 01 2013. doi:10.1051/mmnp/20138107.

[HH83] Gerald L Hazelbauer and Shigeaki Harayama. Sensory transduction in bacterila chemotaxis. International Review of Cytology, 81:33–70, 1983.

[HHT13] Sabrina Hock, Jan Hasenauer, and Fabian J Theis. Modeling of 2d diffusion processes based on microscopy data: parameter estimation and practical identifiability analysis. BMC bioinformatics, 14:1–6, 2013.

[HJ85] Roger A. Horn and Charles R. Johnson. Matrix Analysis. Cambridge University Press, 1985. doi:10.1017/CBO9780511810817.

[HJM24] Xun Huan, Jayanth Jagalur, and Youssef Marzouk. Optimal experimental design: Formulations and computations. Acta Numerica, 33:715–840, 2024.

[HKL24] Kathrin Hellmuth, Christian Klingenberg, and Qin Li. Preserving positivity of gauss-newton hessian through random sampling, 2024. arXiv:2409.15906.

[HKLT21] Kathrin Hellmuth, Christian Klingenberg, Qin Li, and Min Tang. Multiscale convergence of the inverse problem for chemotaxis in the bayesian setting. Computation, 9(11):119, 2021.

[HKLT24] Kathrin Hellmuth, Christian Klingenberg, Qin Li, and Min Tang. Kinetic chemotaxis tumbling kernel determined from macroscopic quantities. SIAM Journal on Mathematical Analysis, 56(1):568–587, 2024. doi:10.1137/22M1499911.

[HKLT25] Kathrin Hellmuth, Christian Klingenberg, Qin Li, and Min Tang. Reconstructing the kinetic chemotaxis kernel using macroscopic data: well-posedness and ill-posedness. SIAM Journal on Applied Mathematics, accepted, 2025. arXiv:2309.05004.

[HKS05] Hyung Ju Hwang, Kyungkeun Kang, and Angela Stevens. Global solutions of nonlinear transport equations for chemosensitive movement. SIAM Journal on Mathematical Analysis, 36(4):1177–1199, 2005. doi:10.1137/S0036141003431888.

[HKS06] Hyung Hwang, Kyungkeun Kang, and Angela Stevens. Global existence of classical solutions for a hyperbolic chemotaxis model and its parabolic limit. Indiana University Mathematics Journal, v.55, 289-316 (2006), 55, 01 2006. doi:10.1512/iumj.2006.55.2677.

[HL04] Didier Henrion and J-B Lasserre. Solving nonconvex optimization problems. IEEE Control Systems Magazine, 24(3):72–83, 2004.

[HLH21] Sunwoo Hwang, Seongwon Lee, and Hyung Ju Hwang. Neural network approach to data-driven estimation of chemotactic sensitivity in the keller-segel model. Mathematical Biosciences and Engineering, 18(6):8524–8534, 2021.

[HM97] Sybille Handrock-Meyer. Identifiability of distributed parameters for a class of quasilinear differential equations. 1997.

[HO00] Thomas Hillen and Hans Othmer. The diffusion limit of transport equations derived from velocity jump processes. SIAM Journal on Applied Mathematics, 61, 10 2000. doi:10.1137/S0036139999358167.

[Hor03] Dirk Horstmann. From 1970 until present: the keller-segel model in chemotaxis and its consequences. 2003.

[HP07] Andreas Hofinger and Hanna K Pikkarainen. Convergence rate for the bayesian approach to linear inverse problems. Inverse Problems, 23(6):2469, 2007.

[HP09] Thomas Hillen and Kevin J Painter. A user's guide to pde models for chemotaxis. Journal of mathematical biology, 58(1-2):183–217, 2009.

[HPUU08] Michael Hinze, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. Optimization with pde constraints. In Mathematical Modelling, 2008.

[HRL01] Thomas Hillen, Christian Rohde, and Frithjof Lutscher. Existence of weak solutions for a hyperbolic model of chemosensitive movement. Journal of mathematical analysis and applications, 260(1):173–199, 2001.

[HRP+21] Jack D Hywood, Gregory Rice, Sophie V Pageon, Mark N Read, and Maté Biro. Detection and characterization of chemotaxis without cell tracking. Journal of the Royal Society Interface, 18(176):20200879, 2021.

[HS00] T. Hillen and A. Stevens. Hyperbolic models for chemotaxis in 1-d. Nonlinear Analysis: Real World Applications, 1(3):409–433, 2000. doi:10.1016/S0362-546X(99)00284-9.

[HV19] Michael Herty and Giuseppe Visconti. Kinetic methods for inverse problems. Kinetic and Related Models, 12(5):1109–1130, 2019.

[HW07] Sander C Hille and Daniël TH Worm. Global existence of positive mild solutions for a class of kinetic chemotaxis models, 2007.

[ILS14] Marco A Iglesias, Kui Lin, and Andrew M Stuart. Well-posed bayesian geometric inverse problems arising in subsurface flow. Inverse Problems, 30(11):114001, 2014.

[IMB+06] Donald E Ingber, Van C Mow, David Butler, Laura Niklason, Johnny Huard, Jeremy Mao, Ioannis Yannas, David Kaplan, and Gordana Vunjak-Novakovic. Tissue engineering and developmental biology: going biomimetic. Tissue engineering, 12(12):3265–3283, 2006.

[JDBC12] Thomas Julou, Nicolas Desprat, David Bensimon, and Vincent Croquette. Monitoring microbial population dynamics at low densities. Review of Scientific Instruments, 83(7):074301, 2012. doi:10.1063/1.4729796.

[JJH+19] Hannah Jeckel, Eric Jelli, Raimo Hartmann, Praveen K. Singh, Rachel Mok, Jan Frederik Totz, Lucia Vidakovic, Bruno Eckhardt, Jörn Dunkel, and Knut Drescher. Learning the space-time phase diagram of bacterial swarm expansion. Proceedings of the National Academy of Sciences, 116(5):1489–1494, 2019. doi:10.1073/pnas.1811722116.

[JK+17] Prateek Jain, Purushottam Kar, et al. Non-convex optimization for machine learning. Foundations and Trends® in Machine Learning, 10(3-4):142–363, 2017.

[JLNS24] Ruhui Jin, Qin Li, Anjali Nair, and Samuel Stechmann. Unique reconstruction for discretized inverse problems: a random sketching approach via subsampling, 2024. arXiv:2403.05935.

