RECONSTRUCTING THE KINETIC CHEMOTAXIS KERNEL USING MACROSCOPIC DATA: WELL-POSEDNESS AND ILL-POSEDNESS*

KATHRIN HELLMUTH†, CHRISTIAN KLINGENBERG†, QIN LI‡, AND MIN TANG§

Abstract. Bacterial motion is guided by external stimuli (chemotaxis), and the motion described on the mesoscopic scale is uniquely determined by a parameter K that models velocity change response from the bacteria. This parameter is termed chemotaxis kernel. In a practical setting, experimental data was collected to infer this kernel. In this article, a PDE-constrained optimization framework is deployed to perform this reconstruction using velocity-averaged, localized data taken in the interior of the domain. The problem can be well-posed or ill-posed depending on the data preparation and the experimental setup. In particular, we propose one specific design that guarantees numerical reconstructability and local convergence. This design is adapted to the discretization of K in space and decouples the reconstruction of local values of K into smaller cell problems, opening up parallelization opportunities. Numerical evidences support the theoretical findings.

Key words. mathematical biology, kinetic chemotaxis model, parameter reconstruction, macroscopic data, PDE-constrained optimization, well- and ill-posedness, inverse problem

MSC codes.
35R30; 65M32; 92C17; 49M41; 49K40

1. Introduction.

Kinetic chemotaxis equation is one of the classical equations that describes the collective behavior of bacteria motion. Presented on the phase space, the equation describes the "run-and-tumble" bacteria motion [3, 19, 39, 40]

$$\partial_t f + v \cdot \nabla_x f = \mathcal{K}(f) := \int_V K(x,v,v') f(x,t,v') - K(x,v',v) f(x,t,v) \, \mathrm{d}v', \tag{1.1}$$

$$f(t=0,x,v) = \phi(x,v)\,. \tag{1.2}$$

The solution $f(t,x,v)$ represents the density of bacteria at any given time t for any location x moving with velocity v. The two terms describe different aspects of the motion. The $v\cdot\nabla_x f$ term characterizes the "run"-part: bacteria move in a straight line with velocity v, and the terms on the right characterize the "tumble"-part: bacteria change from having velocity v' to v using the transitional rate $K(x,v,v') \geq 0$. This transition rate thus is termed the tumbling kernel. Initial data is given at $t = 0$ and is denoted by $\phi(x,v)$. The equation contains phase-space information, and thus compared to the macroscopic models, such as the Keller Segel model, it offers more details and has the greater potential to capture the fine motion of the bacteria. Indeed, it is observed that the dynamics predicted by the model is in high agreement with real measurements, see [7, 17, 48, 47].

*Submitted to the editors DATE.

Funding: K.H. acknowledges support by the German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes) and the Marianne-Plehn-Program. Q.L. is partially supported by Vice Chancellor for Research and Graduate Education, DMS-2308440 and ONR-N000142112140. M.T. is partially supported by the Strategic Priority Research Program of Chinese Academy of Sciences, XDA25010401 and NSFC12031013.

†Department of Mathematics, University of Würzburg, Würzburg, Germany (kathrin.hellmuth@uni-wuerzburg.de, https://www.mathematik.uni-wuerzburg.de/en/fluidmechanics/team/kathrin-hellmuth/; klingen@mathematik.uni-wuerzburg.de, https://www.mathematik.uni-wuerzburg.de/en/fluidmechanics/team/christian-klingenberg/).

‡Department of Mathematics, University of Wisconsin-Madison, Madison, WI, USA(qinli@math.wisc.edu, https://sites.google.com/view/qinlimadison/home).

§School of Mathematics and Institute of Natural Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai, China(tangmin@sjtu.edu.cn, https://ins.sjtu.edu.cn/people/mtang/).

This manuscript is for review purposes only.

It is noteworthy that these comparisons are conducted in the forward-simulation setting. Guesses are made about parameters, and simulations are run to be compared with experimental measurements. To fully reveal the bacteria's motion and its interaction with the environment, inverse perspectives have to be taken. This is to take measurements to infer K. The data can be collected at the individual level or the population level: biophysicists can use a high-resolution camera and trace each single bacterium for a long time to obtain single particle trajectory information, or take photos and record the density changes on a cell cultural dish. Such data should be used to unveil the true interaction between particles [35].

In this article, we frame this problem into a finite dimensional PDE-constrained optimization and study the unique and stable reconstructability of the kernel. In particular, we study different types of initial condition and measurement schemes and show that different experimental setups provide different stability of the reconstruction.

As more physics models derived from first-principles get deployed in applications, kinetic models are becoming more important in various scientific domains, see modeling of neutrons [14], photons or electrons [45] and rarefied gas [10]. The applications on biological and social science have also been put forward in [39] for cell motion, in [52] for animal (birds) migration or in [1, 9, 13, 38, 54] for opinion formation. In most of these models, parameters are included to characterize the interactions among agents or those between agents and the media. It is typical that these interactions cannot be measured directly, and it prompts the use of inverse solvers.

The most prominent application of inverse problem within the domain of kinetic systems is the optical tomography emerged from medical imaging, where non-intrusive boundary data is deployed to map out the optical properties of the interior. Mathematically the technique called the singular decomposition is deployed to conduct the inversion [6, 12, 33, 36, 51], and these studies have their numerical counterparts in [5, 11, 16, 43, 44], just to mention a few references.

Back to our current model, we notice that tracing the trajectory of every single bacterium is much more difficult than measuring the evolution of the macroscopic density [30, 57], so we are tasked to unveil the interaction between bacteria and the environment using the density measurement. A series of new results by biophysicists [32, 58] studies this experimental setting for a similar kinetic model and exhibits significance for practitioners. Compared with classical inverse problem originated from optical tomography, we encounter some new mathematical challenges. In particular, in our setup, our measurements are taken in the interior of the domain instead of on the boundary, and interior data is richer than boundary measurements. Meanwhile, our data is velocity independent, as compared to that in optical tomography that contains velocity information, so we also lose some richness in data.

In [27] the authors examined the theoretical aspect of this reconstruction problem. It was shown that trading off the microscopic information for the interior data still gives us sufficient information to recover the transition kernel, but the experiments need to be carefully crafted. In this theoretical work we assumed that the transition kernel is an unknown function, and thus an infinitely dimensional object, and the available data is the full map (from initial condition to density for all time and space), and thus an infinite dimensional object as well. This infinite-to-infinite setup is hard to be implemented in a practical setting, rendering the theoretical results only a guidance for direct use. The current paper can be seen as the practical counterpart of [27]. In particular, our goal is to study the same question on the discrete level: when measurement data are finite in size, and the to-be-reconstructed transition kernel is

This manuscript is for review purposes only.

also represented by a finite dimensional vector, can one still successfully recover the unknowns?

It turns out that the numerical issue is significantly more convoluted. In particular, when the dimension of K, the transition kernel, changes from infinite to finite, the amount of data needed to recover this parameter is expected to be reduced. The way of the reduction, however, is not clear. We will present below two different scenarios to argue:

- when data is prepared well, a stable reconstruction is expected;
- when the data "degenerates," it loses information, and the reconstruction does not hold.

Such coexistence of well-posedness and ill-posedness are presented respectively in two subsections of Section 3. Then in Section 5 we present the numerical evidence to showcase the theoretical prediction.

It should be noted that it is well within anticipation that different data preparation gives different conditioning for parameter reconstruction. This further prompts the study of experimental design. In the context of reconstructing the transition kernel in the chemotaxis equation, in Section 4 we will design a particular experimental setup that guarantees a unique reconstruction. This verifies existence of the situation of data being well-prepared.

We should further mention that reconstructing parameters for bacterial motion using the inversion perspective is not entirely new. Until recently, existing literature followed two different approaches: the first involves the utilization of statistical information at the individual level to extrapolate the microscopic transition kernel [41, 49], whereas the second entails employing density data at a macroscopic scale to reconstruct certain parameters associated with a macroscopic model through an optimization framework [23, 24, 46, 55]. To our knowledge, these available studies focus on a preset low-dimensional set of unknowns. The idea to infer parameters of kinetic descriptions from macroscopic type data emerged more recently [27, 32, 58]. The viewpoint we take in the current article significantly differs from those in the existing literature: Similar as was done in [15, 22] for a macroscopic model, we also recover the discretized version of the kinetic parameter. This brings more flexibility in application, at the cost of potentially high dimension of the unknown parameter. In contrast to existing results, our focus lies on the study of identifiability of the parameter in the proposed optimization setting, and thus its well- and ill-conditioning. Noise would introduce an additional layer of parameter uncertainty that we specifically seek to exclude from this stage of analysis. Numerical examples are thus presented in a noise-free and non regularized manner. This allows investigation of structural identifiability as well as suitability of specific experimental set ups to generate informative data for reconstruction in the sense of practical identifiability.

2. Framing a PDE-constrained optimization problem.

The problem is framed as a PDE-constrained optimization, which is to reconstruct K that fits data as much as possible, conditioned on the fact that the kinetic chemotaxis model is satisfied.

We reduce the dimension of the original kinetic chemotaxis model (1.1)–(1.2) for $t > 0$ from $(x, v) \in \mathbb{R}^3 \times \mathbb{S}^2$ to $(x, v) \in \mathbb{R}^1 \times \{\pm 1\}$ [24, 48, 47], i.e. the bacteria either moves to the left or to the right, and x is 1D in space. This simple setting reflects how experiments are conducted in the labs: bacteria are cultured in a tube, and the motion is one-dimensional. More details will be discussed in the subsequent part.

In a numerical setting, we first represent K as a finite dimensional parameter:

$$K(x,v,v') = \sum_{r=1}^{R} K_r(v,v')\mathbb{1}_{I_r}(x)\,. \tag{2.1}$$

This means dividing the domain into $\mathbb{R}^1 = \cup_r I_r$ with $I_r = [a_{r-1}, a_r)$, for $r = 2, ..., R-1$, and $I_1 = (-\infty, a_1)$, $I_R = [a_{R-1}, \infty)$, we approximate the function $K(x,v,v')$ within the cell I_r by $K_r(v,v')$, constant in space. Since $V = \{\pm 1\}$, there are only two parameters: $K_r(1,-1)$ and $K_r(-1,1)$ for each cell, so in total there are $2R$ free values to represent K. Throughout the paper we abuse the notation and denote $K \in \mathbb{R}^{2R}$ as the unknown vector to be reconstructed, and denote:

$$K_r = [K_{r,1}, K_{r,2}] \quad \text{with} \quad K_{r,i} = K_r(v_i, v_i') \quad \text{and} \quad (v_i, v_i') = \left((-1)^{i+1}, (-1)^i\right) \tag{2.2}$$

for $i = 1, 2$. The dataset is also finite in size. In particular, we mathematically represent the measurement as a reading of the bacteria density using a test function $\mu_l \in L^1(\mathbb{R})$ for some l, so the measurement is:

$$M_l(K) = \int_{\mathbb{R}} \int_V f_K(x,T,v)\,\mathrm{d}v\; \mu_l(x)\,\mathrm{d}x, \qquad l = 1, ..., L\,, \tag{2.3}$$

where f_K denotes the solution to (1.1) with kernel K. In case μ_l is a characteristic function, this corresponds to the pixel reading of a photo.

For simplicity of the presentation, the the ground-truth kernel denoted by $K_\star$ is assumed to be of form (2.1) as well. Consideration of continuous in space ground truths would require additional approximation error estimates, as presented in [31] for a diffusion coefficient reconstruction in elliptic and parabolic equations, which would go beyond the scope of this article. Then the true data is:

$$y_l = M_l(K_\star), \qquad l = 1, ..., L\,. \tag{2.4}$$

Since K is represented by a finite dimensional vector, we expect the amount of data needed is also finite. Given the nonlinear nature of the problem, it is unclear $L = 2R$ leads to a unique reconstruction. One ought to dive in the intricate dependence on the form of $\{\mu_l\}_{l=1,...,L}$.

To conduct such inversion, we deploy a PDE-constrained optimization formulation. This is to minimize the square loss between the simulated data $M(K)$ and the data y:

$$\begin{aligned} \min_K \quad \mathcal{C}(K) = \min \frac{1}{2L}\sum_{l=1}^{L}\left(M_l(K) - y_l\right)^2 \\ \text{subject to} \quad (1.1), \text{ and } (1.2). \end{aligned} \tag{2.5}$$

Many algorithms can be deployed to solve this minimization problem, and we are particularly interested in the application of gradient-based solvers. The simple gradient descent method gives:

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

with a suitable step size $\eta_n \in \mathbb{R}_+$. It is a standard practice of calculus-of-variation to derive the partial differentiation against the (r,i)-th ($i = 1, 2$, $r = 1, \cdots, R$) entry in the gradient $\nabla_K \mathcal{C}$:

$$\frac{\partial \mathcal{C}}{\partial K_{r,i}} = \int_0^T \int_{I_r} f(t,x,v_i')\left(g(t,x,v_i') - g(t,x,v_i)\right)\mathrm{d}x\,\mathrm{d}t\,, \tag{2.7}$$

Detailed are placed in Appendix A. In the formulation, (v_i, v_i') is given in (2.2) and g is the adjoint state that solves the adjoint equation

$$-\partial_t g - v \cdot \nabla g = \tilde{\mathcal{K}}(g) := \int_V K(x, v', v)(g(x,t,v') - g(x,t,v))\,\mathrm{d}v', \tag{2.8}$$

$$g(x, t = T, v) = -\frac{1}{L}\sum_{l=1}^{L} \mu_l(x)\left(M_l(K) - y_l\right) . \tag{2.9}$$

The convergence of GD in (2.6) is guaranteed for a suitable step size if the objective function is convex. Denoting $H_K\mathcal{C}$ the Hessian function of the loss function, we need $H_K\mathcal{C} > 0$ at least in a small neighborhood around $K_\star$. In [56], a constant step size $\eta_n = \eta = \frac{2\lambda_{\min}}{\lambda_{\max}^2}$ is recommended with $\lambda_{\min}, \lambda_{\max}$ denoting the smallest and largest eigenvalues of $H_K\mathcal{C}(K_\star)$. More sophisticated methods include line search for the step size or higher order methods are also possible, see e.g. [44, 56].

To properly set up the problem, we make some general assumptions and fix some notations.

Assumption 2.1. We make assumptions to ensure the well-posedness of the forward problem in a feasible set, in particular:

- We will work locally in K, so we assume in a neighbourhood $\mathcal{U}_{K_\star}$ of $K_\star$, there is a constant C_K so that for all $K \in \mathcal{U}_{K_\star}$:

$$0 < \|K\|_\infty \le C_K . \tag{2.10}$$

- Assume the initial data ϕ be in the space $L^\infty_{+,c}(\mathbb{R} \times V)$ of non negative, compactly supported functions with essential bound

$$\|\phi\|_{L^\infty(\mathbb{R}\times V)} =: C_\phi .$$

- The test functions $\{\mu_l\}_{l=1}^L$ are supposed to be selected from the space $L^1(\mathbb{R})$ with uniform L^1 bound

$$\int_{\mathbb{R}} |\mu_l|\,\mathrm{d}x \le C_\mu, \quad l = 1, ..., L .$$

These assumptions are satisfied in a realistic setting. They allow us to operate f and g in the right spaces. In particular, we can establish existence of mild solutions and upper bounds for both the forward and adjoint solution, see Lemma B.1 and B.3 in Appendix B.

3. Well-posedness vs. ill-posedness. As many optimization algorithms are designed to produce minimizing sequences, we study well-posedness in the sense of Tikhonov.