[JS02] Christine Josenhans and Sebastian Suerbaum. The role of motility as a virulence factor in bacteria. International Journal of Medical Microbiology, 291(8):605–614, 2002. doi:10.1078/1438-4221-00173.

[KA21a] Omer Karin and Uri Alon. Temporal fluctuations in chemotaxis gain implement a simulated-tempering strategy for efficient navigation in complex environments. Iscience, 24(7), 2021.

[KA21b] O Krivorotko and D Andornaya. Sensitivity analysis and practical identifiability of the mathematical model for partial differential equations. In Journal of Physics: Conference Series, volume 2092, page 012012. IOP Publishing, 2021.

[Kac74] Mark Kac. A stochastic model related to the telegrapher's equation. Rocky Mountain Journal of Mathematics, 4(3):497 – 510, 1974. doi:10.1216/RMJ-1974-4-3-497.

[Kie59] J. Kiefer. Optimum experimental designs. Journal of the Royal Statistical Society. Series B (Methodological), 21(2):272–319, 1959. URL: http://www.jstor.org/stable/2983802.

[Kir21] Andreas Kirsch. An introduction to the mathematical theory of inverse problems, volume 120. Springer, 3rd edition, 2021.

[KN77] S. Kitamura and S. Nakagiri. Identifiability of spatially-varying and constant parameters in distributed systems of parabolic type. SIAM Journal on Control and Optimization, 15(5):785–802, 1977. doi:10.1137/0315050.

[Kob80] Toshihiro Kobayashi. Parameter identifiability for distributed parameter systems of hyperbolic type. International Journal of Systems Science, 11(2):247–259, 1980. doi:10.1080/00207728008967011.

[Koo49] Tjalling C Koopmans. Identification problems in economic model construction. Econometrica, Journal of the Econometric Society, pages 125–144, 1949.

[KP08] Michael V. Klibanov and Sergey E. Pamyatnykh. Global uniqueness for a coefficient inverse problem for the non-stationary transport equation via carleman estimate. Journal of Mathematical Analysis and Applications, 343(1):352–365, 2008. doi:10.1016/j.jmaa.2008.01.071.

[KR50] Tjalling C Koopmans and Olav Reiersol. The identification of structural characteristics. The Annals of Mathematical Statistics, 21(2):165–181, 1950.

[Kra88] C. Kravaris. Identifiability of the nonlinearity in a quasilinear parabolic system (heat conduction). In Proceedings of the 27th IEEE Conference on Decision and Control, pages 245–248 vol.1, 1988. doi:10.1109/CDC.1988.194303.

[KS70] Evelyn F. Keller and Lee A. Segel. Initiation of slime mold aggregation viewed as an instability. Journal of Theoretical Biology, 26(3):399–415, 1970. doi:10.1016/0022-5193(70)90092-5.

[KS71] Evelyn F. Keller and Lee A. Segel. Traveling bands of chemotactic bacteria: A theoretical analysis. Journal of Theoretical Biology, 30(2):235–248, 1971. doi:10.1016/0022-5193(71)90051-8.

[KS86] Costas Kravaris and John H. Seinfeld. Identifiability of spatially-varying conductivity from point observation as an inverse sturm–liouville problem. SIAM Journal on Control and Optimization, 24(3):522–542, 1986. doi:10.1137/0324030.

[KS06] Jari Kaipio and Erkki Somersalo. Statistical and computational inverse problems, volume 160. Springer Science & Business Media, 2006.

[KS19] Nikola B Kovachki and Andrew M Stuart. Ensemble kalman inversion: a derivative-free technique for machine learning tasks. Inverse Problems, 35(9):095005, aug 2019. doi:10.1088/1361-6420/ab1c3a.

[KW86] Karl Kunisch and Luther W White. Parameter identifiability under approximation. Quarterly of applied mathematics, 44(3):475–486, 1986.

[KZZ+24] Christina Kurzthaler, Yongfeng Zhao, Nan Zhou, Jana Schwarz-Linek, Clemence Devailly, Jochen Arlt, Jian-Dong Huang, Wilson C. K. Poon, Thomas Franosch, Julien Tailleur, and Vincent A. Martinez. Characterization and control of the run-and-tumble dynamics of escherichia coli. Phys. Rev. Lett., 132:038302, Jan 2024. doi:10.1103/PhysRevLett.132.038302.

[Lah87] A. Lahouaoula. Identifiability and identification of linear distributed systems via double general orthogonal polynomials. International Journal of Control, 46(5):1771–1782, 1987. doi:10.1080/00207178708934009.

[Lan96] Cornelius Lanczos. Linear differential operators. SIAM, 1996.

[Lan08] Ian Langmore. The stationary transport problem with angularly averaged measurements. Inverse problems, 24(1):015024, 2008.

[Lat23] Jonas Latz. Bayesian inverse problems are usually well-posed. SIAM Review, 65(3):831–865, 2023.

[LC89] A. Lahouaoula and M. Courdesses. Identifiability and identification of a class of parabolic distributed systems under approximation. International Journal of Systems Science, 20(4):683–697, 1989. doi:10.1080/00207728908910160.

[LDM22] Nicholas N Lam, Paul D Docherty, and Rua Murray. Practical identifiability of parametrised models: A review of benefits and limitations of various approaches. Mathematics and Computers in Simulation, 199:202–216, 2022.

[LEI00] D. Lesnic, L. Elliott, and D. B. Ingham. Identifiability of distributed parameters for high-order quasi-linear differential equations. Journal of Inverse and Ill-posed Problems, 8(1):1–22, 2000. doi:10.1515/jiip.2000.8.1.1.

[Les00] D Lesnic. Identifiability of distributed parameters in beam-type systems. Journal of Inverse and Ill-posed Problems, 8(4):379–397, 2000.

[LF01] Paul Lewus and Roseanne M Ford. Quantification of random motility and chemotaxis bacterial transport coefficients using individual-cell and population-scale assays. Biotechnology and bioengineering, 75(3):292–304, 2001.

[LFKS13] Juliane Liepe, Sarah Filippi, Michał Komorowski, and Michael PH Stumpf. Maximizing the information content of experiments in systems biology. PLoS computational biology, 9(1):e1002888, 2013.

[LG94] Lennart Ljung and Torkel Glad. On global identifiability for arbitrary model parametrizations. automatica, 30(2):265–276, 1994.

[LHL10] Mark P Little, Wolfgang F Heidenreich, and Guangquan Li. Parameter identifiability and redundancy: theoretical considerations. PloS one, 5(1):e8915, 2010.

[LL92] Randall J LeVeque and Randall J Leveque. Numerical methods for conservation laws, volume 214. Springer, 1992.

[LL18] Benno Liebchen and Hartmut Löwen. Synthetic chemotaxis and collective behavior in active matter. Accounts of Chemical Research, 51(12):2982–2990, 2018. doi:10.1021/acs.accounts.8b00215.

[LL23] Quanjun Lang and Fei Lu. Identifiability of interaction kernels in mean-field equations of interacting particles. Foundations of Data Science, 5, 01 2023. doi:10.3934/fods.2023007.

[LL24] Yuhan Li and Catharine WK Lo. On the simultaneous recovery of environmental factors in the 3d chemotaxis-navier-stokes models. Communications on Analysis and Computation, 2(1):30–47, 2024.

[LLM+21] Zhongyang Li, Fei Lu, Mauro Maggioni, Sui Tang, and Cheng Zhang. On the identifiability of interaction functions in systems of interacting particles. Stochastic Processes and their Applications, 132:135–163, 2021. doi:10.1016/j.spa.2020.10.005.

[LLU19] Ru-Yu Lai, Qin Li, and Gunther Uhlmann. Inverse problems for the stationary transport equation in the diffusion scaling. SIAM Journal on Applied Mathematics, 79(6):2340–2358, 2019.

[LM84] Elmer Eugene Lewis and Warren F Miller. Computational methods of neutron transport. 1984.

[LMT15] Quan Long, Mohammad Motamed, and Raúl Tempone. Fast bayesian optimal experimental design for seismic source inversion. Computer Methods in Applied Mechanics and Engineering, 291:123–145, 2015.

[LNJ18] Seongjin Lim, Hyeono Nam, and Jessie S Jeon. Chemotaxis model for breast cancer cells based on signal/noise ratio. Biophysical journal, 115(10):2034–2043, 2018.

[LNW22] Qin Li, Kit Newton, and Li Wang. Bayesian instability of optical imaging: Ill conditioning of inverse linear and nonlinear radiative transfer equation in the fluid regime. Computation, 10(2):15, 2022.

[LO23] Li Li and Zhimeng Ouyang. Determining the collision kernel in the boltzmann equation near the equilibrium. Proceedings of the American Mathematical Society, 151(11):4855–4865, 2023.

[Log11] Daryl L Logan. A first course in the finite element method, volume 4. Thomson, 2011.

[LPP+19] Jason D Lee, Ioannis Panageas, Georgios Piliouras, Max Simchowitz, Michael I Jordan, and Benjamin Recht. First-order methods almost always avoid strict saddle points. Mathematical programming, 176:311–337, 2019.

[LR05] Peter D Lax and Robert D Richtmyer. Survey of the stability of linear finite difference equations. In Selected Papers Volume I, pages 125–151. Springer, 2005.

[LS82] Kenneth R Lutchen and Gerald M Saidel. Sensitivity analysis and experimental design techniques: application to nonlinear, dynamic lung models. Computers and Biomedical Research, 15(5):434–454, 1982.

[LS20] Qin Li and Weiran Sun. Applications of kinetic tools to inverse transport problems. Inverse Problems, 36(3):035011, Feb 2020. doi:10.1088/1361-6420/ab59b8.

[LSM+24] Yue Liu, Kevin Suh, Philip K Maini, Daniel J Cohen, and Ruth E Baker. Parameter identifiability and model selection for partial differential equation models of cell invasion. Journal of the Royal Society Interface, 21(212):20230607, 2024. doi:10.1098/rsif.2023.0607.

[LSTW13] Quan Long, Marco Scavino, Raúl Tempone, and Suojin Wang. Fast estimation of expected information gains for bayesian experimental designs based on laplace approximations. Computer Methods in Applied Mechanics and Engineering, 259:24–39, 2013.

[LTD+15] Shu Li, Xiaoling Tan, Nicolas Desneux, Giovanni Benelli, Jing Zhao, Xinhai Li, Fan Zhang, Xiwu Gao, and Su Wang. Innate positive chemotaxis to pollen from crops and banker plants in predaceous biological control agents: towards new field lures? Scientific Reports, 5(1):12729, 2015.

[LUY21] Ru-Yu Lai, Gunther Uhlmann, and Yang Yang. Reconstruction of the collision kernel in the nonlinear boltzmann equation. SIAM Journal on Mathematical Analysis, 53(1):1049–1069, 2021.

[LW64] Peter D. Lax and Burton Wendroff. Difference schemes for hyperbolic equations with high order of accuracy. Communications on Pure and Applied Mathematics, 17(3):381–398, 1964. doi:10.1002/cpa.3160170311.

[LW19] Jun Liu and Zhu Wang. Non-commutative discretize-then-optimize algorithms for elliptic pde-constrained optimal control problems. Journal of Computational and Applied Mathematics, 362:596–613, 2019. doi:10.1016/j.cam.2018.07.028.

[LZWZ16] Chang Liu, Chao Zhou, Wei Wang, and H. P. Zhang. Bimetallic microswimmers speed up in confining channels. Phys. Rev. Lett., 117:198001, Nov 2016. doi:10.1103/PhysRevLett.117.198001.

[Mah16] Michael W. Mahoney. Lecture notes on randomized linear algebra, 2016. arXiv:1608.04481.

[Maj97] Armando Majorana. Existence and uniqueness of positive solutions to a linear transport equation in a metric space. Applied Mathematics Letters, 10:49–53, 11 1997. doi:10.1016/S0893-9659(97)00104-3.

[MBBVO03] Nikhil Mittal, Elena O Budrene, Michael P Brenner, and Alexander Van Oudenaarden. Motility of escherichia coli cells in clusters formed by chemotactic aggregation. Proceedings of the National Academy of Sciences, 100(23):13259–13263, 2003.

[MBC+12] Vincent A Martinez, Rut Besseling, Ottavio A Croze, Julien Tailleur, Mathias Reufer, Jana Schwarz-Linek, Laurence G Wilson, Martin A Bees, and Wilson CK Poon. Differential dynamic microscopy: A high-throughput method for characterizing the motility of microorganisms. Biophysical journal, 103(8):1637–1647, 2012.

[MBJPM85] Cedric Minkin, David J Bannon Jr, Selma Pokress, and Michael Melnick. Multiwell chamber chemotaxis assays: improved experimental design and data analysis. Journal of immunological methods, 78(2):307–321, 1985.