DEFINITION 3.1 (Tikhonov well-posedness [53]). *A minimization problem is Tikhonov well-posed, if a unique minimum point exists towards which every minimizing sequence converges.*

The well-posedness of the inversion heavily depends on the data preparation. If a suitable experimental setting is arranged, the optimization problem is expected to provide local well-posedness around the ground-truth parameter $K_\star$, so the classical GD can reconstruct the ground-truth. However, if data becomes degenerate, we also expect ill-conditioning and the GD will find it hard to converge to the global minimum. We spell out the two scenarios in the two theorems below.

This manuscript is for review purposes only.

THEOREM 3.2. *Assume the Hessian matrix of the cost function is positive definite at $K_\star$ and let the remaining assumptions of Proposition 3.4 hold, then there exists a neighbourhood U of $K_\star$, in which the optimization problem* (2.5) *is Tikhonov well-posed. In particular, the gradient descent algorithm* (2.6) *with initial value $K_0 \in U$ converges.*

This theorem provides the well-posedness of the problem. To be specific, it spells out the sufficient condition for GD to find the global minimizer $K_\star$. The condition of the Hessian being positive definite at $K_\star$ may seem strong. In Section 4, we will carefully craft a setting for which we can ensure this to hold.

On contrary to the previous well-posedness discussion, we also provide a negative result below on ill-conditioning.

THEOREM 3.3. *Let $L = 2R$ and let Assumption 2.1 hold for all considered quantities. Consider a sequence $(\mu_1^{(m)})_m$ of test functions for the first measurement $M_1(K)$ for which one of the following scenarios holds:*

1. *$\mu_1^{(m)} \to \mu_2$ in L^1 as $m \to \infty$.*
2. *$(\mu_1^{(m)})_m$ and μ_2, as defined in* (3.12)*, are mollifications of singular point-measurements in measurement points $\{(x_1^{(m)})_m, x_2\}$ such that $x_1^{(m)} \to x_2$ as $m \to \infty$. Furthermore, let the assumptions of Proposition 3.10 hold.*

Then, as $m \to \infty$, i.e. as the measurement test functions become close in one of the above senses, strong convexity of the loss function decays, and the convergence of the gradient descent algorithm (2.6) *to $K_\star$ cannot be guaranteed. In scenario (2), this holds independently of the mollification parameter.*

The two theorems, to be proved in detail in Section 3.1 and 3.2 respectively, hold vast contrast to each other. The core difference between the two theorems is the data selection. The former guarantees the convexity of the objective function, and the latter shows degeneracy. The analysis comes down to evaluating the Hessian, a $2R \times 2R$ matrix:

$$H_K \mathcal{C}(K) = \frac{1}{L} \sum_{l=1}^{L} \left(\nabla_K M_l(K) \otimes \nabla_K M_l(K) + (M_l(K) - y_l) H_K M_l(K) \right). \tag{3.1}$$

It is a well-known fact [42] that a positive definite Hessian provides the strong convexity of the loss function, and is a sufficient criterion that permits the convergence in the parameter space. If $H_K\mathcal{C}(K_\star)$ is known to be positive and the Hessian matrix does not change much under small perturbation of K, then convexity of the cost function can be guaranteed in a small environment around $K_\star$. Such boundedness of perturbation in the Hessian is spelled out in Proposition 3.4, and Theorem 3.2 naturally follows.

Theorem 3.3 orients the opposite side. In particular, it examines the degeneracy when two data collection points get very close. The guiding principle for such degeneracy is that when two measurements can get too close, they offer no additional information. Mathematically, this amounts to rank deficiency of the Hessian (3.1), prompting the collapse of convexity in the landscape of the objective function. The closeness of two measurements can be quantified through different manners, and we specifically examine two types:

- the two test functions μ_1, μ_2 are close in L^1;
- the measurement locations are close: setting μ_1 and μ_2 as mollifiers from direct Dirac-δ centered at x_1 and x_2, then the closeness is quantified by $|x_1 - x_2|$.

corresponding to the two bullet points in Theorem 3.3. These two scenarios of deficient ranks are presented in Proposition 3.10 and 3.9 respectively.

3.1. Local well-posedness of the optimization problem. Generally speaking, it would not be easy to characterize the landscape of the distribution and thus hard to prescribe conditions for obtaining global convergence. However, suppose the data is prepared well enough so to guarantee the positive definiteness for the Hessian $H_K\mathcal{C}(K_\star)$ evaluated at the ground-truth $K_\star$, then the following results provide that in a small neighborhood of this ground-truth, positive-definiteness persists. Therefore, GD that starts within this neighborhood, finds the global minimum to (2.5). This gives us a local well-posedness.

This local behavior is characterized in the following proposition.

PROPOSITION 3.4. *Let Assumption 2.1 hold. Assume the Hessian $H_K\mathcal{C}(K_\star)$ is positive definite at $K_\star$, and that there is a uniform bound for the Hessian of the measurements in the neighborhood $\mathcal{U}_{K_\star}$ in the sense that $\|H_K M_l(K)(v,v')\|_F \le C_{H_K M}$ for all $l = 1,\ldots,L$ and $K \in \mathcal{U}_K$ in the Frobenius norm. Then there exists a (bounded) neighbourhood $U \subset \mathcal{U}_{K_\star}$ of $K_\star$, where $H_K\mathcal{C}(K)$ is positive definite for all $K \in U$. Moreover, the minimal eigenvalues $\lambda_{\min}(H_K\mathcal{C})$ satisfies*

$$|\lambda_{\min}(H_K\mathcal{C}(K_\star)) - \lambda_{\min}(H_K\mathcal{C}(K))| \le \|K_\star - K\|_\infty C', \tag{3.2}$$

where the constant C' depends on the measurement time T, R, and the bounds C_μ, C_ϕ, C_K in Assumption 2.1 and $C_{H_K M}$. As a consequence, the radius of U can be chosen as $\lambda_{\min}(H_K\mathcal{C}(K_\star))/C'$.

The proposition is hardly surprising. Essentially it suggests the Hessian term is Lipschitz continuous with respect to its argument. This is expected if the solution to the equation is somewhat smooth. Such strategy will be spelled out in detail in the proof. Now Theorem 3.2 is immediate.

Proof for Theorem 3.2. By Proposition 3.4, there exists a neighbourhood U of $K_\star$ in which the Hessian is positive definite, $H_K\mathcal{C}(K) > 0$ for all $K \in U$. Without loss of generality, we can assume that U is a convex set. By the strong convexity of $\mathcal{C}$ in U, the minimizer $K_\star \in U$ of $\mathcal{C}$ is unique and thus the finite dimension of the parameter space $K \in \mathbb{R}^{2R}$ guarantees Tikhonov well-posedness of the optimization problem (2.5) [20, Prop.3.1]. Convergence of GD follows from strong convexity of $\mathcal{C}$ in U. ∎

Now we give the proof for Proposition 3.4. It mostly relies on the matrix perturbation theory [29, Cor. 6.3.8] and continuity of equation (1.1) with respect to the parameter K.

Proof for Proposition 3.4. According to the matrix perturbation theory, the minimal eigenvalue is continuous with respect to a perturbation to the matrix, we have

$$
\begin{aligned}
|\lambda_{\min}(H_K\mathcal{C}(K_\star)) - \lambda_{\min}(H_K\mathcal{C}(K))| &\le \|H_K\mathcal{C}(K_\star) - H_K\mathcal{C}(K)\|_F \\
&\le \frac{1}{L}\sum_l \Bigg(\|(\nabla_K M_l \otimes \nabla_K M_l)(K_\star) - (\nabla_K M_l \otimes \nabla_K M_l)(K)\|_F \\
&\qquad + \|(M_l(K) - y_l) H_K M_l(K)\|_F \Bigg) \qquad (3.3)\\
&\le \frac{1}{L}\sum_l \Bigg(\|\nabla_K M_l(K_\star) - \nabla_K M_l(K)\|_F \left(\|\nabla_K M_l(K_\star)\|_F + \|\nabla_K M_l(K)\|_F\right) \\
&\qquad + |M_l(K) - y_l| \|H_K M_l(K)\|_F \Bigg)
\end{aligned}
$$

where we used the Hessian form (3.1), triangle inequality and sub-multiplicativity for Frobenius norms. To obtain the bound (3.2) now amounts to quantifying each term on the right hand side of (3.3) and bounding them by $\|K_\star - K\|_\infty$. This is respectively achieved in Lemmas 3.5, 3.7 and 3.8 that give controls to $M_l(K) - y_l$, $\|\nabla_K M_l(K)\|_F$ and $\|\nabla_K M_l(K_\star) - \nabla_K M_l(K)\|_F$. Putting these results together, we have:

$$
\begin{aligned}
&|\lambda_{\min}(H_K\mathcal{C}(K_\star)) - \lambda_{\min}(H_K\mathcal{C}(K))| \le \|H_K\mathcal{C}(K_\star) - H_K\mathcal{C}(K)\|_F \\
&\le 2\|K_\star - K\|_\infty C_\mu C_\phi e^{2C_K|V|T} \Bigg[8RC_\phi C_\mu e^{2|V|C_K T} T \left(|V|T^2 + \frac{1}{C_K}\left(\frac{e^{2C_K|V|T} - 1}{2C_K|V|} - T \right)\right) \\
&\qquad + |V|^2 T C_{H_K M} \Bigg] \\
&=: \|K_\star - K\|_\infty C'.
\end{aligned}
$$

The positive definiteness in a small neighborhood of $K_\star$ now follows. Finally, given $\|K_\star - K\|_\infty < \lambda_{\min}(H_K\mathcal{C}(K_\star))/C'$, the triangle inequality shows

$$
\lambda_{\min}(H_K\mathcal{C}(K)) \ge \lambda_{\min}(H_K\mathcal{C}(K_\star)) - |\lambda_{\min}(H_K\mathcal{C}(K_\star)) - \lambda_{\min}(H_K\mathcal{C}(K))| > 0.
$$

We note the form of C' is complicated but the dependence is spelled out in the following lemmas and summarized in the theorem statement. ∎

As can be seen from the proof, Proposition 3.4 strongly relies on the boundedness of the terms in (3.3). We present the estimates below.

LEMMA 3.5. *Let Assumptions 2.1 holds, then the measurement difference is upper bounded by:*

$$
|M_l(K) - y_l| \le |V| C_\mu \|(f_{K_\star} - f_K)(T)\|_{L^\infty(\mathbb{R}\times V)} \le \|K_\star - K\|_\infty 2|V|^2 C_\mu C_\phi T e^{2C_K|V|T}.
$$

Proof. Apply Lemma B.1 to the difference equation for $\bar{f} := f_{K_\star} - f_K$

$$
\partial_t \bar{f} + v\cdot\nabla_x \bar{f} = \mathcal{K}_K(\bar{f}) + \mathcal{K}_{(K_\star - K)}(f_{K_\star}) \qquad (3.4)
$$

with initial condition 0 and source $h = \mathcal{K}_{(K_\star - K)}(f_{K_\star}) \in L^1((0,T); L^\infty(\mathbb{R}\times V))$ by the regularity (B.1) of $f_{K_\star}$. This leads to

$$\operatorname*{ess\,sup}_{v,x} |\bar{f}|(x,t,v) \le \int_0^t e^{2|V|C_K(t-s)} \operatorname*{ess\,sup}_{v,x} |\mathcal{K}_{(K_\star - K)}(f_{K_\star})(s)| \, \mathrm{d}s$$
$$\le 2|V| \|K_\star - K\|_\infty e^{2|V|C_K t} C_\phi t, \tag{3.5}$$

where we used the estimate $\|f_{K_\star}(s)\|_{L^\infty(\mathbb{R}\times V)} \le e^{2|V|C_K s} \|\phi\|_{L^\infty(\mathbb{R}\times V)}$ from Lemma B.1 in the last step. ∎

To estimate the gradient $\nabla_K M_l(K)$ and its difference, we first recall the form in (2.7) with $\mathcal{C}$ changed to M_l here. Analogously, we can use the adjoint equation to explicitly represent the gradient:

LEMMA 3.6. *Let Assumption 2.1 hold. Denote by f_K the mild solution of* (1.1) *and by $g_l \in C^0\left([0,T]; L^\infty(V; L^1(\mathbb{R}))\right)$ the mild solution of*

$$-\partial_t g_l - v \cdot \nabla g_l = \tilde{\mathcal{K}}(g_l) := \int_V K(x,v',v)(g_l(x,t,v') - g_l(x,t,v)) \, dv', \tag{3.6}$$
$$g_l(t = T, x, v) = -\mu_l(x) \, .$$

Then

$$\frac{\partial M_l(K)}{\partial K_{r,i}} = \int_0^T \int_{I_r} f'(g_l' - g_l) \, dx \, dt \, , \tag{3.7}$$

where we used the abbreviated notation $h := h(t,x,v_i)$ and $h' := h(t,x,v_i')$ for $h = f, g_l$, with (v_i, v_i') defined as in (2.7).

We omit explicitly writing down the x,t dependence when it is not controversial. The proof for this lemma is the application of calculus-of-variation and will be omitted from here. We are now in the position to derive the estimates of the gradient norms.