[MBKB18] Krithika Manohar, Bingni W. Brunton, J. Nathan Kutz, and Steven L. Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. IEEE Control Systems Magazine, 38(3):63–86, 2018. doi:10.1109/MCS.2018.2810460.

[McK67] Jr. McKean, H. P. Chapman-Enskog-Hilbert Expansion for a Class of Solutions of the Telegraph Equation. Journal of Mathematical Physics, 8(3):547–552, 03 1967. doi:10.1063/1.1705230.

[MISO76] Kayo Maeda, Yasuo Imae, Jun-Ichi Shioi, and FUMIO Oosawa. Effect of temperature on motility and chemotaxis of escherichia coli. Journal of bacteriology, 127(3):1039–1046, 1976.

[Mit00] Toby J Mitchell. An algorithm for the construction of "d-optimal" experimental designs. Technometrics, 42(1):48–54, 2000.

[MMS24] Ryan J Murphy, Oliver J Maclaren, and Matthew J Simpson. Implementing measurement error models with mechanistic mathematical models in a likelihood-based framework for estimation, identifiability analysis and prediction in the life sciences. Journal of the Royal Society Interface, 21(210):20230402, 2024.

[MRCW01] Gabriella Margaria, Eva Riccomagno, Michael J Chappell, and Henry P Wynn. Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. Mathematical biosciences, 174(1):1–26, 2001.

[MRR+53] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. The journal of chemical physics, 21(6):1087–1092, 1953.

[MS21] Oren Mangoubi and Aaron Smith. Mixing of hamiltonian monte carlo on strongly log-concave distributions: Continuous dynamics. The Annals of Applied Probability, 31(5):2019–2045, 2021.

[MST10] Stephen McDowall, Plamen Stefanov, and Alexandru Tamasan. Stability of the gauge equivalent classes in inverse stationary transport.

Inverse Problems, 26(2):025006, jan 2010. doi:10.1088/0266-5611/26/2/025006.

[MT14] Sebastien Motsch and Eitan Tadmor. Heterophilious dynamics enhances consensus. SIAM Review, 56(4):577–621, 2014. doi:10.1137/120901866.

[MT20] Per-Gunnar Martinsson and Joel A. Tropp. Randomized numerical linear algebra: Foundations and algorithms. Acta Numerica, 29:403–572, 2020. doi:10.1017/S0962492920000021.

[MXPW11] Hongyu Miao, Xiaohua Xia, Alan S Perelson, and Hulin Wu. On identifiability of nonlinear ode models and applications in viral dynamics. SIAM review, 53(1):3–39, 2011.

[Nak93] Shin-ichi Nakagiri. Review of japanese work of the last ten years on identifiability in distributed parameter systems. Inverse Problems, 9(2):143, apr 1993. doi:10.1088/0266-5611/9/2/001.

[Nak97] Shin-ichi Nakagiri. Regional identifiability of spatially-varying parameters in distributed parameter systems of parabolic type. IFAC Proceedings Volumes, 30(11):353–358, 1997.

[NBCS97] Martin A Nowak, Maarten C Boerlijst, Jonathan Cooke, and John Maynard Smith. Evolution of genetic redundancy. Nature, 388(6638):167–171, 1997.

[NLS20] Kit Newton, Qin Li, and Andrew M Stuart. Diffusive optical tomography in the bayesian framework. Multiscale Modeling & Simulation, 18(2):589–611, 2020.

[NP08] Andreas Neubauer and Hanna K Pikkarainen. Convergence results for the bayesian inversion theory. Journal of Inverse and Ill-posed Problems, 16:601–613, 10 2008. doi:10.1515/JIIP.2008.032.

[NPVW19] Ira Neitzel, Konstantin Pieper, Boris Vexler, and Daniel Walter. A sparse control approach to optimal sensor placement in pde-constrained parameter estimation problems. Numerische Mathematik, 143(4):943–984, 2019.

[OB00] Yuri Orlov and J. Bentsman. Adaptive distributed parameter systems identification with enforceable identifiability conditions and reduced-order spatial differentiation. Automatic Control, IEEE Transactions on, 45:203 – 216, 03 2000. doi:10.1109/9.839944.

[ODA88] Hans Othmer, S Dunbar, and W Alt. Models of dispersal in biological systems. Journal of mathematical biology, 26:263–98, 02 1988. doi:10.1007/BF00277392.

[OFD16] Nuno M. Oliveira, Kevin R. Foster, and William M. Durham. Single-cell twitching chemotaxis in developing biofilms. Proceedings of the National Academy of Sciences, 113(23):6532–6537, 2016. doi:10.1073/pnas.1600760113.

[OH98] Masood Otarod and John Happel. Identifiability of kinetic parameters of methanol synthesis in plug-flow tracing. American Institute of Chemical Engineers journal, 44(8):1897–1902, 1998.

[OH02] Hans Othmer and Thomas Hillen. The diffusion limit of transport equations ii: Chemotaxis equations. SIAM Journal of Applied Mathematics, 62:1222–1250, 04 2002. doi:10.1137/S0036139900382772.

[OHC24] Rafael Orozco, Felix J Herrmann, and Peng Chen. Probabilistic bayesian optimal experimental design using conditional normalizing flows. arXiv preprint arXiv:2402.18337, 2024.

[OL02] Katsutoshi Ozaki and Warren J Leonard. Cytokine and cytokine receptor pleiotropy and redundancy. Journal of Biological Chemistry, 277(33):29355–29358, 2002.

[OY05] Joost J Oppenheim and De Yang. Alarmins: chemotactic activators of immune responses. Current opinion in immunology, 17(4):359–365, 2005.

[Pat53] C.S. Patlak. Random walk with persistence and external bias: A mathematical contribution to the study of orientation of organisms. Bulletin of Mathematical Biophysics, 15:311–338, 1953. doi:10.1007/BF02476407.

[Paz12] Amnon Pazy. Semigroups of linear operators and applications to partial differential equations, volume 44. Springer Science & Business Media, 2012.

[PD16] Kernel Prieto and Oliver Dorn. Sparsity and level set regularization for diffuse optical tomography using a transport model in 2d. Inverse Problems, 33(1):014001, nov 2016. doi:10.1088/0266-5611/33/1/014001.

[Per04] Benoît Perthame. Mathematical tools for kinetic equations. Bulletin of the American Mathematical Society, 41(2):205–244, 2004.

[PH02] Kevin Painter and Thomas Hillen. Volume-filling and quorum-sensing in models for chemosensitive movement. Can. Appl. Math. Q., 10:501–543, 01 2002.