LEMMA 3.7. *Under Assumption 2.1, the gradient is uniformly bounded*

$$\|\nabla_K M_l(K)\|_F \le \sqrt{2R} 2C_\phi C_\mu e^{2C_K|V|T} T, \qquad \textit{for all } K \in \mathcal{U}_K.$$

Proof. The Frobenius norm is bounded by the entries

$$\|\nabla M_l(K)\|_F \le \sqrt{2R} \max_{r,i} \left| \frac{\mathrm{d}M_l(K)}{\mathrm{d}K_{r,i}} \right| .$$

Representation (3.7) together with (B.2) then gives the bound

$$\left| \frac{\mathrm{d}M_l}{\mathrm{d}K_{r,i}} \right| \le 2C_\phi \int_0^T e^{2|V|C_K t} \max_v \left(\int_{\mathbb{R}} |g_l| \, \mathrm{d}x \right) \mathrm{d}t, \tag{3.8}$$

Application of lemma B.3 to $g = g_l, h = 0$ and $\psi = -\mu_l$ yields

$$\max_v \int_{\mathbb{R}} |g_l| \, \mathrm{d}x \ (t) \le \int_{\mathbb{R}} |-\mu_l(x)| \, \mathrm{d}x \ e^{2C_K|V|(T-t)} \le C_\mu e^{2C_K|V|(T-t)}, \tag{3.9}$$

which, when plugged into (3.8), gives

$$\left| \frac{\partial M_l}{\partial K_{r,i}} \right| \le 2C_\phi C_\mu e^{2C_K|V|T} T \, .$$ ∎

LEMMA 3.8. *In the setting of Theorem 3.2 and under Assumption 2.1, the gradient difference is uniformly bounded in $K \in \mathcal{U}_K$ by*

$$
\begin{aligned}
&\|\nabla M_l(K_\star) - \nabla M_l(K)\|_F \\
&\le \sqrt{2R}\|K_\star - K\|_\infty 2C_\phi C_\mu e^{2C_K|V|T}\left(|V|T^2 + \frac{1}{C_K}\left(\frac{e^{2C_K|V|T}-1}{2C_K|V|} - T\right)\right).
\end{aligned}
$$

Proof. Now consider the entries of $\nabla M_l(K_\star) - \nabla M_l(K)$ to show smallness of $\|\nabla M_l(K_\star) - \nabla M_l(K)\|_F$. Rewrite, using lemma 3.6 and (B.2)

$$
\begin{aligned}
\left|\frac{\partial M_l(K_\star)}{\partial K_{r,i}} - \frac{\partial M_l(K)}{\partial K_{r,i}}\right| &= \left|\int_0^T \int_{I_r} f_{K_\star}({g'}_{l,K_\star} - g_{l,K_\star}) - f_K({g'}_{l,K} - g_{l,K})\,\mathrm{d}x\,\mathrm{d}t\right| \\
&\le \int_0^T \|(f_{K_\star} - f_K)(t)\|_{L^\infty(\mathbb{R}\times V)} 2\max_v \int_{\mathbb{R}} |g_{l,K_\star}(t)|\,\mathrm{d}x\,\mathrm{d}t \\
&\quad + 2C_\phi \int_0^T e^{2|V|C_K t} \max_v \int_{\mathbb{R}} |(g_{l,K_\star} - g_{l,K})(t)|\,\mathrm{d}x\,\mathrm{d}t.
\end{aligned}
$$

The first summand can be bounded by (3.5) and (3.9). To estimate the second summand, apply Lemma B.3 to $\bar{g} \coloneqq g_{l,K_\star} - g_{l,K}$ with evolution equation

$$
\begin{aligned}
-\partial_t \bar{g} - v\cdot\nabla_x \bar{g} &= \tilde{\mathcal{K}}_{K_\star}(\bar{g}) + \tilde{\mathcal{K}}_{(K_\star - K)}(g_{l,K}), \\
\bar{g}(t=T) &= 0,
\end{aligned}
$$

and $h = \tilde{\mathcal{K}}_{(K_\star - K)}(g_{l,K}) \in L^1((0,T); L^\infty(V; L^1(\mathbb{R})))$ by the regularity (B.6) of $g_{l,K} \in C^0\left((0,T); L^\infty(V; L^1(\mathbb{R}))\right)$. This leads to

$$
\begin{aligned}
\max_v \int_{\mathbb{R}} |\bar{g}|\,\mathrm{d}x &\le e^{2|V|C_K(T-t)} \int_0^{T-t} \max_v \|\tilde{\mathcal{K}}_{(K_\star - K)}(g_{l,K})(T-s,v)\|_{L^1(\mathbb{R})}\,\mathrm{d}s \\
&\le 2|V|\|K_\star - K\|_\infty e^{2|V|C_K(T-t)} \int_0^{T-t} \max_v \|g_{l,K}(T-s,v)\|_{L^1(\mathbb{R})}\,\mathrm{d}s \\
&\le \|K_\star - K\|_\infty \frac{C_\mu}{C_K} e^{2|V|C_K(T-t)}(e^{2C_K|V|(T-t)} - 1),
\end{aligned}
$$

where we used (3.9) in the last line. In summary, one obtains

$$
\begin{aligned}
&\left|\frac{\partial M_l(K_\star)}{\partial K_{r,i}} - \frac{\mathrm{d}M_l(K)}{\mathrm{d}K_{r,i}}\right| \\
&\le \|K_\star - K\|_\infty \Bigg[\int_0^T 2|V|C_\phi t e^{2C_K|V|t}\cdot 2C_\mu e^{2C_K|V|(T-t)}\,\mathrm{d}t \\
&\qquad + 2C_\phi \int_0^T e^{2|V|C_K t}\frac{C_\mu}{C_K} e^{2C_K|V|(T-t)}(e^{2C_K|V|(T-t)} - 1)\,\mathrm{d}t\Bigg] \\
&\le \|K_\star - K\|_\infty 2C_\phi C_\mu e^{2C_K|V|T}\left(|V|T^2 + \frac{1}{C_K}\left(\frac{e^{2C_K|V|T}-1}{2C_K|V|} - T\right)\right). \qquad \square
\end{aligned}
$$

Together with the boundedness of the gradient (3.8), this shows that the first summands in (3.3) are Lipschitz continuous in K around $K_\star$ which concludes the proof of Proposition 3.4.

3.2. Ill-conditioning for close measurements. While the positive Hessian at $K_\star$ guarantees local convergence, such positive-definiteness will disappear when data are not prepared well. In particular, if $L = 2R$, meaning the number of measurements equals the number of parameters to be recovered, and that two measurements, $M_1(K)$ and $M_2(K)$ are close, we will show that the Hessian degenerates. Then strong convexity is lost, and the convergence to $K_\star$ is no longer guaranteed.

We will study how the Hessian degenerates in the two scenarios in Theorem 3.3. This comes down to examining the two terms in (3.1). Applying Lemma 3.5, we already see the second part in (3.1) is negligible when K is close to $K_\star$ and the rank structure of the Hessian is predominantly controlled by the first term. It is a summation of L rank 1 matrices $\nabla_K M_l(K) \otimes \nabla_K M_l(K)$. When two measurements (μ_1 and μ_2) get close, we will argue that $\nabla_K M_1(K)$ is almost parallel to $\nabla_K M_2(K)$, making the Hessian lacking at least one rank, and the strong convexity is lost. Mathematically, this means we need to show $\|\nabla_K M_1(K) - \nabla_K M_2(K)\|_2 \approx 0$ when $\mu_1 \approx \mu_2$.

Throughout the derivation, the following formula is important. Recalling (3.7), we have for every $r \in \{1, \cdots, R\}$ and $i \in \{1, 2\}$

$$\begin{aligned}\frac{\partial M_1(K)}{\partial K_{r,i}} - \frac{\partial M_2(K)}{\partial K_{r,i}} &= \int_0^T \int_{I_r} f'((g_1 - g_2)' - (g_1 - g_2))\,\mathrm{d}x\,\mathrm{d}t \\ &= \int_0^T \int_{I_r} f'(\bar{g}' - \bar{g})\,\mathrm{d}x\,\mathrm{d}t\,, \qquad (3.10)\end{aligned}$$

where $\bar{g} := g_1 - g_2$ solves (2.8) with final condition $\bar{g}(t = T, x, v) = \mu_2(x) - \mu_1(x)$. The two subsections below serve to quantify the smallness of (3.10) in terms of the smallness of $\mu_1(x) - \mu_2(x)$.

3.2.1. L^1 measurement closeness. The following proposition states the loss of strong convexity as $\mu_2 - \mu_1 \to 0$ in $L^1(\mathbb{R})$. In particular, the requirement of Proposition 3.4 that $H_K\mathcal{C}(K_\star)$ is positive definite is no longer satisfied, so local well-posedness of the optimization problem and thus the convergence of the algorithm can no longer be guaranteed.

PROPOSITION 3.9. *Let Assumption 2.1 hold. Then, as $\mu_1^{(m)} \xrightarrow{m\to\infty} \mu_2$ in $L^1(\mathbb{R})$, one eigenvalue of the Hessian $H_K\mathcal{C}(K_\star)$ vanishes.*

This proposition immediately allows us to prove scenario 1 in Theorem 3.3:

Proof of Theorem 3.3. Propositions 3.9 establishes one eigenvalue of $H_K\mathcal{C}(K_\star)$ vanishes as $m \to \infty$. This lack of positive definiteness and thus strong convexity of $\mathcal{C}$ around $K_\star$ means that it cannot be guaranteed that the minimizing sequences of $\mathcal{C}$ converge to $K_\star$. ∎

We now give the proof of the proposition.

Proof. As argued above, we show $\|\nabla_K M_1^{(m)}(K) - \nabla_K M_2(K)\|_2 \to 0$ as $m \to \infty$. Recall (3.10), we need to show:

$$\frac{\partial M_1^{(m)}(K)}{\partial K_{r,i}} - \frac{\partial M_2(K)}{\partial K_{r,i}} \xrightarrow{m\to\infty} 0 \quad \forall (r,i) \in \{1, \cdots, R\} \times \{1, 2\}\,. \qquad (3.11)$$

where $\bar{g} := g_1 - g_2$ solves (2.8) with final condition $\bar{g}(t = T, x, v) = \mu_2(x) - \mu_1^{(m)}(x)$. Application of Lemma B.3 gives

$$\|\bar{g}(t)\|_{L^\infty(V;L^1(\mathbb{R}))} \le e^{2C_K|V|(T-t)}\|\mu_2 - \mu_1^{(m)}\|_{L^1(\mathbb{R})}.$$

by independence of μ_1, μ_2 with respect to v. Plug the above into (3.10) and estimate f by (B.2) to obtain

$$\left|\frac{\partial(M_1^{(m)} - M_2)(K)}{\partial K_{r,i}}\right| \leq 2C_\phi \int_0^T e^{2C_K|V|t} \|\bar{g}(t)\|_{L^\infty(V;L^1(\mathbb{R}))} \,\mathrm{d}t$$
$$\leq 2C_\phi e^{2C_K|V|T} T \|\mu_2 - \mu_1^{(m)}\|_{L^1(\mathbb{R})}.$$

Since every entry (r,i) converges, the gradient difference vanishes $\|\nabla_K M_1^{(m)}(K) - \nabla_K M_2(K)\|_2 \to 0$ as $m \to \infty$.

We utilize this fact to show the degeneracy of the Hessian. Noting:

$$H_K\mathcal{C}(K_\star) = \underbrace{\left[\sum_{l=3}^{2R} \nabla M_l \otimes \nabla M_l + 2\nabla M_2 \otimes \nabla M_2\right]}_{A} + \underbrace{\left[\nabla M_1^{(m)} \otimes \nabla M_1^{(m)} - \nabla M_2 \otimes \nabla M_2\right]}_{B^{(m)}}.$$

It is straightforward that the rank of A is at most $2R-1$, so the j-th largest eigenvalue $\lambda_j(A) = 0$ vanishes for some j. Moreover, since $\|\nabla_K M_1^{(m)}(K) - \nabla_K M_2(K)\|_2 \to 0$, we have $\|B^{(m)}\|_F \to 0$. Using the continuity of the minimal eigenvalue with respect to a perturbation of the matrix, the j-th largest eigenvalue of $H_K\mathcal{C}(K_\star)$ vanishes

$$|\lambda_j(H_K\mathcal{C}(K_\star))| = |\lambda_j(H_K\mathcal{C}(K_\star)) - \lambda_j(A)| \leq \|B^{(m)}\|_F \to 0, \quad \text{as } m \to \infty. \qquad \square$$

3.2.2. Pointwise measurement closeness. We now study the second scenario of Theorem 3.3 and consider μ_1, μ_2 as mollifications of a singular pointwise testing. For this purpose, let $\xi \in C_c^\infty(\mathbb{R})$ be a smooth function, compactly supported in the unit ball $B_1(0)$ with $0 \leq \xi \leq 1$ and $\xi(0) = 1$. In the following, we consider the measurement test functions

$$\mu_i^\eta(x) = \frac{1}{\eta}\xi\left(\frac{x - x_i}{\eta}\right), \quad i = 1, 2. \tag{3.12}$$

Our aim is to show that the assertion of Theorem 3.3 is true independently of the mollification parameter $\eta > 0$. This shows that in the limit as $\eta \to 0$, i.e. in the pointwise measurement case, we still lose strong convexity around $K_\star$.

PROPOSITION 3.10. *Let μ_1^η, μ_2^η be of the form (3.12) with measurement locations $x_2 \notin \{a_r\}_{r=1,\dots,R}$ for the partition of $\mathbb{R}$ from (2.1). Consider a small neighbourhood of $K_\star$ and let Assumption 2.1 hold. Additionally, let the measurement time T and locations be chosen such that*

$$(e^{T|V|C_K} - 1) < 1, \qquad \min_r |x_2 - a_r| - T > \eta_0 > 0.$$

If the initial condition ϕ is uniformly continuous in x, uniformly in v, then

$$\nabla_K M_1(K) \to \nabla_K M_2(K) \qquad \textit{as } x_1 \to x_2 \textit{ in the standard Euclidean norm,}$$

and the convergence is independent of $\eta \leq \eta_0$.

This proposition explains the breakdown of well-posedness presented in Theorem 3.3 in the second scenario. Since the proof for the theorem is rather similar to that of the first scenario, we omit it from here.

Similar to the previous scenario, we need to show smallness of the gradient difference (3.10). This time, we have to distinguish two sources of smallness: For singular parts of the adjoint $\bar{g}$, the smallness of the corresponding gradient difference is generated by testing it on a sufficiently regular f at close measuring locations. So it is small in the weak sense. The regular parts $\bar{g}_{>N}$ of $\bar{g}$ represent the difference of $\bar{g}$ and its singular parts and evolve form the integral operator on the right hand side of (2.8), which exhibits a diffusive effect. Smallness is obtained by adjusting the cut off regularity N.

Let us mention, however, that the time constraint is mostly induced for a technical reason. In order to bound the size of the regular parts of the adjoint solution, we use the plain Grönwall inequality which leads to an exponential growth that we counterbalance by a small measuring time T. The spatial requirement $\min_r |x_2 - a_r| - T > \eta_0 > 0$ is a reflection of the fact that we need the measuring blob (support of μ) to be somewhat centered in the constant pieces of the piecewise-constant function K. This helps to force the measuring to precisely pick up only the information from that particular piece. This specific design will later be discussed in Section 4 as well.

To put the above considerations into a mathematical framework, we deploy the singular decomposition approach, and we are to decompose

$$\bar{g} = \sum_{n=0}^{N} \bar{g}_n + \bar{g}_{>N}\,, \tag{3.13}$$

where the regularity of $\bar{g}_n$ increases with n. Here, we define $\bar{g}_0$ as the solution to

$$\begin{aligned} -\partial_t \bar{g}_0 - v \cdot \nabla_x \bar{g}_0 &= -\sigma \bar{g}_0\,,\\ \bar{g}_0(t = T, x, v) &= \mu_2^\eta(x) - \mu_1^\eta(x)\,, \end{aligned}$$

for $\sigma(x,v) := \int_V K(x,v',v)\,\mathrm{d}v'$, and $\bar{g}_n$ are inductively defined by

$$\begin{aligned} -\partial_t \bar{g}_n - v \cdot \nabla_x \bar{g}_n &= -\sigma \bar{g}_n + \tilde{\mathcal{L}}(\bar{g}_{n-1})\,,\\ \bar{g}_n(t = T, x, v) &= 0\,, \end{aligned} \tag{3.14}$$

where we used the notation $\tilde{\mathcal{L}}(\bar{g}) := \int K(x,v',v)\bar{g}(x,t,v')\,\mathrm{d}v'$. The remainder $\bar{g}_{>N}$ satisfies

$$\begin{aligned} -\partial_t \bar{g}_{>N} - v \cdot \nabla_x \bar{g}_{>N} &= -\sigma \bar{g}_{>N} + \tilde{\mathcal{L}}(\bar{g}_N + \bar{g}_{>N})\,,\\ \bar{g}_{>N}(t = T, x, v) &= 0\,. \end{aligned} \tag{3.15}$$

It is a straightforward calculation that

$$(3.10) = \sum_{n=0}^{N} \int_0^T \int_{I_r} f'(\bar{g}_n' - \bar{g}_n)\,\mathrm{d}x\,\mathrm{d}t + \int_0^T \int_{I_r} f'(\bar{g}_{>N}' - \bar{g}_{>N})\,\mathrm{d}x\,\mathrm{d}t\,. \tag{3.16}$$

We are to show, in the two lemmas below, that both terms are small when $x_1 \to x_2$. To be more specific:

Lemma 3.11. *Let the assumptions of Proposition 3.10 be satisfied. For any $\varepsilon > 0$, and any $n \in \mathbb{N}_0$, there exists a $\delta_n(\varepsilon) > 0$ such that*

$$\left| \int_0^T \int_{I_r} f' \bar{g}_n \,dx\,dt \right| \le \varepsilon\,, \quad \textit{if} \quad |x_1 - x_2| < \delta_n(\varepsilon)\,. \tag{3.17}$$

The remainder can be bounded similarly.

LEMMA 3.12. *Under the assumptions of Proposition 3.10, one has*

$$\left|\int_0^T \int_{I_r} f' \bar{g}_{>N} \, dx \, dt\right| \leq T^2 |V| C_K C_\phi e^{2|V|C_K T} (e^{C_K |V| T} - 1)^N C_\mu,$$

which becomes arbitrarily small for large N.

The proofs for both lemmas exploit the continuity of f by choice of ϕ, and the smallness of the higher regularity components of the g term. Since it is not keen to the core of the paper, we leave the details to Appendix C. The application of the two lemmas gives Proposition 3.10:

Proof of Proposition 3.10. Let $\varepsilon > 0$. Because $e^{C_K|V|T} - 1 < 1$ by assumption, we can choose $N \in \mathbb{N}$ large enough such that $2T^2|V|C_K C_\phi e^{2|V|C_K T}(e^{C_K|V|T} - 1)^N < \frac{\varepsilon}{2}$. Furthermore, let $|x_1 - x_2| < \min_{n \leq N} \delta_n(\frac{\varepsilon}{4(N+1)})$. Then with the triangle inequality and Lemmas 3.11 and 3.12, we obtain from (3.16)

$$\begin{aligned}
\left|\frac{\partial (M_1 - M_2)(K)}{\partial K_{r,i}}\right| \leq & \sum_{n=0}^N \left|\int_0^T \int_{I_r} f'(\bar{g}_n' - \bar{g}_n) \, \mathrm{d}x \, \mathrm{d}t\right| + \left|\int_0^T \int_{I_r} f'(\bar{g}_{>N}' - \bar{g}_{>N}) \, \mathrm{d}x \, \mathrm{d}t\right| \\
\leq & 2N \frac{\varepsilon}{4(N+1)} + 2T^2 |V| C_K C_\phi e^{2|V|C_K T} (e^{C_K|V|T} - 1)^N C_\mu \\
\leq & \varepsilon . \qquad \square
\end{aligned}$$

4. Experimental Design.

We now provide an explicit experimental setup that ensures well-posedness. Recalling that Proposition 3.4 requires the positive-definiteness of the Hessian term at $K_\star$, we are to design a special experimental setup that validates this assumption. We propose to use the following:

DESIGN 4.1. *We divide the domain $I = [a_0, a_R)$ into R intervals $I = \dot\bigcup_{r=1}^R I_r$ with $I_r = [a_{r-1}, a_r)$, and the center for each interval is denoted by $a_{r-1/2} := \frac{a_{r-1}+a_r}{2}$. The spatial supports of the values $K_r(v, v')$ takes on the form of (2.1). The design is:*

- *<u>initial condition</u> $\phi(x, v) = \sum_{r=1}^R \phi_r(x)$ is a sum of R positive functions ϕ_r that are compactly supported in $a_{r-1/2} + [-d, d]$ with $d < \min\left(\frac{a_r - a_{r-1}}{4}\right)$, symmetric and monotonously decreasing in $|x - a_{r-1/2}|$ (for instance, a centered Gaussian with a cut-off tail);*
- *<u>measurement test functions</u> $\mu_{l_i^r} = \bar{C}_\mu \mathbb{1}_{[(-1)^i T - d_\mu, (-1)^i T + d_\mu] + a_{r-1/2}}$, $i = 1, 2$, for some $\bar{C}_\mu > 0$, centered around $a_{r-1/2} \pm T$ with $d_\mu \leq d$;*
- *<u>measurement time</u> T such that*

$$T < \min\left((1 - \delta)\frac{0.09}{C_K |V|}, \min_r \left(\frac{a_r - a_{r-1}}{4} - \frac{d}{2}\right)\right) \tag{4.1}$$

$$\text{for} \quad \delta = (d + d_\mu)/T < e^{-TC_K|V|}. \tag{4.2}$$

Remark 4.2. Note that this design requires a delicate balancing between T and d and d_μ. Requirement (4.1) prescribes that T must not be too large. On the other hand, (4.2) requires that it must not be too small compared to d, d_μ. An exemplary choice of $d = d_\mu = cT^2$ for some $c > 0$, for instance, automatically verifies requirement (4.2) for small enough T.

This particular design of initial data and measurement is to respond to the fact that the equation has a characteristic and particles moves along the trajectories. The measurement is set up to single out the information we would like to reconstruct along the propagation. The visualization of this design is plotted in Figure 1.

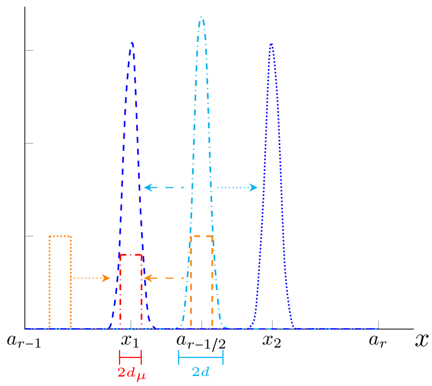

Fig. 1: Motion of the ballistic parts $f^{(0)}(t=0,v)$ (cyan, dashdotted) to $f^{(0)}(t=T,v=+1)$ (blue, dotted) and $f^{(0)}(t=T,v=-1)$ (blue, dashed) and $g_1^{(0)}(t=0,v=+1)$ (orange, dotted) and $g_1^{(0)}(t=0,v=-1)$ (orange, dashed) to $g_1^{(0)}(t=T,v)$ (red, dashdotted), compare also (4.5).

Under this design, we have the following proposition:

Proposition 4.3. *The design (D) decouples the reconstruction of K_r. To be more specific, recall* (2.2)

$$K = [K_r], \quad \textit{with} \quad K_r = [K_{r,1}, K_{r,2}].$$

The Hessian $H_K\mathcal{C}$ has a block diagonal structure with each of the blocks is a 2×2 matrix given by the Hessian $H_{K_r}\mathcal{C}$.

Proof. By the linearity of (1.1) and (2.8), their solutions $f=\sum_{r=1}^R f_r$ and $g=\sum_{r=1}^R\sum_{i=1}^2 g_{l_i^r}$ decompose into solutions f_r of (1.1) with initial conditions ϕ_r and $g_{l_i^r}$ with final condition $-(M_{l_i^r}-y_{l_i^r})\mu_{l_i^r}/2R$, the summands of the final condition (2.9), correspondingly. By construction of T and the constant speed of propagation $|v|=1$, the spatial supports of f_r and $g_{l_1^r}$, $g_{l_2^r}$ are is fully contained only in I_r for all $t\in[0,T], v\in V$. As such, only f_r and $g_{l_j^r}$ carry information about K_r, and no information for other K_s with $s\neq r$. ∎

This not only makes boundary conditions superfluous, but also translates the problem of finding a $2R$ valued vector K into R individual smaller problems of finding the two-constant pair $(K_{r,1},K_{r,2})$ within I_r. This comes with the cost of prescribing very detailed measurements depending on the experimental scales I_r and d, but opens the door for parallelized computation.

Furthermore, under mild conditions, this design ensures the local reconstructability of the inverse problem.

Theorem 4.4. *Let Assumption 2.1 hold. Given the Hessian $H_K M_l(K)$ is bounded in Frobenius norm in a neighbourhood of $K_\star$, Design (D) generates a locally well-posed optimization problem* (2.5).

The proof is layed out in the subsequent subsection 4.1.

Remark 4.5. Let us mention that the bounds for T in Design (D) are not optimal. In the proof of theorem 4.4 we used crude estimates, and we believe these estimates can potentially be relaxed to allow for longer measurement times T. Furthermore, the setup can easily be modified to use different measurement times for different intervals I_r, for instance. In this case, again, the bounds on T can be relaxed.

Remark 4.6. Design (D) shares similarities with the theoretical reconstruction setting in [27], where a pointwise reconstruction of a continuous kernel $\tilde{K}$ was obtained by a sequence of experiments where the measurement time T became small and the measurement location gets close to the initial location. The situation is also seen here. As we refine the discretization for the underlying K-function using higher dimensional vector, measurement time has to be shortened to honor the refined discretization. However, we should also note the difference. In [27], we studied the problem in higher dimension and thus explicitly excluded the ballistic part of the data from the measurement

4.1. Proof of Theorem 4.4. According to Theorem 3.2, one only needs to show $H_K\mathcal{C}(K_\star) > 0$. As the Hessian attains a block diagonal structure (Proposition 4.3), we are to study the 2×2-blocks

$$H_{K_r}\mathcal{C}(K_\star) = \nabla_{K_r} M_{l_1^r}(K_\star) \otimes \nabla_{K_r} M_{l_1^r}(K_\star) + \nabla_{K_r} M_{l_2^r}(K_\star) \otimes \nabla_{K_r} M_{l_2^r}(K_\star). \tag{4.3}$$

Here the two measurements $M_{l_1^r}$, $M_{l_2^r}$ are inside I_r, and $\nabla_{K_r} = [\partial_{K_{r,1}}, \partial_{K_{r,2}}]$. The positive definiteness of the full $H_K\mathcal{C}(K_\star)$ is equivalent to the positive definiteness of each individual $H_{K_r}\mathcal{C}(K_\star)$. This is established in the subsequent proposition.

Proposition 4.7. *Let Assumption 2.1 hold. If the Hessian $H_K M_l(K)$ is bounded in Frobenius norm in a neighbourhood of $K_\star$, then the Design (D) produces a positive-definite Hessian $H_K\mathcal{C}(K_\star)$.*

According to (4.3), $H_{K_1}\mathcal{C}(K_\star)$ is positive definite if

$$\left|\frac{\partial M_1(K_\star)}{\partial K_{1,1}}\right| > \left|\frac{\partial M_1(K_\star)}{\partial K_{1,2}}\right| \quad \text{and} \quad \left|\frac{\partial M_2(K_\star)}{\partial K_{1,1}}\right| < \left|\frac{\partial M_2(K_\star)}{\partial K_{1,2}}\right| \tag{4.4}$$

holds true for the measurements M_1, M_2 corresponding to K_1. Due to design symmetry, it is sufficient to study the first inequality. Consider the difference $\frac{\partial M_1(K_\star)}{\partial K_{1,1}} - \frac{\partial M_1(K_\star)}{\partial K_{1,2}}$. Similar to (3.13) and (3.16), we are to decompose the equation for f and g ((1.1) and (3.6) respectively, with $K = K_\star$) into the ballistic parts $g_1^{(0)}$ and $f^{(0)}$ and the remainder terms. Namely, let $g_1^{(0)}$ and $f^{(0)}$ satisfy

$$\begin{cases} -\partial_t g_1^{(0)} - v\cdot\nabla_x g_1^{(0)} &= -\sigma g_1^{(0)} \\ g_1^{(0)}(t=T,x,v) &= \mu_1(x) \end{cases} \quad \text{and} \quad \begin{cases} \partial_t f^{(0)} - v\cdot\nabla_x f^{(0)} &= -\sigma f^{(0)} \\ f^{(0)}(t=0,x,v) &= \phi(x,v). \end{cases} \tag{4.5}$$

Then the following two lemmas are in place with $\mu_1(x)$ and $\phi(x,v)$ as in Design (D).

Lemma 4.8. *In the setting of Proposition 4.7, for $(v,v') = (+1,-1)$, the ballistic part*

$$\begin{aligned} B := &\left|\int_0^T \int_{I_1} f^{(0)}(v')(g_1^{(0)}(v') - g_1^{(0)}(v))\,dx\,dt\right| \\ &- \left|\int_0^T \int_{I_1} f^{(0)}(v)(g_1^{(0)}(v) - g_1^{(0)}(v'))\,dx\,dt\right| \end{aligned} \tag{4.6}$$

satisfies

$$B \geq C_{\phi\mu}\left(e^{-TC_K|V|}T - (d_\mu + d)\right) > 0, \tag{4.7}$$

where $C_{\phi\mu} = \int_{I_1} \phi_1(x)\mu_1(-T+x)\,dx = \max_{a,b}\int_{I_1}\phi_1(x+a)\mu_1(-T+x+b)\,dx$ *by construction of* ϕ_1, μ_1.

At the same time, the remainder term is small.

LEMMA 4.9. *In the setting of Proposition 4.7, the remaining scattering term*

$$S := \int_0^T \int_{I_1} f(v')(g_1(v') - g_1(v))\,dx\,dt - \int_0^T \int_{I_1} f^{(0)}(v')(g_1^{(0)}(v') - g_1^{(0)}(v))\,dx\,dt$$

is bounded uniformly in (v, v') *by*

$$|S| \leq 4C_{\phi\mu}T\frac{C_K|V|T}{(1-C_K|V|T)^2}. \tag{4.8}$$

Proposition 4.7 is a corollary of Lemmas 4.8, 4.9.

Proof of Proposition 4.7. By the bounds obtained in lemmas 4.8, 4.9, one has

$$\begin{aligned}
\left|\frac{\partial M_1(K_\star)}{\partial K_{1,1}}\right| - \left|\frac{\partial M_1(K_\star)}{\partial K_{1,2}}\right| &\geq B - 2|S| \\
&\geq C_{\phi\mu}\left(e^{-TC_K|V|}T - (d_\mu + d)\right) - 8C_{\phi\mu}T\frac{C_K|V|T}{(1-C_K|V|T)^2} \\
&\geq C_{\phi\mu}T\left(1 - TC_K|V| - \delta - 8\frac{0.09(1-\delta)}{(1-0.09)^2}\right).
\end{aligned}$$

By assumption $0 < T < (1-\delta)\frac{0.09}{C_K|V|}$ with $\delta = \frac{d+d_\mu}{T} < 1$, the last line is positive. In total, this shows the first part of inequality (4.4). As the second part can be treated in analogy, it follows that $H_{K_1}\mathcal{C}(K_\star)$ is positive definite. ∎

Finally, Theorem 4.4 is a direct consequence of Proposition 4.7.

Proof of Theorem 4.4. Repeated application of the arguments to all $H_{K_r}\mathcal{C}(K_\star)$, $r = 1, \ldots, R$, shows that $H_K\mathcal{C}(K_\star) > 0$. By the assumption of boundedness of the Hessian $H_K M_l(K)$ in a neighbourhood of $K_\star$, theorem 3.2 proves local well-posedness of the inverse problem. ∎

The proofs for the Lemmas 4.8-4.9 are rather technical and we leave them to Appendix D. Here we only briefly present the intuition. According to Figure 1, $f^{(0)}(v' = -1)$ and $g_1^{(0)}(v' = -1)$ have a fairly large overlapping support, whereas $g_1^{(0)}(v = +1)$ overlaps with $f^{(0)}(v' = -1)$ and $g_1^{(0)}(v' = -1)$ with $f^{(0)}(v = +1)$ only for a short time spans $T \approx T$ and $T \approx 0$ respectively. Recalling (4.6), we see the negative components of the term B are small, making B positive overall. The smallness of S is a result of small measurement time T.

5. Numerical experiments. As a proof of concept for the prediction given by the theoretical results in Section 3, we present some numerical evidence.