[PHA+17] Oliver Pohl, Marius Hintsche, Zahra Alirezaeizanjani, Maximilian Seyrich, Carsten Beta, and Holger Stark. Inferring the chemotactic strategy of p. putida and e. coli using modified kramers-moyal coefficients. PLoS computational biology, 13(1):e1005329, 2017.

[PHF22] Susanne Pieschner, Jan Hasenauer, and Christiane Fuchs. Identifiability analysis for models of the translation kinetics after mrna transfection. Journal of Mathematical Biology, 84(7):56, 2022.

[PKG+18] Saehong Park, Dylan Kato, Zach Gima, Reinhardt Klein, and Scott Moura. Optimal experimental design for parameterization of an electrochemical lithium-ion battery model. Journal of The Electrochemical Society, 165(7):A1309, 2018.

[PLCT11] Antoine Perasso, Béatrice Laroche, Yacine Chitour, and Suzanne Touzeau. Identifiability analysis of an epidemiological model in a structured population. Journal of Mathematical Analysis and Applications, 374(1):154–165, 2011. doi:10.1016/j.jmaa.2010.08.072.

[PMMA02] Joseph A Pedit, Randall B Marx, Cass T Miller, and Michael D Aitken. Quantitative analysis of experiments on bacterial chemotaxis to naphthalene. Biotechnology and bioengineering, 78(6):626–634, 2002.

[Poh78] Hannu Pohjanpalo. System identifiability based on the power series expansion of the solution. Mathematical biosciences, 41(1-2):21–33, 1978.

[Pol63] Boris Teodorovich Polyak. Gradient methods for minimizing functionals. Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki, 3(4):643–653, 1963.

[Pom73] Gerald C Pomraning. The equations of radiation hydrodynamics. 1973.

[Pon12] Augusto C Ponce. Selected problems on elliptic equations involving measures. arXiv preprint arXiv:1204.0668, 2012.

[PR16] A. Perasso and U. Razafison. Identifiability problem for recovering the mortality rate in an age-structured population dynamics model. Inverse Problems in Science and Engineering, 24(4):711–728, 2016. doi:10.1080/17415977.2015.1061522.

[PST18] Benoît Perthame, Weiran Sun, and Min Tang. The fractional diffusion limit of a kinetic model with biochemical pathway. Zeitschrift für angewandte Mathematik und Physik, 69(3):67, 2018.

[PTV16] Benoît Perthame, Min Tang, and Nicolas Vauchelet. Derivation of the bacterial run-and-tumble kinetic equation from a model with biochemical pathway. Journal of mathematical biology, 73:1161–1178, 2016.

[Puk06] Friedrich Pukelsheim. Optimal design of experiments. SIAM, 2006.

[PY18] Benoît Perthame and Shugo Yasuda. Stiff-response-induced instability for chemotactic bacteria and flux-limited keller–segel equation. Nonlinearity, 31(9):4065–4089, jul 2018. doi:10.1088/1361-6544/aac760.

[RBV23] Xabier Rey Barreiro and Alejandro F Villaverde. Benchmarking tools for a priori identifiability analysis. Bioinformatics, 39(2):btad065, 2023.

[RDMP16] Elizabeth G Ryan, Christopher C Drovandi, James M McGree, and Anthony N Pettitt. A review of modern computational algorithms for bayesian optimal design. International Statistical Review, 84(1):128–154, 2016.

[Red93] JN Reddy. An Introduction to the Finite Element Method. McGraw-Hill, 1993.

[Rei11] Sebastian Reich. A dynamical systems framework for intermittent data assimilation. BIT Numerical Mathematics, 51:235–249, 2011.

[Ren10] Kui Ren. Recent developments in numerical techniques for transport-based medical imaging methods. Communications in Computational Physics, 8(1):1–50, 2010. URL: https://global-sci.org/intro/article_detail/cicp/7562.html#.

[RFBCEB07] M. Rodríguez-Fernández, E. Balsa-Canto, J.A. Egea, and J.R. Banga. Identifiability and robust parameter estimation in food process modeling: Application to a drying model. Journal of Food Engineering, 83(3):374–383, 2007. doi:10.1016/j.jfoodeng.2007.03.023.

[RFIBS24] Tom Rainforth, Adam Foster, Desi R Ivanova, and Freddie Bickford Smith. Modern bayesian experimental design. Statistical Science, 39(1):100–114, 2024.

[RFKPS07] Maria Rodriguez-Fernandez, Sergei Kucherenko, Costas Pantelides, and Nilay Shah. Optimal experimental design based on global sensitivity analysis. In Computer Aided Chemical Engineering, volume 24, pages 63–68. Elsevier, 2007.

[RG22] Benjamin Rhodes and Michael Gutmann. Enhanced gradient-based MCMC in discrete spaces. arXiv preprint arXiv:2208.00040, 2022.

[RHL86] Mercedes Rivero-Hudec and Douglas A Lauffenburger. Quantification of bacterial chemotaxis by measurement of model parameters using the capillary assay. Biotechnology and bioengineering, 28(8):1178–1190, 1986.

[RKE22] Marissa Renardy, Denise Kirschner, and Marisa Eisenberg. Structural identifiability analysis of age-structured pde epidemic models. Journal of Mathematical Biology, 84(1):9, 2022.

[RKGF23] Konstantin Riedl, Timo Klock, Carina Geldhauser, and Massimo Fornasier. Gradient is all you need?, 2023. arXiv:2306.09778.

[RKM+09] Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics, 25(15):1923–1929, 2009.

[RKM+11] Andreas Raue, Clemens Kreutz, T Maiwald, Ursula Klingmüller, and Jens Timmer. Addressing parameter identifiability by model-based experimentation. IET systems biology, 5(2):120–130, 2011.

[RL86] George B. Rybicki and Alan P. Lightman. Radiative Processes in Astrophysics. Wiley, 1986.

[RL19] Angela Re and Paola Lecca. Determining structural parameter identifiability in biological dynamical models by analysing the statistical properties of the likelihood behaviour. In 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), pages 1–8. IEEE, 2019.

[RLS+19] Sabrina Rashid, Zhicheng Long, Shashank Singh, Maryam Kohram, Harsh Vashistha, Saket Navlakha, Hanna Salman, Zoltán N Oltvai, and Ziv Bar-Joseph. Adjustment in tumbling rates improves bacterial chemotaxis on obstacle-laden terrains. Proceedings of the National Academy of Sciences, 116(24):11770–11775, 2019.