An explicit finite difference scheme is used for the discretization of (1.1) and (2.8). In particular, the transport operator is discretized by the Lax-Wendroff method and the operator $\mathcal{K}$ is treated explicitly in time. The scheme can be shown to be consistent

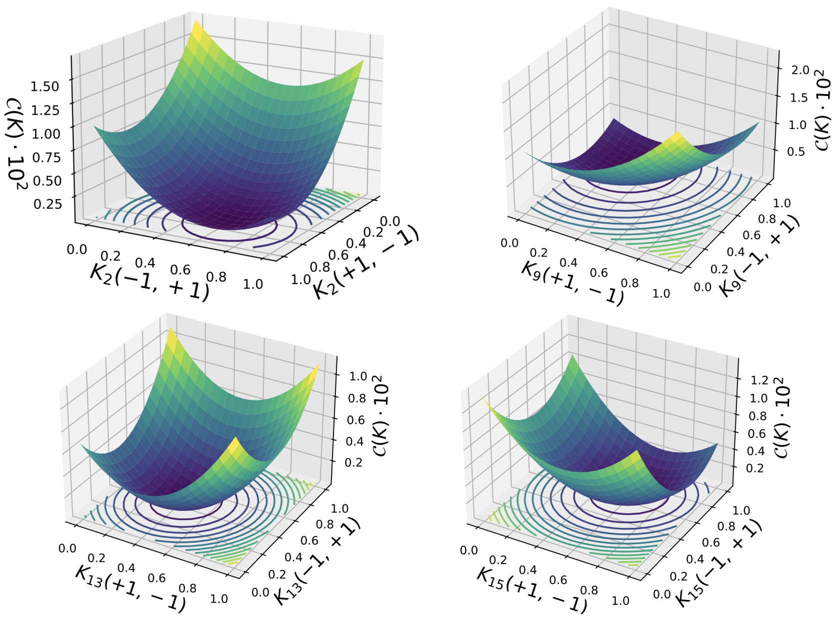

Fig. 2: (Marginal) loss functions $\mathcal{C}(K)$ for $R = 20$: For a fixed $r \in \{2, 9, 13, 15\}$, we plot $\mathcal{C}$ as a function of K_r with all $K_{s \neq r}$ set to be the ground-truth $(K_\star)_s$.

and stable when $\Delta t \leq \min(\Delta x, C_K^{-1})$, and thus it converges according to the Lax-Equivalence theorem. More sophisticated solvers for the forward model [21] can be deployed when necessary. Also, when a compatible solver [4] for the adjoint equation exists, these pairs of solvers can readily be incorporated in the inversion setting.

All subsequent experiments were conducted with noise free synthetic data $y_l = M_l(K_\star)$ that was generated by a forward computation with the true underlying parameter $K_\star$.

5.1. Illustration of well-posedness. In Section 4, it was suggested a specific design of initial data and measurement mechanism can provide a successful reconstruction of the kernel K, and that the loss function is expected to be strongly convex. We observe it numerically as well. In particular, we set $R = 20$ and use Gaussian initial data, and plot the (marginal) loss function in Figure 2. Figure 3 depicts the convergence of some parameter values $K_r(v, v')$ in this scenario against the corresponding loss function value. An exponential decay of the loss function, as expected from theory [42, Th.3], can be observed.

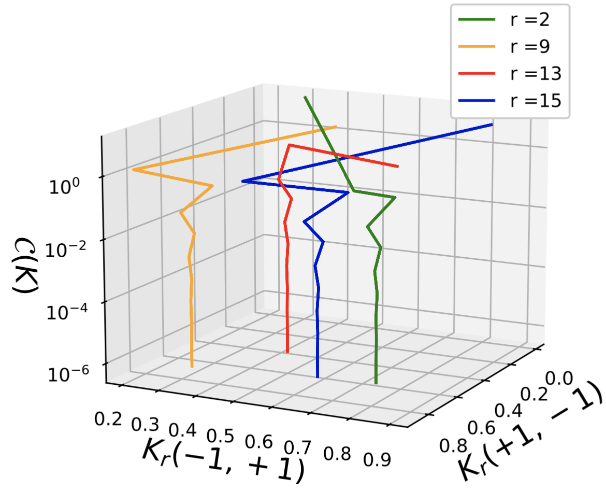

Fig. 3: Convergence of the parameter values $K_r(v, v')$ from (2.1) for $r = 2, 9, 13, 15$ to the ground truth as the cost function converges.

The strictly positive-definiteness feature persists in a small neighborhood of the optimal solution $K_\star$. This means adding a small perturbation to $K_\star$, the minimal eigenvalue of the Hessian matrix $H_K\mathcal{C}(K)$ stays above zero. In Figure 4 we present, for two distinct experimental setups, the minimum eigenvalue as a function of the perturbation to $K_r(v, v')$. In both cases, the green spot (the ground-truth) is positive, and it enjoys a small neighborhood where the minimum eigenvalue is also positive, as predicted by Theorem 3.2. In the right panel, we do observe, as one moves away from the ground-truth, the minimal eigenvalue takes on a negative value, suggesting the loss of convexity. This numerically verifies that the well-posedness result in Theorem 3.2 is local in nature. The panel on the left deploys the experiment design provided by Section 4. The simulation is ran over the entire domain of $[0, 1]^2$ and the positive-definiteness stays throughout the domain, hinting the proposed experimental design (D) can potentially be globally well-posed.To generate the plots, a simplified setup with $R = 2$ and constant initial data was considered.

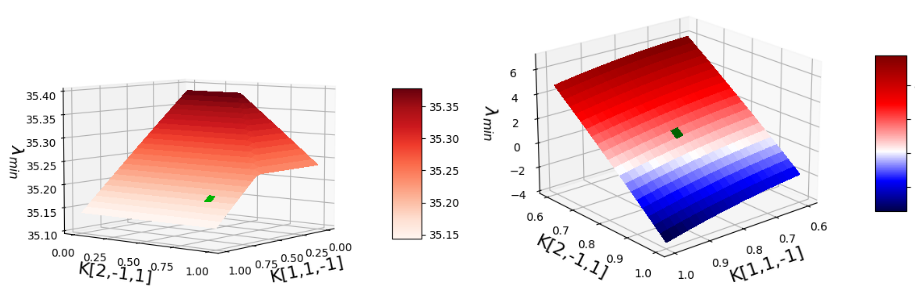

Fig. 4: Minimal eigenvalues of the Hessian $H_K\mathcal{C}(K)$ around the true parameter $K_\star$ for two experimental designs. We perturb K by changing values in $K_1(1, -1)$ and $K_2(-1, 1)$. The ground-truth is marked green in both plots.

5.2. Ill-conditioning for close measurement locations . We now provide numerical evidence to reflect the assertion from 3.2. In particular, the strong convexity of the loss function would be lost if measurement location x_1 becomes close to x_2.

We summarize the numerical evidence in Figure 5. Here we still use $R = 20$ and constant initial data but vary the detector positions. To be specific, we assign values to x_1 using $\{x_1^{(0)} = c_1 - T, x_1^{(1)} = c_1 + \frac{T}{2}, x_1^{(2)} = c_1 + \frac{4}{5}T, x_1^{(3)} = x_2 = c_1 + T\}$. As the superindex grows, $x_1 \to x_2$ with $x_1^{(3)} = x_2$ when the two measurements exactly coincide. For $x_1 = x_2$, the cost function is no longer strongly convex around the ground truth $K_\star$, as its Hessian is singular. The thus induced vanishing learning rate $\eta = \frac{2\lambda_{\min}}{\lambda_{\max}^2}$ was exchanged by the learning rate for $x_1 = x_1^{(2)}$ in this case to observe the effect of the gradient.

In the first, third and fourth panel of Figure 5, we observe that the cost function as well as the parameter reconstructions for K_9 and K_{15} nevertheless converge,but convergence rates that slow down significantly comparing purple (for $x_1^{(0)}$), blue (for $x_1^{(1)}$), green (for $x_1^{(2)}$) and orange (for $x_1^{(3)}$) due to smaller learning rates. The overlap of the parameter reconstructions for $x_1 \in \{x_1^{(2)}, x_1^{(3)}\}$ is due to the coinciding choice of the learning rate and a very similar gradient for parameters K_9, K_{15} whose information is not reflected in the measurement in x_1.

As parameter K_1 directly affects the measurement at x_1, Panel 2 showcases the degenerating effect of the different choices of x_1 on the reconstruction. Whereas convergence is still obtained in the blue curve (for $x_1^{(1)}$), the reconstructions of K_1 from measurements at $x_1^{(2)}$ (green) and $x_1^{(3)}$ (orange) clearly fail to converge to the true parameter value in black. This offset seems to grow with stronger degeneracy in the measurements.

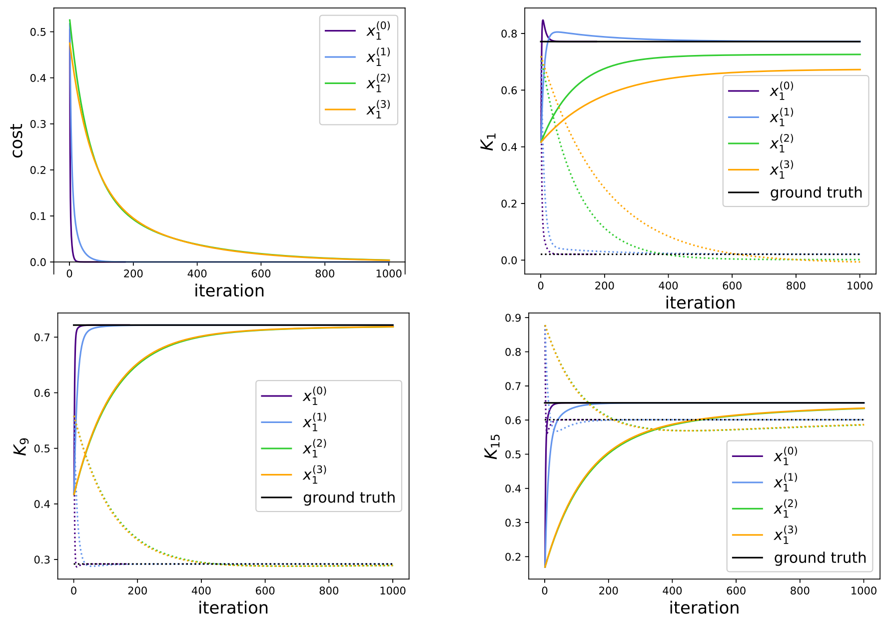

Fig. 5: Cost function and reconstructions of $K_r(+1,-1)$ (solid lines) and $K_r(-1,+1)$ (dotted lines) for $r = 1, 9, 15$ and $R = 20$ under different measurement locations for x_1. x_1 takes the values of $\{x_1^{(0)} = c_1 - T\,, x_1^{(1)} = c_1 + \frac{T}{2}\,, x_1^{(2)} = c_1 + \frac{4}{5}T\,, x_1^{(3)} = c_1 + T\}$ with $x_1^{(3)} = x_2$.

6. Discussion. As discussed in [32, 58], to accurately extract tumbling statistics, it is necessary to track single-cell trajectories, which necessitates a low cell concentration and is constrained to shorter trajectories. This will result in insufficient statistical accuracy for reliable extraction of velocity jump statistics. In this paper we present an optimization framework for the reconstruction of the velocity jump parameter K in the chemotaxis equation (1.1) using velocity averaged measurements (2.3) from the interior domain. The velocity-averaged measurements do not require tracking single-cell trajectories, thus allowing for the measurement of higher cell density over a longer period of time. This may provide a new and reliable way of determining the microscopic statistics. In the numerical setting when PDE-constrained optimization is deployed, depending on the experimental setup, the problem is can be either locally well-posedness or ill-conditioned. We further propose a specific experimental design that is adaptive to the discretization of K. This design decouples the reconstruction of local values of the parameter K using the corresponding measurements. The design thus opens up opportunities to parallelization. As a proof of concept, numerical evidence were presented. They are in good agreement with the theoretical predictions

A natural extension of the results presented in the current paper is the algorithmic performance in higher dimensions. The theoretical findings seem to apply in a straightforward manner, but details need to be evaluated. Numerically one can certainly also refine the solver implementation. For example, it is possible that higher order numerical PDE solvers that preserve structures bring extra benefit. More sophisticated optimization methods such as the (Quasi-)Newton method or Sequential Quadratic Programming are possible alternatives for conducting the inversion [8, 26, 44, 50]. Furthermore, we adopted a first optimize, then discretize approach in this article. Suggested in [4, 25, 37], a first discretize, then optimize framework could be bring automatic compatibility of forward and adjoint solvers, but extra difficulties [28] need to be resolved. Error estimates for continuous in space ground truth parameters as in [31] could help practitioners to select a suitable space-discretization.

Our ultimate goal is to form a collaboration between practitioners to solve the real-world problem related to bacteria motion reconstruction [34]. To that end, experimental design is non avoidable. A class of criteria proposed under the Bayesian perspective shed light on this topic, see [2] and references therein. In our context, it translates to whether the design proposed in Section 4 satisfies these established optimality criteria.

Appendix A. Derivation of the gradient (2.7)**.** This section justifies formula (2.7) for the gradient of the cost function $\mathcal{C}$ with respect to K. Let us first introduce some notation: Denote by

$$\mathcal{J}(f) := \frac{1}{2L}\sum_{l=1}^{L}\left(\int_{\mathbb{R}}\int_V f(T,x,v)\,\mathrm{d}v\ \mu_l(x)\,\mathrm{d}x - y_l\right)^2$$

the loss for $f \in \mathcal{Y} = \{h \mid h, \partial_t h + v\cdot\nabla h \in C^0([0,T];L^\infty(\mathbb{R}\times V))\}$. Note that mild solutions of (1.1) are contained in $\mathcal{Y}$, since $\mathcal{K}(f) \in C^0([0,T];L^\infty(\mathbb{R}\times V))$ by regularity of f from Lemma B.1. Then $\mathcal{C}(K) := \mathcal{J}(f_K)$ in the notation of (2.3). The Lagrangian function for the PDE constrained optimization problem (2.5) reads

$$\mathcal{L}(K,f,g,\lambda) = \mathcal{J}(f) + \langle g, \partial_t f + v\cdot\nabla f - \mathcal{K}(f)\rangle_{x,v,t} + \langle \lambda, f(t=0) - \phi\rangle_{x,v},$$

for $f \in \mathcal{Y}$ and $g \in \mathcal{Z} = \{h \mid h, \partial_t h + v\cdot\nabla h \in C^0([0,T];L^\infty(V;L^1(\mathbb{R})))\}$. For $f = f_K$, our cost function $\mathcal{C}(K) = \mathcal{J}(f_K) = \mathcal{L}(K, f_K, g, \lambda)$ and

$$\frac{\mathrm{d}\mathcal{C}(\hat{K})}{\mathrm{d}K} = \frac{\partial\mathcal{L}}{\partial K}\bigg|_{\substack{K=\hat{K},\\ f=f_{\hat{K}}}} + \frac{\partial\mathcal{L}}{\partial f}\bigg|_{\substack{K=\hat{K},\\ f=f_{\hat{K}}}}\frac{\partial f_K}{\partial K}\bigg|_{K=\hat{K}}$$

To avoid the calculation of $\frac{\partial f_K}{\partial K}$, choose the Lagrange multipliers g, λ such that $\frac{\partial\mathcal{L}}{\partial f}\big|_{\substack{K=\hat{K},\\ f=f_{\hat{K}}}} = 0$. Then

$$\frac{\mathrm{d}\mathcal{C}(\hat{K})}{\mathrm{d}K_r} = \frac{\partial\mathcal{L}}{\partial K_r}\bigg|_{\substack{K=\hat{K},\\ f=f_{\hat{K}}}} = -\frac{\partial\langle g, \mathcal{K}_K(f)\rangle_{x,t,v}}{\partial K_r}\bigg|_{\substack{K=\hat{K},\\ f=f_{\hat{K}}}}$$
$$= \int_0^T\int_{I_r} f_{\hat{K}}(x,t,v')\big(g(x,t,v') - g(x,t,v)\big)\,\mathrm{d}x\,\mathrm{d}t.$$

To compute the gradient, g has to be specified. Recall the requirement

$$0 = \frac{\partial L}{\partial f}\bigg|_{\substack{K=\hat{K},\\ f=f_{\hat{K}}}}$$
$$(A.1)\qquad = \frac{1}{L}\sum_{l=1}^{L}\left(\int_{\mathbb{R}}\int_V f(T)\,\mathrm{d}v\ \mu_l\,\mathrm{d}x - y_l\right)\frac{\partial}{\partial f}\,\langle\mu_l, f(T)\rangle_{x,v}\bigg|_{\substack{K=\hat{K},\\ f=f_{\hat{K}}}}$$
$$+ \frac{\partial}{\partial f}\Big[\langle g, \partial_t f + v\cdot\nabla f - \mathcal{K}_K(f)\rangle_{x,t,v} + \langle\lambda, f(t=0)\rangle_{x,v}\Big]\bigg|_{\substack{K=\hat{K},\\ f=f_{\hat{K}}}}$$

We will motivate the choice of g such that the derivatives cancel out each other. Because we are dealing with mild solutions where integration by parts in time and space cannot be used right away, we approximate f and g by sequences of functions

- $(f^n)_n \subset C^1([0,T];L^\infty(\mathbb{R}\times V)) \cap C^0([0,T];W^{1,\infty}(\mathbb{R};L^\infty(V)))$ that converge $f_n \to f$ with $\partial_t f_n + v\cdot\nabla f_n \to \partial_t f + v\cdot\nabla f$ in $C^0([0,T];L^\infty(\mathbb{R}\times V))$ and
- $(g^n)_n \subset C^1([0,T];L^\infty(V;L^1(\mathbb{R}))) \cap C^0([0,T];L^\infty(V;W^{1,1}(\mathbb{R})))$ with $g_n \to g$ with $-\partial_t g_n - v\cdot\nabla g_n \to -\partial_t g - v\cdot\nabla g$ in $C^0([0,T];L^\infty(V;L^1(\mathbb{R})))$.

This is possible, because the respective spaces for f_n and g_n are dense in $\mathcal{Y}$ and $\mathcal{Z}$. Replacing f by f_n and g by g_n in $\langle g, \partial_t f + v \cdot \nabla f - \mathcal{K}(f)\rangle_{x,t,v}$, we obtain

$$
\begin{aligned}
\langle g, \partial_t f + v \cdot \nabla f - \mathcal{K}(f)\rangle_{x,t,v} &= \lim_n \langle g_n, \partial_t f_n + v \cdot \nabla f_n - \mathcal{K}(f_n)\rangle_{x,t,v} \\
&= \lim_n \Big(\langle -\partial_t g_n - v \cdot \nabla g_n - \tilde{\mathcal{K}}(g_n), f_n \rangle_{x,t,v} + \langle f_n(t=T), g_n(t=T)\rangle_{x,v} \\
&\qquad\qquad - \langle f_n(t=0), g_n(t=0)\rangle_{x,v} \Big) \\
&= \langle -\partial_t g - v \cdot \nabla g - \tilde{\mathcal{K}}(g), f\rangle_{x,t,v} + \langle f(t=T), g(t=T)\rangle_{x,v} - \langle f(t=0), g(t=0)\rangle_{x,v},
\end{aligned}
$$

where

$$
\tilde{\mathcal{K}}_K(g) := \int_V K(x, v', v)(g(x,t,v') - g(x,t,v))\,\mathrm{d}v'.
$$

Now, collect all terms in (A.1) with the same integration domain and equate them to 0. This leads to

$$
\begin{aligned}
&-\partial_t g - v \cdot \nabla g - \tilde{\mathcal{K}}_K(g) = 0 && \text{in } x \in \mathbb{R}, v \in V, t \in (0,T), \\
&g(x, t=T, v) = -\frac{1}{L}\sum_{l=1}^{L} \left(\int_{\mathbb{R}} \int_V f(T,x,v)\,\mathrm{d}v\ \mu_l(x)\,\mathrm{d}x - y_l \right) \mu_l(x) && \text{in } x \in \mathbb{R}, v \in V, \\
&\lambda = g(t=0) && \text{in } x \in \mathbb{R}, v \in V.
\end{aligned}
$$

Note that since g reflects the measurement procedure, it makes sense that $g(t=T)$ is isotropic in v. For computation of $\frac{\mathrm{d}\mathcal{C}(\hat{K})}{\mathrm{d}K_r}$, use the solution g to the first two equations with kernel $K = \hat{K}$ and $f = f_{\hat{K}}$.

Appendix B. Some a-priori estimates.

By Assumption 2.1, semigroup theory yields the existence of a mild solution to (1.1)–(1.2).

Lemma B.1. *Let Assumption 2.1 hold and assume $h \in L^1((0,T); L^\infty(\mathbb{R} \times V))$. Then there exists a mild solution*

$$
f \in C^0\left([0,T]; L^\infty(\mathbb{R} \times V)\right) \tag{B.1}
$$

to

$$
\begin{aligned}
\partial_t f + v \cdot \nabla_x f &= \mathcal{K}(f) + h, \\
f(t=0, x, v) &= \phi(x,v) \in L^\infty_+(\mathbb{R} \times V)
\end{aligned}
$$

that is bounded

$$
\max_v \|f(t)\|_{L^\infty(\mathbb{R})} \le e^{2|V|C_K t} C_\phi + \int_0^t e^{2|V|C_K(t-s)} \|h(s)\|_{L^\infty(\mathbb{R}\times V)}\, ds.
$$

We carry out the proof once to make clear, how the constant in the bound is derived.

Proof. Rewrite (1.1) as

$$
\partial_t f = \mathcal{A} f + \mathcal{B} f + h
$$

with operators $\mathcal{A} : \mathcal{D}(\mathcal{A}) \to \mathcal{X}, f \mapsto -v \cdot \nabla_x f$ and $\mathcal{B} : \mathcal{X} \to \mathcal{X}, f \mapsto \mathcal{K}(f)$, where the function spaces $\mathcal{D}(\mathcal{A}) := W^{1,\infty}(\mathbb{R}; L^\infty(V))$ and $\mathcal{X} := L^\infty(\mathbb{R} \times V)$ are used. The transport operator $\mathcal{A}$ generates a strongly continuous semigroup $T(t)u(x) = u(x - vt)$

with operator norm $\|T(t)\| \le 1$. Clearly, $\mathcal{B}$ is bounded in operator norm by $2|V|C_K$. The bounded perturbation theorem, see e.g. [18], shows that $\mathcal{A}+\mathcal{B}$ generates a strongly continuous semigroup S with $\|S(t)\| \le e^{2|V|C_K t}$. As $\phi \in \mathcal{X}$, (1.1) admits a mild solution

$$f(t) = S(t)\phi + \int_0^t S(t-s)h(s)\,\mathrm{d}s. \qquad \square$$

The regularity of the solution of (1.1)–(1.2) is improved by more regular initial data. This is exploited in the proof of ill-conditioning for pointwise measurement closeness in Theorem 3.3.

COROLLARY B.2. *Let Assumption 2.1 hold.*

a) Equation (1.1) *has a mild solution f is bounded*

$$\max_v \|f(t)\|_{L^\infty(\mathbb{R})} \le e^{2|V|C_K t} C_\phi \le e^{2|V|C_K T} C_\phi =: C_f. \tag{B.2}$$

b) If, additionally, the initial data ϕ is uniformly continuous in x, uniformly in v, then f is uniformly continuous in x, uniformly in v, t, i.e. $\max_v |f(t,x,v) - f(t,y,v)| < \varepsilon$ *for all $t \in [0,T]$, if $|x-y| < \delta(\varepsilon)$.*

Proof. Assertion a) is a direct consequence of lemma B.1. We focus on proving assertion b). Let $\varepsilon > 0$. By uniform continuity of ϕ in x, one can choose δ' such that

$$\operatorname*{ess\,sup}_{|x-y|<\delta', v} |\phi(x,v) - \phi(y,v)| < e^{-2C_K|V|T}\varepsilon/2. \tag{B.3}$$

Now consider $\delta := \min\left(\delta', \frac{\varepsilon e^{-2C_K|V|T}}{8C_f|V|C_K(R-1)}\right)$. Integration along characteristics yields

$$\begin{aligned}
&\operatorname*{ess\,sup}_{|x-y|<\delta, v} |f(t,x,v) - f(t,y,v)| \\
&\le \operatorname*{ess\,sup}_{|x-y|<\delta, v} |\phi(x-vt,v) - \phi(y-vt,v)| \\
&\quad + \operatorname*{ess\,sup}_{|x-y|<\delta, v} \left| \int_0^t \mathcal{K}(f)(t-s, x-vs, v) - \mathcal{K}(f)(t-s, y-vs, v)\,\mathrm{d}s \right| \\
&\le \operatorname*{ess\,sup}_{|x-y|<\delta, v} |\phi(x,v) - \phi(y,v)| \\
&\quad + 2C_K|V| \int_0^t \operatorname*{ess\,sup}_{|x-y|<\delta, v'} |f(s,x,v') - f(s,y,v')|\,\mathrm{d}s \\
&\quad + 2C_f|V| \operatorname*{ess\,sup}_{|x-y|<\delta, v} \int_0^t \max_{v',v''} |K(x-vs,v',v'') - K(y-vs,v',v'')|\,\mathrm{d}s \\
&=: (i) + (ii) + (iii),
\end{aligned}$$

where $(i) \le \frac{\varepsilon}{2} e^{-2C_K|V|T}$ by (B.3). By symmetry, $(iii) = 2\cdot(iv)$ where (iv) coincides with (iii), but $x \ge y$. As K is a step function in space (2.1), its difference can only be non zero if a jump occurred between $x - vs$ and $y - vs$. Boundedness of K in (2.10) then lead to the estimate

$$\begin{aligned}
(iii) = 2\cdot(iv) &\le 2\cdot 2C_f|V| \operatorname*{ess\,sup}_{|x-y|<\delta, v} \int_0^t C_K \sum_{r=1}^{R-1} \mathbb{1}_{(x-vs,y-vs]}(a_r)\,\mathrm{d}s \\
&\le 4C_f|V|C_K(R-1)\delta \le \frac{\varepsilon}{2} e^{-2C_K|V|T}.
\end{aligned} \tag{B.4}$$

In summary, Gronwall's lemma yields

$$\operatorname*{ess\,sup}_{|x-y|<\delta,v} |f(t,x,v) - f(t,y,v)| \le \varepsilon e^{-2C_K|V|(T-t)} \le \varepsilon. \qquad \square$$

Again, semigroup theory shows existence of the adjoint equation (2.8) with corresponding bounds.

LEMMA B.3. *Let $h \in L^1((0,T);L^\infty(V;L^1(\mathbb{R})))$, $\psi \in L^1(\mathbb{R})$ and let (2.10) hold. Then the equation*

$$\begin{aligned} -\partial_t g - v\cdot\nabla_x g &= \alpha\tilde{\mathcal{L}}(g) - \sigma g + h, \\ g(t=T) &= \psi(x) \end{aligned} \tag{B.5}$$

with $\alpha \in \{0,1\}$ and $\tilde{\mathcal{L}}(g) := \int K(x,v',v)g(x,t,v')\,dv'$ and $\sigma(x,v) := \int K(x,v',v)\,dv'$ has a mild solution

$$g \in C^0\left([0,T];L^\infty(V;L^1(\mathbb{R}))\right) \tag{B.6}$$

that satisfies

$$\|g(t)\|_{L^\infty(V;L^1(\mathbb{R}))} \le e^{(1+\alpha)|V|C_K(T-t)}\left(\|\psi\|_{L^1(\mathbb{R})} + \int_0^{T-t}\max_v \|h(T-s,v)\|_{L^1(\mathbb{R})}\,ds\right). \tag{B.7}$$

If, additionally, $h \in L^\infty([0,T]\times V;L^1(\mathbb{R}))$, then

$$\begin{aligned} &\|g(t)\|_{L^\infty(V;L^1(\mathbb{R}))} \\ &\le e^{(1+\alpha)|V|C_K(T-t)}\|\psi\|_{L^1(\mathbb{R})} + \frac{e^{(1+\alpha)|V|C_K(T-t)}-1}{(1+\alpha)|V|C_K}\operatorname*{ess\,sup}_{t,v}\|h(t,v)\|_{L^1(\mathbb{R})}. \end{aligned} \tag{B.8}$$

Proof. Repeating the arguments in the proof of Lemma B.1, semigroup theory yields the existence of a mild solution

$$g(t) = S(T-t)\psi + \int_0^{T-t} S(T-t-s)h(T-s)\,\mathrm{d}s$$

for the semigroup $S(t)$ generated by the operator $v\cdot\nabla_x + \alpha\tilde{\mathcal{L}} - \sigma$ with $\|S(t)\| \le e^{(1+\alpha)|V|C_K t}$. This yields (B.7) and (B.8). $\square$

Appendix C. Proof of Lemma 3.11-3.12. In this section, we provide the proof for the two Lemmas in section 3.2. In particular, Lemma 3.11 discusses the smallness of the first term in (3.16).