[Rot71] Thomas J Rothenberg. Identification in parametric models. Econometrica: Journal of the Econometric Society, pages 577–591, 1971. doi:10.2307/1913267.

[RS08] Alejandro Ribes and Francis Schmitt. Linear inverse problems in imaging. IEEE Signal Processing Magazine, 25(4):84–99, 2008.

[RS10] Karl Rieger and Kurt Schlacher. A group-theoretic approach to parameter identifiability of pde systems. PAMM, 10(1):619–620, 2010.

[RS13] Mathias Rousset and Giovanni Samaey. Individual-based models for bacterial chemotaxis in the diffusion asymptotics. Mathematical Models and Methods in Applied Sciences, 23(11):2005–2037, 2013.

[RT96] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of langevin distributions and their discrete approximations. Bernoulli, 2(4):341–363, 1996. URL: http://www.jstor.org/stable/3318418.

[SABS18] Maximilian Seyrich, Zahra Alirezaeizanjani, Carsten Beta, and Holger Stark. Statistical parameter inference of bacterial swimming strategies. New Journal of Physics, 20(10):103033, oct 2018. doi:10.1088/1367-2630/aae72c.

[SAD03] Maria Pia Saccomani, Stefania Audoly, and Leontina D'Angiò. Parameter identifiability of nonlinear systems: the role of initial conditions. Automatica, 39(4):619–632, 2003.

[SBVM20] Matthew J Simpson, Ruth E Baker, Sean T Vittadello, and Oliver J Maclaren. Practical parameter identifiability for spatio-temporal models of

cell invasion. Journal of the Royal Society Interface, 17(164):20200055, 2020.

[SCB+10] Jonathan Saragosti, Vincent Calvez, Nikolaos Bournaveas, Axel Buguin, Pascal Silberzan, and Benoît Perthame. Mathematical description of bacterial traveling pulses. PLOS Computational Biology, 6(8):1–12, 08 2010. doi:10.1371/journal.pcbi.1000890.

[SCB+11] J. Saragosti, V. Calvez, N. Bournaveas, B. Perthame, A. Buguin, and P. Silberzan. Directional persistence of chemotactic bacteria in a traveling concentration wave. Proceedings of the National Academy of Sciences, 108(39):16235–16240, 2011. doi:10.1073/pnas.1101996108.

[SCF+19] Mehdi Salek, Francesco Carrara, Vicente Fernandez, Jeffrey Guasto, and Roman Stocker. Bacterial chemotaxis in a microfluidic t-maze reveals strong phenotypic heterogeneity in chemotactic sensitivity. Nature Communications, 10, 04 2019. doi:10.1038/s41467-019-09521-2.

[SCM19] Shyam Srinivasan, William R Cluett, and Radhakrishnan Mahadevan. A scalable method for parameter identification in kinetic models of metabolism using steady-state data. Bioinformatics, 35(24):5216–5225, 2019.

[SF97] Gilbert Strang and George Fix. An analysis of the finite element method second edition. 1997.

[SI19] Timothy E Saunders and Philip W Ingham. Open questions: how to get developmental biology into shape? BMC biology, 17:1–3, 2019.

[SK13] Eric Sonnendrücker and K Kormann. Numerical methods for vlasov equations. Lecture notes, 107:108, 2013.

[SLAJ+16] Jana Schwarz-Linek, Jochen Arlt, Alys Jepson, Angela Dawson, Teun Vissers, Dario Miroli, Teuta Pilizota, Vincent Martinez, and Wilson Poon. Escherichia coli as a model active colloid: A practical introduction. Colloids and Surfaces B: Biointerfaces, 137:2–16, 01 2016. doi:10.1016/j.colsurfb.2015.07.048.

[SLS17] Matthew J Simpson, Kai-Yin Lo, and Yung-Shin Sun. Quantifying the roles of random motility and directed motility using advection-diffusion theory for a 3t3 fibroblast cell migration assay stimulated with an electric field. BMC Systems Biology, 11:1–9, 2017.

[SM95] Zhenming Shun and Peter McCullagh. Laplace approximation of high dimensional integrals. Journal of the Royal Statistical Society Series B: Statistical Methodology, 57(4):749–760, 1995.

[SM23] Matthew J. Simpson and Oliver J. Maclaren. Profile-wise analysis: A profile likelihood-based workflow for identifiability analysis, estimation, and prediction with mechanistic mathematical models. PLOS Computational Biology, 19(9):1–31, 09 2023. doi:10.1371/journal.pcbi.1011515.

[SMM24] Matthew J. Simpson, Ryan J. Murphy, and Oliver J. Maclaren. Modelling count data with partial differential equation models in biology. Journal of Theoretical Biology, 580:111732, 2024. doi:10.1016/j.jtbi.2024.111732.

[SO08] Rajveer Singh and Mira S. Olson. Application of Bacterial Swimming and Chemotaxis for Enhanced Bioremediation, pages 149–172. Springer Netherlands, Dordrecht, 2008. doi:10.1007/978-1-4020-8786-8_7.

[Spi75] M. Spivak. A Comprehensive Introduction to Differential Geometry. Number Bd. 4 in A Comprehensive Introduction to Differential Geometry. Publish or Perish, Incorporated, 1975. URL: https://books.google.de/books?id=clbvAAAAMAAJ.

[SSW13] Jean-Marie Swiecicki, Olesksii Sliusarenko, and Douglas B. Weibel. From swimming to swarming: Escherichia coli cell motility in two-dimensions. Integrative Biology, 5(12):1490–1494, 10 2013. doi:10.1039/c3ib40130h.

[ST09] Plamen Stefanov and Alexandru Tamasan. Uniqueness and non-uniqueness in inverse radiative transfer. Proceedings of the American Mathematical Society, 137(7):2335–2344, 2009.

[ST16] Weiran Sun and Min Tang. Macroscopic limits of pathway-based kinetic models for e. coli chemotaxis in large gradient environments. Multiscale Modeling & Simulation, 15, 05 2016. doi:10.1137/16M1074011.

[ST18] Maria Pia Saccomani and Karl Thomaseth. The union between structural and practical identifiability makes strength in reducing oncological model complexity: a case study. Complexity, 2018:1–10, 2018.

[Ste98] GW Stewart. Afternotes goes to graduate school: lectures on advanced numerical analysis. SIAM, Philadelphia, PA, 1998.