Proof for Lemma 3.11. By the assumption on the initial data and Corollary B.2 b), f is uniformly continuous in x, uniformly in v,t. For $n=0$, the boundedness (3.17) is a consequence of the explicit representation

$$\bar{g}_0(t,x,v_0) = e^{-\int_0^{T-t}\sigma(x+v_0\tau,v_0)\,\mathrm{d}\tau}(\mu_2^\eta - \mu_1^\eta)(x+v_0(T-t)) \tag{C.1}$$

together with the step function shape (2.1) of K, the continuity of f and our assumptions: Write $p_0(t,x,v_0,v') := f(x,t,v')e^{-\int_0^{T-t}\sigma(x+v_0\tau,v_0)\,\mathrm{d}\tau}$ and assume without loss of generality $x_1 \geq x_2$, then

$$\begin{aligned}
&\int_{I_r} f' \bar{g}_0 \,\mathrm{d}x \\
&= \int_{I_r} p_0(t,x,v_0,v')(\mu_2^\eta - \mu_1^\eta)(x + v_0(T-t)) \,\mathrm{d}x \\
&= -\int_{a_{r-1}-(x_1-x_2)}^{a_{r-1}} p_0(t, x + (x_1 - x_2), v_0, v') \mu_2^\eta (x + v_0(T-t)) \,\mathrm{d}x \\
&\quad + \int_{a_r-(x_1-x_2)}^{a_r} p_0(t,x,v_0,v') \mu_2^\eta(x + v_0(T-t)) \,\mathrm{d}x \\
&\quad + \int_{a_{r-1}}^{a_r-(x_1-x_2)} (p_0(t,x,v_0,v') - p_0(t, x + (x_1-x_2), v_0, v')) \mu_2^\eta (x + v_0(T-t)) \,\mathrm{d}x,
\end{aligned}$$

where we used the substitution $x \to x - (x_1 - x_2)$ for the integration domain of test function $\mu_1^\eta(x) = \mu_2^\eta(x - (x_1 - x_2))$. By uniform continuity and boundedness of f a similar argumentation as in (B.4) shows that $p_0(t,x,v_0,v')$ is uniformly continuous in x, uniformly in t, v_0, v', as well. The corresponding threshold from the epsilon-delta criterion is denoted by $\delta_{p_0}(\varepsilon)$. Then, for $0 \leq |x_1 - x_2| < \delta_0(\varepsilon) := \min(\min_r |a_r - x_2| - T - \eta_0, \delta_{p_0}(\varepsilon))$, the first two integrals vanish, because $\mu_2^\eta(x + v_0(T-t)) = 0$ for all x in the integration domain. We are left with

$$\begin{aligned}
\left| \int_{I_r} f' \bar{g}_0 \,\mathrm{d}x \right| &\leq \int\limits_{a_{r-1}}^{a_r-(x_1-x_2)} |p_0(t,x,v_0,v') - p_0(t, x + (x_1 - x_2), v_0, v')| \mu_2^\eta (x + v_0(T-t)) \,\mathrm{d}x \\
&\leq \varepsilon \int_{\mathbb{R}} \mu_2^\eta (x + v_0(T-t)) \,\mathrm{d}x = \varepsilon.
\end{aligned}$$

For $n \geq 1$, source iteration shows that the solution to (3.14) has the form

$$\begin{aligned}
\bar{g}_n(t,x,v_0) = \int\limits_0^{T-t} \int_V \dots \int\limits_0^{T-t-\sum_{j=0}^{n-2} s_j} \int_V p_n(t, x, (v_i)_{i=0,\dots,n}, (s_j)_{j=0,\dots,n-1}) \cdot \\
(\mu_2 - \mu_1) \left(x + \sum_{l=0}^{n-1} v_l s_l + v_n \left(T - t - \sum_{l=0}^{n-1} s_l \right) \right) \mathrm{d}v_n \,\mathrm{d}s_{n-1} \dots \mathrm{d}v_1 \,\mathrm{d}s_0 \,.
\end{aligned}$$

The function p_n is bounded $0 \leq p_n \leq C_K^n$ and satisfies

$$\int_0^T |p_n(t, x + v_n t, (v_i)_i, (s_j)_j) - p_n(t, y + v_n t, (v_i)_i, (s_j)_j)| \,\mathrm{d}t < \varepsilon$$

for $|x - y| < \delta_{p_n}(\varepsilon)$, uniformly in $(v_i)_i, (s_j)_j$. The assertion then follows in analogy to the case $n = 0$. ∎

Lemma 3.12 argues the smallness of the second term in (3.16). We provide the proof below. It is a consequence of the smallness of $\bar{g}_{>N}$ by Lemma B.3 and the boundedness of f.

Proof for Lemma 3.12. Application of lemma B.3 to $g = \bar{g}_{>N}, h = \tilde{\mathcal{L}} \bar{g}_N, \alpha = 1$ and $\psi = 0$ yields

$$\begin{aligned}
\max_v \int_{\mathbb{R}} |\bar{g}_{>N}(t)| \,\mathrm{d}x &\leq e^{2C_K|V|(T-t)} \int_0^{T-t} \sup_v \|\tilde{\mathcal{L}}(\bar{g}_N)(T-s,v)\|_{L^1(\mathbb{R})} \,\mathrm{d}s \\
&\leq |V| C_K (T-t) e^{2C_K|V|(T-t)} \operatorname*{ess\,sup}_{s,v} \|\bar{g}_N(s,x,v)\|_{L^1(\mathbb{R})}.
\end{aligned}$$

Now, application of the same lemma to the evolution equation (3.14) for g_n, $n = 1, ..., N$, shows

$$\operatorname*{ess\,sup}_{t,v} \int_{\mathbb{R}} |\bar{g}_n| \,\mathrm{d}x \le (e^{C_K|V|T} - 1) \operatorname*{ess\,sup}_{s,v} \int_{\mathbb{R}} |\bar{g}_{n-1}(s,x,v)| \,\mathrm{d}x.$$

The boundedness of f in (B.2) and repeated application of the above estimate lead to

$$\begin{aligned}
&\left| \int_0^T \max_v \int_{\mathbb{R}} f' \bar{g}_{>N} \,\mathrm{d}x \,\mathrm{d}t \right| \\
&\le \frac{T^2}{2} |V| C_K C_\phi e^{2|V|C_K T} (e^{C_K|V|T} - 1)^N \operatorname*{ess\,sup}_{s,v} \int_{\mathbb{R}} |\bar{g}_0(s,x,v)| \,\mathrm{d}x \\
&\le \frac{T^2}{2} |V| C_K C_\phi e^{2|V|C_K T} \left(e^{C_K|V|T} - 1\right)^N \operatorname*{ess\,sup}_{s,v} \int_{\mathbb{R}} |(\mu_2^\eta - \mu_1^\eta)(x+vs)| \,\mathrm{d}x \\
&\le T^2 |V| C_K C_\phi e^{2|V|C_K T} (e^{C_K|V|T} - 1)^N C_\mu,
\end{aligned}$$

where $|\bar{g}_0(s,x,v)| \le |(\mu_2^\eta - \mu_1^\eta)(x+vs)|$ can be observed from the explicit formula for $\bar{g}_0$ in (C.1). ∎

Appendix D. Proof of Lemmas in Section 4. We provide proofs for Lemma 4.8-4.9 in this section.

Proof of Lemma 4.8. Use the explicit representations

$$g_1^{(0)}(t,x,v) = e^{-(T-t)\sigma_1(v)} \mu_1(x + v(T-t)), \tag{D.1}$$
$$f^{(0)}(t,x,v) = e^{-t\sigma_1(v)} \phi(x - vt) \tag{D.2}$$

with $\sigma_1(v) = \int_V K_1(v',v)\,\mathrm{d}v'$ and set without loss of generality $c_1 = 0$. Since $f^{(0)}|_{I_1} = f_1^{(0)}$ in the notation of the proof of Proposition 4.3, one obtains for $(v,v') = (+1,-1)$

$$\begin{aligned}
&\int_0^T \int_{I_1} f^{(0)}(v')(g_1^{(0)}(v') - g_1^{(0)}(v)) \,\mathrm{d}x \,\mathrm{d}t \\
&= \int_0^T \int_{I_1} e^{-t\sigma_1(v')} \phi_1(x - v't) \Big(e^{-(T-t)\sigma_1(v')} \mu_1(x + v'(T-t)) \\
&\qquad\qquad - e^{-(T-t)\sigma_1(v)} \mu_1(x + v(T-t))\Big) \,\mathrm{d}x \,\mathrm{d}t \\
&\ge e^{-T\sigma_1(-1)} T \int_{a_0+T}^{a_1} \phi_1(x) \mu_1(-T+x) \,\mathrm{d}x - \int_{T-\frac{d_\mu+d}{2}}^T \int_{I_1} \phi_1(x) \mu_1(-T+x) \,\mathrm{d}x \,\mathrm{d}t \\
&\ge e^{-TC_K|V|} T C_{\phi\mu} - \frac{d_\mu + d}{2} C_{\phi\mu},
\end{aligned}$$

where the first inequality is due to the fact that $\phi_1(x - v't)\mu_1(x + v(T-t)) = \phi_1(x + t)\mu_1(x + (T-t)) \ne 0$ only for $x \in [-t-d, -t+d] \cap [-2T+t-d_\mu, -2T+t+d_\mu] \subset I_1$ which is empty for $t \le T - \frac{d_\mu+d}{2}$.

For $(v', v) = (-1, +1)$, instead, we obtain

$$\begin{aligned}
&\left|\int_0^T \int_{I_1} f^{(0)}(v)(g_1^{(0)}(v) - g_1^{(0)}(v'))\,\mathrm{d}x\,\mathrm{d}t\right| \\
&= \left| \int_0^T \int_{I_1} e^{-t\sigma_1(v)}\phi_1(x - vt)\big(e^{-(T-t)\sigma_1(v)}\mu_1(x + v(T - t)) \right. \\
&\qquad\qquad \left. - e^{-(T-t)\sigma_1(v')}\mu_1(x + v'(T - t))\big)\,\mathrm{d}x\,\mathrm{d}t\right| \\
&\le C_{\phi\mu}\frac{d + d_\mu}{2}
\end{aligned}$$

since
- $\phi_1(x - vt)\mu_1(x + v(T - t)) = \phi_1(x - t)\mu_1(x + T - t)$ vanishes, as its support $[t - d, t + d] \cap [-2T + t - d_\mu, -2T + t + d_\mu] = \varnothing$ is empty by construction of $T > d \ge d_\mu$ and
- the support $[t - d, t + d] \cap [-t - d_\mu, -t + d_\mu]$ of $\phi_1(x - vt)\mu_1(x + v'(T - t)) = \phi_1(x - t)\mu_1(x - (T - t))$ is non-empty only for $t \le \frac{d+d_\mu}{2}$.

Since $e^{-TC_K|V|} - \frac{d_\mu+d}{T} > 0$ by assumption, this proves the assertion. ∎

To show inequality (4.8) in Lemma 4.9, decompose for some $N \in \mathbb{N}$ to be determined later

$$\begin{aligned}
S = &\sum_{\substack{n,k=0\\ n+k\ge 1}}^{N} \int_0^T \int_{I_1} f^{(k)}(v')(g_1^{(n)}(v') - g_1^{(n)}(v))\,\mathrm{d}x\,\mathrm{d}t \\
&+ \int_0^T \int_{I_1} f(v')(g_1^{(>N)}(v') - g_1^{(>N)}(v))\,\mathrm{d}x\,\mathrm{d}t \qquad\text{(D.3)}\\
&+ \sum_{n=0}^{N} \int_0^T \int_{I_1} f^{(>N)}(v')(g_1^{(n)}(v') - g_1^{(n)}(v))\,\mathrm{d}x\,\mathrm{d}t\,,
\end{aligned}$$

where $g_1^{(n)}$ and $g_1^{(>N)}$ solve (3.14) and (3.15) respectively and $f^{(k)}$ are solutions to

$$\begin{aligned}
\partial_t f^{(k)} - v \cdot \nabla_x f^{(k)} &= \mathcal{L}(f^{(k-1)}) - \sigma f^{(k)}, \\
f^{(k)}(t = 0, x, v) &= 0,
\end{aligned}$$

with $\mathcal{L}(h) \coloneqq \int_V K(v, v')h(t, x, v')\,\mathrm{d}v'$, and $f^{(>N)}$ satisfies

$$\begin{aligned}
\partial_t f^{(>N)} - v \cdot \nabla_x f^{(>N)} &= \mathcal{L}(f^{(N)} + f^{(>N)}) - \sigma f^{(>N)}, \\
f^{(>N)}(t = 0, x, v) &= 0.
\end{aligned}$$

Each part of S in representation (D.3) is estimated separately in the subsequent three lemmas.

LEMMA D.1. *In the setting of proposition 4.7,*

$$\begin{aligned}
&\left|\int_0^T \int_{I_1} f^{(k)}(v')(g_1^{(n)}(v') - g_1^{(n)}(v))\,dx\,dt\right| \le 2\max_{v,v'} \int_0^T \int_{I_1} f^{(k)}(v')g_1^{(n)}(v)\,dx\,dt \\
&\le 2\,(C_K|V|)^{n+k}\,T^{n+k+1}C_{\phi\mu}
\end{aligned}$$

Proof. Source iteration

$$
\begin{aligned}
g_1^{(n)}(t,x,v_0) &= \int_0^{T-t}\int_V e^{-s_0\sigma(v_0)} K_1(\hat v_1, v_0) g_1^{(n-1)}(t+s_0, x+v_0 s_0, \hat v_1)\,\mathrm{d}\hat v_1\,\mathrm{d}s_0 \\
&\le |V| \int_0^{T-t} e^{-s_0\sigma(v_0)} K_1(v_1, v_0) g_1^{(n-1)}(t+s_0, x+v_0 s_0, v_1)\,\mathrm{d}s_0, \\
f^{(k)}(t,x,v_0) &= \int_0^{t}\int_V e^{-s_0\sigma(v_0)} K(v_0,\hat v_1) f^{(k-1)}(t-s_0, x-v_0 s_0, \hat v_1)\,\mathrm{d}\hat v_1\,\mathrm{d}s_0 \\
&\le |V| \int_0^{t} e^{-s_0\sigma(v_0)} K(v_0, v_1) f^{(k-1)}(t-s_0, x-v_0 s_0, v_1)\,\mathrm{d}s_0,
\end{aligned}
$$

where $v_1 = -v_0$, together with the explicit formulas (D.1)–(D.2) leads to estimates

(D.4)

$$
\begin{aligned}
0 \le g_1^{(n)}(x,t,v_0) \le (C_K|V|)^n \int_0^{T-t} \dots \int_0^{T-t-\sum_{i=0}^{n-2} s_i} \mu_1\left(x + \sum_{i=0}^{n-1} v_i s_i + v_n\left(T-t-\sum_{i=0}^{n-1} s_i\right)\right) \\
\mathrm{d}s_{n-1}\dots\mathrm{d}s_0,
\end{aligned}
$$

$$
0 \le f^{(k)}(x,t,v_0) \le (C_K|V|)^k \int_0^{t} \dots \int_0^{t-\sum_{i=0}^{k-2} s_i} \phi\left(x - \sum_{i=0}^{k-1} v_i s_i + v_k\left(t-\sum_{i=0}^{k-1} s_i\right)\right)\mathrm{d}s_{k-1}\dots\mathrm{d}s_0.
$$

Using again $f^{(k)}|_{I_1} = f_1^{(k)}$ with initial condition ϕ_1 in the notation of the proof of Porposition 4.3, this proves

$$
\begin{aligned}
&\left|\int_0^T\int_{I_1} f^{(k)}(v')(g_1^{(n)}(v') - g_1^{(n)}(v))\,\mathrm{d}x\,\mathrm{d}t\right| \le 2\max_{v,v'}\int_0^T\int_{I_1} f_1^{(k)}(v') g_1^{(n)}(v)\,\mathrm{d}x\,\mathrm{d}t \\
&\le 2\left(C_K|V|\right)^{n+k} T^{n+k+1} C_{\phi\mu}.
\end{aligned}
$$

∎

The following bound for the second summand in (D.3) is obtained in analogy to Lemma 3.12.

Lemma D.2. *In the setting of Proposition 4.7,*

$$
\begin{aligned}
&\max_v\left|\iint f(v')(g_1^{(>N)}(v') - g_1^{(>N)}(v))\,dx\,dt\right| \\
&\le 4T^2|V|C_K C_\phi e^{2|V|C_K T}(e^{C_K|V|T}-1)^N \bar C_\mu d_\mu =: C'(T)(e^{C_K|V|T}-1)^N
\end{aligned}
$$

For the third term in (D.3), one establishes the following bound.