[Ste00a] Angela Stevens. The derivation of chemotaxis equations as limit dynamics of moderately interacting stochastic many-particle systems. SIAM Journal on Applied Mathematics, 61(1):183–212, 2000. doi:10.1137/S0036139998342065.

[Ste00b] Angela Stevens. A stochastic cellular automaton modeling gliding and aggregation of myxobacteria. SIAM Journal on Applied Mathematics, 61(1):172–182, 2000. URL: http://www.jstor.org/stable/3061864.

[STLH21] Danny Smyl, Tyler N. Tallman, Dong Liu, and Andreas Hauptmann. An efficient quasi-newton method for nonlinear inverse problems via learned singular values. IEEE Signal Processing Letters, 28:748–752, 2021. doi:10.1109/LSP.2021.3063622.

[Str04] John C Strikwerda. Finite difference schemes and partial differential equations. SIAM, 2004.

[Stu10] A. M. Stuart. Inverse problems: A bayesian perspective. Acta Numerica, 19:451–559, 2010. doi:10.1017/S0962492910000061.

[STY14] Guangwei Si, Min Tang, and Xu Yang. A pathway-based mean-field model for e. coli chemotaxis: Mathematical derivation and its hyperbolic and parabolic limits. Multiscale Modeling & Simulation, 12:907–926, 06 2014. doi:10.1137/130944199.

[SU03] Plamen Stefanov and Gunther Uhlmann. Optical tomography in two dimensions. Methods and Applicaitons of Analysis, (1):001–010, March 2003.

[Suz83] Takashi Suzuki. Uniqueness and nonuniqueness in an inverse problem for the parabolic equation. Journal of Differential Equations, 47(2):296–316, 1983.

[SWZ23] Björn Sprungk, Simon Weissmann, and Jakob Zech. Metropolis-adjusted interacting particle sampling. arXiv preprint arXiv:2312.13889, 2023.

[SZ22] Plamen Stefanov and Yimin Zhong. Inverse boundary problem for the two photon absorption transport equation. SIAM Journal on Mathematical Analysis, 54(3):2753–2767, 2022.

[TA77] Andrey N Tikhonov and Vasiliy Y Arsenin. Solutions of ill-posed problems. New York, pages 1–30, 1977.

[TGCM16] Necibe Tuncer, Hayriye Gulbudak, Vincent L Cannataro, and Maia Martcheva. Structural and practical identifiability issues of immuno-epidemiological vector–host models with application to rift valley fever. Bulletin of mathematical biology, 78:1796–1827, 2016.

[Tho19] Godfrey H Thomson. The proof or disproof of the existence of general ability. British Journal of Psychology, 9(3):321, 1919.

[TKL18] Thomas Trunk, Hawzeen S. Khalil, and Jack C. Leo. Bacterial autoaggregation. AIMS Microbiology, 4(1):140–164, 2018. doi:10.3934/microbiol.2018.1.140.

[TKLRC14] Jake Taylor-King, E Loon, Gabriel Rosser, and Stephen Chapman. From birds to bacteria: Generalised velocity jump processes with resting states. Bulletin of mathematical biology, 77:1213–1236, 07 2014. doi:10.1007/s11538-015-0083-7.

[TMPA08] Marcus J Tindall, Philip K Maini, Steven L Porter, and Judith P Armitage. Overview of mathematical approaches used to model bacterial chemotaxis ii: bacterial populations. Bulletin of mathematical biology, 70:1570–1607, 2008.

[Tos06] Giuseppe Toscani. Kinetic models of opinion formation. Commun. Math. Sci., 4:481–496, 09 2006. doi:10.4310/CMS.2006.v4.n3.a1.

[TPM+08] Marcus J Tindall, SL Porter, PK Maini, G Gaglia, and Judith P Armitage. Overview of mathematical approaches used to model bacterial chemotaxis i: the single cell. Bulletin of mathematical biology, 70:1525–1569, 2008.

[TT15] Alex Townsend and Lloyd N Trefethen. Continuous analogues of matrix factorizations. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 471(2173):20140585, 2015.

[Tu13] Yuhai Tu. Quantitative modeling of bacterial chemotaxis: signal amplification and accurate adaptation. Annual review of biophysics, 42(1):337–359, 2013.

[TW82] CC Travis and LW White. Parameter identifiability for partial differential equations. Technical report, Oak Ridge National Lab., TN (USA); Oklahoma Univ., Norman (USA). Dept. of . . . , 1982.

[TW85] CC Travis and LW White. Parameter identification of distributed parameter systems. Mathematical Biosciences, 77(1):341–352, 1985. doi:10.1016/0025-5564(85)90105-1.

[TZL88] R. T. Tranquillo, S. H. Zigmond, and D. A. Lauffenburger. Measurement of the chemotaxis coefficient for human neutrophils in the under-agarose migration assay. Cell Motility, 11(1):1–15, 1988. doi:10.1002/cm.970110102.

[Vae24] Urbain Vaes. Sharp propagation of chaos for the ensemble langevin sampler. Journal of the London Mathematical Society, 110(5):e13008, 2024.

[Vau10] Nicolas Vauchelet. Numerical simulation of a kinetic model for chemotaxis. Kinetic and Related Models, 3(3):501–528, 2010.

[VC99] Paolo Vicini and Claudio Cobelli. A priori identifiability of distributed models of blood–tissue exchange. Annals of biomedical engineering, 27:200–207, 1999.

[VDVJB06] Nathalie Verdiere, Lilianne Denis-Vidal, and Ghislaine Joly-Blanchard. Identifiability of a pollution source: The distributed model and the semi-discretized differential model. In 2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control, pages 3318–3323. IEEE, 2006.

[Vla68] Anatoliĭ Aleksandrovich Vlasov. The vibrational properties of an electron gas. Soviet Physics Uspekhi, 10(6):721, 1968.

[VMZDV20] Nathalie Verdière, David Manceau, Shousheng Zhu, and Lilianne Denis-Vidal. Inverse problem for a coupling model of reaction-diffusion and ordinary differential equations systems. application to an epidemiological model. Applied Mathematics and Computation, 375:125067, 2020. doi:10.1016/j.amc.2020.125067.

[Vog02] Curtis R Vogel. Computational methods for inverse problems. SIAM, 2002.

[VRWL89] Sandor Vajda, Herschel Rabitz, Eric Walter, and Yves Lecourtier. Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. Chemical Engineering Communications, 83(1):191–219, 1989.

[W+14] David P Woodruff et al. Sketching as a tool for numerical linear algebra. Foundations and Trends® in Theoretical Computer Science, 10(1–2):1–157, 2014.

[WA04] George H Wadhams and Judith P Armitage. Making sense of it all: bacterial chemotaxis. Nature reviews Molecular cell biology, 5(12):1024–1037, 2004.

[Wal19] Daniel Walter. On sparse sensor placement for parameter identification problems with partial differential equations. PhD thesis, Technische Universität München, 2019.

[WB22] Navish Wadhwa and Howard C Berg. Bacterial motility: machinery and mechanisms. Nature reviews microbiology, 20(3):161–173, 2022.

[WBJK04] Eric Walter, Isabelle Braems, Luc Jaulin, and Michel Kieffer. Guaranteed numerical computation as an alternative to computer algebra for testing models for identifiability. In Numerical Software with Result Verification: International Dagstuhl Seminar, Dagstuhl Castle, Germany, January 19-24, 2003. Revised Papers, pages 124–131. Springer, 2004.

[WECW97] Mark T Widman, David Emerson, Chichia C Chiu, and R Mark Worden. Modeling microbial chemotaxis in a diffusion gradient chamber. Biotechnology and bioengineering, 55(1):191–205, 1997.

[Wei63] Steven Weinberg. Quasiparticles and the born series. Physical Review, 131(1):440, 1963.

[WHR+21] Franz-Georg Wieland, Adrian L Hauber, Marcus Rosenblatt, Christian Tönsing, and Jens Timmer. On structural and practical identifiability. Current Opinion in Systems Biology, 25:60–69, 2021.

[WL82] Eric Walter and Yves Lecourtier. Global approaches to identifiability testing for linear and nonlinear state space models. Mathematics and Computers in Simulation, 24(6):472–482, 1982. doi:10.1016/0378-4754(82)90645-0.

[WP96] Eric Walter and Luc Pronzato. On the identifiability and distinguishability of nonlinear parametric models. Mathematics and computers in simulation, 42(2-3):125–134, 1996. doi:10.1016/0378-4754(95)00123-9.

[WP01] Henry P Wynn and Neil Parkin. Sensitivity analysis and identifiability for differential equation models. In Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No. 01CH37228), volume 4, pages 3116–3121. IEEE, 2001.

[WR22] Stephen J. Wright and Benjamin Recht. Optimization for Data Analysis. Cambridge University Press, 2022. doi:10.1017/9781009004282.

[WZMP08] Hulin Wu, Haihong Zhu, Hongyu Miao, and Alan S Perelson. Parameter identifiability and estimation of hiv/aids dynamic models. Bulletin of mathematical biology, 70:785–799, 2008.

[XM03] Xiaohua Xia and Claude H Moog. Identifiability of nonlinear systems with application to hiv/aids models. IEEE transactions on automatic control, 48(2):330–336, 2003.

[XO09] Chuan Xue and Hans G. Othmer. Multiscale models of taxis-driven patterning in bacterial populations. SIAM Journal on Applied Mathematics, 70(1):133–167, 2009. doi:10.1137/070711505.

[XT21] Xiaoru Xue and Min Tang. Individual based models exhibiting lévy-flight type movement induced by intracellular noise. Journal of Mathematical Biology, 83, 09 2021. doi:10.1007/s00285-021-01651-w.

[Yas17] Shugo Yasuda. Monte carlo simulation for kinetic chemotaxis model: An application to the traveling population wave. Journal of Computational Physics, 330:1022–1042, Feb 2017. doi:10.1016/j.jcp.2016.10.066.

[YHZZ21] Siyuan Yang, Mingji Huang, Yongfeng Zhao, and H. P. Zhang. Controlling cell motion and microscale flow with polarized light fields. Phys. Rev. Lett., 126:058001, Feb 2021. doi:10.1103/PhysRevLett.126.058001.

[ZBFS10] He-Peng Zhang, Avraham Be'er, E-L Florin, and Harry L Swinney. Collective motion and density fluctuations in bacterial colonies. Proceedings of the National Academy of Sciences, 107(31):13626–13630, 2010.

[ZK20] Artur O Zaporozhets and Vladyslav V Khaidurov. Mathematical models of inverse problems for finding the main characteristics of air pollution sources. Water, Air, & Soil Pollution, 231(12):563, 2020.

[ZK23] Tatiana Zvonareva and Olga Krivorotko. Identifiability analysis for source problem of quasi-hyperbolic equation. In 2023 5th International Conference on Problems of Cybernetics and Informatics (PCI), pages 1–4. IEEE, 2023.

[ZKZ+24] Yongfeng Zhao, Christina Kurzthaler, Nan Zhou, Jana Schwarz-Linek, Clemence Devailly, Jochen Arlt, Jian-Dong Huang, Wilson C. K. Poon, Thomas Franosch, Vincent A. Martinez, and Julien Tailleur. Quantitative characterization of run-and-tumble statistics in bulk bacterial suspensions. Phys. Rev. E, 109:014612, Jan 2024. doi:10.1103/PhysRevE.109.014612.

[ZLCZ15] Chao Zhang, Qiang Liao, Rong Chen, and Xun Zhu. Locomotion of bacteria in liquid flow and the boundary layer effect on bacterial attachment. Biochemical and Biophysical Research Communications, 461(4):671–676, 2015. doi:10.1016/j.bbrc.2015.04.089.

[ZVDVK18] Shousheng Zhu, Nathalie Verdière, Lilianne Denis-Vidal, and Djalil Kateb. Identifiability analysis and parameter estimation of a chikungunya model in a spatially continuous domain. Ecological Complexity, 34:80–88, 2018. doi:10.1016/j.ecocom.2017.12.004.

[ZWWC+14] Jiang Zhuang, Guopeng Wei, Rika Wright Carlsen, Matthew R Edwards, Radu Marculescu, Paul Bogdan, and Metin Sitti. Analytical modeling and experimental characterization of chemotaxis in serratia marcescens. Physical Review E, 89(5):052704, 2014.

[ZZ19] Hongkai Zhao and Yimin Zhong. Instability of an inverse problem for the stationary radiative transport near the diffusion limit. SIAM Journal on Mathematical Analysis, 51(5):3750–3768, 2019. doi:10.1137/18M1222582.

[ZZZ21] Lan Zeng, Zhifei Zhang, and Ruizhao Zi. Suppression of blow-up in patlak-keller-segel-navier-stokes system via the couette flow. Journal of Functional Analysis, 280(10):108967, 2021.