Lemma D.3. *In the setting of Proposition 4.7,*

$$
\begin{aligned}
&\max_v\left|\iint f^{(>N)}(v')(g^{(n)}(v') - g^{(n)}(v))\,dx\,dt\right| \\
&\le 4|V|C_K T^2 e^{2|V|C_K T}(e^{C_K|V|T}-1)^N C_\phi \left(C_K|V|T\right)^n \bar C_\mu d_\mu \\
&=: C''(T)(e^{C_K|V|T}-1)^N \left(C_K|V|T\right)^n
\end{aligned}
$$

Proof. An estimate for $f^{(>N)}$ can be derived analogously as the estimate for $\bar g_{>N}$ in Lemma 3.12 from Lemma B.1

$$
\|f^{(>N)}\|_{L^\infty([0,T]\times\mathbb{R}\times V)} \le |V|C_K T e^{2|V|C_K T}(e^{C_K|V|T}-1)^N C_\phi.
$$

Together with (D.4), this proves the lemma. ∎

Lemma 4.9 can now be assembled from the previous lemmas.

Proof of Lemma 4.9. Lemmas D.1, D.2 and D.3 yield the (v, v') independent bound

$$\begin{aligned}
|S| &\le 2C_{\phi\mu}T \sum_{\substack{n,k=0\\ n+k\ge 1}}^{N} (C_K|V|T)^{n+k} + (e^{C_K|V|T}-1)^N \left(C'(T) + C''(T)\sum_{n=0}^{N}(C_K|V|T)^n\right)\\
&\le 4C_{\phi\mu}T \frac{C_K|V|T}{(1-C_K|V|T)^2} + (e^{C_K|V|T}-1)^N \left(C'(T) + C''(T)\frac{1}{1-C_K|V|T}\right)\\
&=: 4C_{\phi\mu}T \frac{C_K|V|T}{(1-C_K|V|T)^2} + (e^{C_K|V|T}-1)^N C(T).
\end{aligned}$$

Because $e^{C_K|V|T} - 1 < 1$ due to the assumption $T < (1-\delta)\frac{0.09}{C_K|V|}$, the second term in the last line becomes arbitrarily small for large $N \in \mathbb{N}$, which shows that $|S|$ is in fact bounded by the first term. ∎

REFERENCES

[1] G. Albi, E. Calzola, and G. Dimarco, A data-driven kinetic model for opinion dynamics with social network contacts, 2023, https://arxiv.org/abs/2307.00906.

[2] A. Alexanderian, Optimal experimental design for infinite-dimensional bayesian inverse problems governed by pdes: a review, Inverse Problems, 37 (2021), p. 043001, https://doi.org/10.1088/1361-6420/abe10c.

[3] W. Alt, Biased random walk models for chemotaxis and related diffusion approximations, Journal of Mathematical Biology, 9 (1980), pp. 147–177, https://doi.org/10.1007/BF00275919.

[4] T. Apel and T. G. Flaig, Crank–nicolson schemes for optimal control problems with evolution equations, SIAM Journal on Numerical Analysis, 50 (2012), pp. 1484–1512, https://doi.org/10.1137/100819333.

[5] S. R. Arridge and J. C. Schotland, Optical tomography: forward and inverse problems, Inverse Problems, 25 (2009), p. 123010, https://doi.org/10.1088/0266-5611/25/12/123010, https://dx.doi.org/10.1088/0266-5611/25/12/123010.

[6] G. Bal, I. Langmore, and F. Monard, Inverse transport with isotropic sources and angularly averaged measurements, Inverse Problems and Imaging, 1 (2008), pp. 23–42, https://doi.org/10.3934/ipi.2008.2.23.

[7] H. Berg, Random Walks in Biology, Princeton paperbacks, Princeton University Press, 1993, https://books.google.com/books?id=DjdgXGLoJY8C.

[8] M. Burger and W. Mühlhuber, Iterative regularization of parameter identification problems by sequential quadratic programming methods, Inverse Problems, 18 (2002), p. 943, https://doi.org/10.1088/0266-5611/18/4/301.

[9] J. Carrillo, M. Fornasier, J. Rosado, and G. Toscani, Asymptotic flocking dynamics for the kinetic cucker–smale model, SIAM Journal on Mathematical Analysis, 42 (2009), https://doi.org/10.1137/090757290.

[10] C. Cercignani, The Boltzmann Equation and Its Applications, Applied Mathematical Sciences, Springer New York, 2012, https://books.google.de/books?id=OcTcBwAAQBAJ.

[11] K. Chen, Q. Li, and J.-G. Liu, Online learning in optical tomography: a stochastic approach, Inverse Problems, 34 (2018), p. 075010, https://doi.org/10.1088/1361-6420/aac220.

[12] M. Choulli and P. Stefanov, Reconstruction of the coefficients of the stationary transport equation from boundary measurements, Inverse Problems, 12 (1996), pp. L19–L23, https://doi.org/10.1088/0266-5611/12/5/001.

[13] W. Chu, Q. Li, and M. A. Porter, Inference of interaction kernels in mean-field models of opinion dynamics, 2022, https://arxiv.org/abs/2212.14489.

[14] B. Davison and J. Sykes, Neutron Transport Theory, International series of monographs on physics, Clarendon Press ; [Oxford University Press], 1958.

[15] H. Egger, J.-F. Pietschmann, and M. Schlottbom, Identification of chemotaxis models with volume-filling, SIAM Journal on Applied Mathematics, 75 (2015), pp. 275–288, https://doi.org/10.1137/140967222.

[16] H. Egger and M. Schlottbom, Numerical methods for parameter identification in stationary radiative transfer, Computational Optimization and Applications, (2013), p. 67–83, https://doi.org/10.1007/s10589-014-9657-9.
[17] C. Emako, C. Gayrard, A. Buguin, L. Neves de Almeida, and N. Vauchelet, Traveling pulses for a two-species chemotaxis model, PLOS Computational Biology, 12 (2016), pp. 1–22, https://doi.org/10.1371/journal.pcbi.1004843, https://doi.org/10.1371/journal.pcbi.1004843.
[18] K.-J. Engel and R. Nagel, One-Parameter Semigroups for Linear Evolution Equations, vol. 63, Springer-Verlag New York, 06 2001, https://doi.org/10.1007/s002330010042.
[19] R. Erban and H. Othmer, From individual to collective behavior in bacterial chemotaxis, SIAM Journal of Applied Mathematics, 65 (2004), pp. 361–391, https://doi.org/10.1137/S0036139903433232.
[20] R. Ferrentino and C. Boniello, On the well-posedness for optimization problems: A theoretical investigation, Applied Mathematics, 10 (2019), pp. 19–38, https://doi.org/10.4236/am.2019.101003.
[21] F. Filbet and C. Yang, Numerical simulations of kinetic models for chemotaxis, SIAM Journal on Scientific Computing, 36 (2014), pp. B348–B366, https://doi.org/10.1137/130910208.
[22] K. Fister and M. Mccarthy, Identification of a chemotactic sensitivity in a coupled system, Mathematical medicine and biology : a journal of the IMA, 25 (2008), pp. 215–32, https://doi.org/10.1093/imammb/dqn015.
[23] R. M. Ford and D. A. Lauffenburger, Measurement of bacterial random motility and chemotaxis coefficients: Ii. application of single-cell-based mathematical model, Biotechnology and Bioengineering, 37 (1991), pp. 661–672, https://doi.org/10.1002/bit.260370708.
[24] A. Giometto, F. Altermatt, A. Maritan, R. Stocker, and A. Rinaldo, Generalized receptor law governs phototaxis in the phytoplankton euglena gracilis, Proceedings of the National Academy of Sciences, 112 (2015), pp. 7045–7050, https://doi.org/10.1073/pnas.1422922112.
[25] M. D. Gunzburger, Perspectives in Flow Control and Optimization, Society for Industrial and Applied Mathematics, 2002, https://doi.org/10.1137/1.9780898718720.
[26] E. Haber, U. M. Ascher, and D. Oldenburg, On optimization techniques for solving nonlinear inverse problems, Inverse Problems, 16 (2000), p. 1263, https://doi.org/10.1088/0266-5611/16/5/309, https://dx.doi.org/10.1088/0266-5611/16/5/309.
[27] K. Hellmuth, C. Klingenberg, Q. Li, and M. Tang, Kinetic chemotaxis tumbling kernel determined from macroscopic quantities, 2022, https://arxiv.org/abs/2206.01629.
[28] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, Optimization with pde constraints, in Mathematical Modelling, 2008.
[29] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, 1985, https://doi.org/10.1017/CBO9780511810817.
[30] H. Jeckel, E. Jelli, R. Hartmann, P. K. Singh, R. Mok, J. F. Totz, L. Vidakovic, B. Eckhardt, J. Dunkel, and K. Drescher, Learning the space-time phase diagram of bacterial swarm expansion, Proceedings of the National Academy of Sciences, 116 (2019), pp. 1489–1494, https://doi.org/10.1073/pnas.1811722116.
[31] B. Jin and Z. Zhou, Error analysis of finite element approximations of diffusion coefficient identification for elliptic and parabolic problems, SIAM Journal on Numerical Analysis, 59 (2021), pp. 119–142.
[32] C. Kurzthaler, Y. Zhao, N. Zhou, J. Schwarz-Linek, C. Devailly, J. Arlt, J.-D. Huang, W. C. K. Poon, T. Franosch, J. Tailleur, and V. A. Martinez, Characterization and control of the run-and-tumble dynamics of escherichia coli, Phys. Rev. Lett., 132 (2024), p. 038302, https://doi.org/10.1103/PhysRevLett.132.038302.
[33] R.-Y. Lai, Q. Li, and G. Uhlmann, Inverse problems for the stationary transport equation in the diffusion scaling, SIAM Journal on Applied Mathematics, 79 (2019), pp. 2340–2358, https://doi.org/10.1137/18M1207582.
[34] D. Le, Dynamics of a bio-reactor model with chemotaxis, Journal of Mathematical Analysis and Applications, 275 (2002), pp. 188–207, https://doi.org/10.1016/S0022-247X(02)00313-X.
[35] H. Li, X. qing Shi, M. Huang, X. Chen, M. Xiao, C. Liu, H. Chaté, and H. P. Zhang, Data-driven quantitative modeling of bacterial active nematics, Proceedings of the National Academy of Sciences, 116 (2019), pp. 777–785, https://doi.org/10.1073/pnas.1812570116.
[36] Q. Li and W. Sun, Applications of kinetic tools to inverse transport problems, Inverse Problems, 36 (2020), p. 035011, https://doi.org/10.1088/1361-6420/ab59b8.
[37] J. Liu and Z. Wang, Non-commutative discretize-then-optimize algorithms for elliptic pde-constrained optimal control problems, Journal of Computational and Applied Mathematics, 362 (2019), pp. 596–613, https://doi.org/https://doi.org/10.1016/j.cam.2018.07.

028.

[38] S. Motsch and E. Tadmor, Heterophilious dynamics enhances consensus, SIAM Review, 56 (2014), pp. 577–621, https://doi.org/10.1137/120901866.

[39] H. Othmer, S. Dunbar, and W. Alt, Models of dispersal in biological systems, Journal of mathematical biology, 26 (1988), pp. 263–98, https://doi.org/10.1007/BF00277392.

[40] H. Othmer and T. Hillen, The diffusion limit of transport equations ii: Chemotaxis equations, SIAM Journal of Applied Mathematics, 62 (2002), pp. 1222–1250, https://doi.org/10.1137/S0036139900382772.

[41] O. Pohl, M. Hintsche, Z. Alirezaeizanjani, M. Seyrich, C. Beta, and H. Stark, Inferring the chemotactic strategy of p. putida and e. coli using modified kramers-moyal coefficients, PLOS Computational Biology, 13 (2017), pp. 1–24, https://doi.org/10.1371/journal.pcbi.1005329.

[42] B. Polyak and P. Shcherbakov, Lyapunov functions: An optimization theory perspective, IFAC-PapersOnLine, 50 (2017), pp. 7456–7461, https://doi.org/https://doi.org/10.1016/j.ifacol.2017.08.1513. 20th IFAC World Congress.

[43] K. Prieto and O. Dorn, Sparsity and level set regularization for diffuse optical tomography using a transport model in 2d, Inverse Problems, 33 (2016), p. 014001, https://doi.org/10.1088/0266-5611/33/1/014001.

[44] K. Ren, Recent developments in numerical techniques for transport-based medical imaging methods, Communications in Computational Physics, 8 (2010), pp. 1–50, https://global-sci.org/intro/article_detail/cicp/7562.html#.

[45] G. B. Rybicki and A. P. Lightman, Radiative Processes in Astrophysics, WILEY-VCH, 1986.

[46] M. Salek, F. Carrara, V. Fernandez, J. Guasto, and R. Stocker, Bacterial chemotaxis in a microfluidic t-maze reveals strong phenotypic heterogeneity in chemotactic sensitivity, Nature Communications, 10 (2019), https://doi.org/10.1038/s41467-019-09521-2.

[47] J. Saragosti, V. Calvez, N. Bournaveas, A. Buguin, P. Silberzan, and B. Perthame, Mathematical description of bacterial traveling pulses, PLOS Computational Biology, 6 (2010), pp. 1–12, https://doi.org/10.1371/journal.pcbi.1000890, https://doi.org/10.1371/journal.pcbi.1000890.

[48] 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 (2011), pp. 16235–16240, https://doi.org/10.1073/pnas.1101996108, https://www.pnas.org/doi/abs/10.1073/pnas.1101996108, https://arxiv.org/abs/https://www.pnas.org/doi/pdf/10.1073/pnas.1101996108.

[49] M. Seyrich, Z. Alirezaeizanjani, C. Beta, and H. Stark, Statistical parameter inference of bacterial swimming strategies, New Journal of Physics, 20 (2018), p. 103033, https://doi.org/10.1088/1367-2630/aae72c.

[50] D. Smyl, T. N. Tallman, D. Liu, and A. Hauptmann, An efficient quasi-newton method for nonlinear inverse problems via learned singular values, IEEE Signal Processing Letters, 28 (2021), pp. 748–752, https://doi.org/10.1109/LSP.2021.3063622.

[51] P. Stefanov and Y. Zhong, Inverse boundary problem for the two photon absorption transport equation, SIAM Journal on Mathematical Analysis, 54 (2022), pp. 2753–2767.

[52] J. Taylor-King, E. Loon, G. Rosser, and S. Chapman, From birds to bacteria: Generalised velocity jump processes with resting states, Bulletin of mathematical biology, 77 (2014), p. 1213–1236, https://doi.org/10.1007/s11538-015-0083-7.

[53] A. Tikhonov, On the stability of the functional optimization problem, USSR Computational Mathematics and Mathematical Physics, 6 (1966), pp. 28–33, https://doi.org/https://doi.org/10.1016/0041-5553(66)90003-6.

[54] G. Toscani, Kinetic models of opinion formation, Commun. Math. Sci., 4 (2006), pp. 481–496, https://doi.org/10.4310/CMS.2006.v4.n3.a1.

[55] 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 (1988), pp. 1–15, https://doi.org/10.1002/cm.970110102.

[56] S. J. Wright and B. Recht, Optimization for Data Analysis, Cambridge University Press, 2022, https://doi.org/10.1017/9781009004282.

[57] H. P. Zhang, A. Be'er, E.-L. Florin, and H. L. Swinney, Collective motion and density fluctuations in bacterial colonies, Proceedings of the National Academy of Sciences, 107 (2010), pp. 13626–13630, https://doi.org/10.1073/pnas.1001651107.

[58] Y. Zhao, C. Kurzthaler, N. Zhou, J. Schwarz-Linek, C. Devailly, J. Arlt, J.-D. Huang, W. C. K. Poon, T. Franosch, V. A. Martinez, and J. Tailleur, Quantitative characterization of run-and-tumble statistics in bulk bacterial suspensions, Phys. Rev. E, 109 (2024), p. 014612, https://doi.org/10.1103/PhysRevE.109.014612.