SIAM J. APPL. MATH.
Vol. 85, No. 2, pp. 613–635

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 a 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 evidence supports the theoretical findings.

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

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

DOI. 10.1137/24M165507X

1. Introduction.

The 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, 18, 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 $v\in V$. 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; 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

*Received by the editors April 18, 2024; accepted for publication (in revised form) November 5, 2024; published electronically March 20, 2025.
https://doi.org/10.1137/24M165507X
Funding: The work of the first author was supported by the German Academic Scholarship Foundation (Studienstiftung des Deutschen Volkes) and by the Marianne-Plehn-Program. The work of the third author was partially supported by Vice Chancellor for Research and Graduate Education grants, and NSF-DMS-2308440 and ONR-N000142112140. The work of the fourth author was partially supported by the Strategic Priority Research Program of Chinese Academy of Sciences grants 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 53705 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/).

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, 47, 48].

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 (bird) 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 problems within the domain of kinetic systems is optical tomography emerging from medical imaging, where nonintrusive boundary data is deployed to map out the optical properties of the interior. Mathematically a 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 the 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 implement 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 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 is 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 inverse 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 [22, 23, 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: Similarly as was done in [15, 21] 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 nonregularized manner. This allows investigation of structural identifiability as well as suitability of specific experimental setups 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 $(x, v) \in \mathbb{R}^1 \times \{\pm 1\}$ [23, 48, 47], i.e., the bacteria moves either to the left or to the right, and x is one-dimensional (1D) in space. This simple setting reflects how experiments are conducted in the labs: bacteria are cultured in a tube, and the motion is 1D. More details will be discussed in the subsequent part.

In a numerical setting, we first represent K as a finite dimensional parameter. After prescribing a partition of the domain $\mathbb{R}^1 = \bigcup_r I_r$ into intervals $I_r = [a_{r-1}, a_r)$, for $r = 2, \dots, R-1$, and $I_1 = (-\infty, a_1)$, $I_R = [a_{R-1}, \infty)$, the function $K(x, v, v')$ is approximated within the cell I_r by the value $K_r(v, v')$, constant in space:

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

where $\mathbb{1}_I$ denotes the characteristic function of a subset $I \subset \mathbb{R}^1$, i.e., $\mathbb{1}_I(x) = 1$ if $x \in I$ and 0 otherwise. For $v = v'$, we set $K_r(v, v) = 0$ since these tumbling events cannot be distinguished from running a straight line and do not effect the motion (1.1). Since $V = \{\pm 1\}$, then 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

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

for $i = 1, 2$. The dataset is also finite in size. In particular, we mathematically represent each 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

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

where f_K denotes the solution to (1.1) with kernel K. Integration in velocity before testing with μ_l means that only the macroscopic density can be accessed. Even though integration amounts in a simple summation in our 1D setting with $V = \{\pm 1\}$, we stick to this notation for conciseness of the representation and set $|V| := \int_V \mathrm{d}v$. In case μ_l is a characteristic function, this corresponds to the pixel reading of a photo.

For simplicity of the presentation, the ground-truth kernel $K_\star$ 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

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

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

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 (GD) 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 ($i = 1,2,\ r = 1,\dots,R$) entry in the gradient $\nabla_K \mathcal{C}$:

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

Details are given in section SM1 in the supplementary materials. In the formulation, (v_i, v_i') is given in (2.2) and g is the adjoint state that solves the adjoint equation

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

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

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 including 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:

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

  $$0 < \|\hat{K}\|_\infty \le C_K \qquad \text{for all } \hat{K} \in \mathcal{U}_{K_\star}. \tag{2.10}$$

- Assume that the initial data ϕ is in the space $L^\infty_+(\mathbb{R}\times V) \cap L^1(\mathbb{R}; L^\infty(V))$ of nonnegative functions with essential bound

  $$\|\phi\|_{L^\infty(\mathbb{R}\times V)}, \|\phi\|_{L^1(\mathbb{R};L^\infty(V))} \le C_\phi\,.$$

- 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

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

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 the adjoint solution; see Lemmas SM2.1 and SM2.3 in supplementary materials section SM2.

3. Well-posedness versus 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 toward 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.

Theorem 3.2. *Assume the Hessian matrix of the cost function is positive definite at $K_\star$ and let the remaining assumptions of Proposition* 3.5 *hold; then there exists a neighborhood U of $K_\star$, in which the optimization problem* (2.5) *is Tikhonov well-posed. In particular, the GD 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, with $L = 2R$, for which we can ensure this to hold.

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 $M_1(K)$ that converges to the test function μ_2 of the second measurement $M_2(K)$ as $m \to \infty$, either*

(1) *$\mu_{1,m} \to \mu_2$ strongly in L^1, or*
(2) *$\mu_{1,m} \rightharpoonup \mu_2$ weakly in L^1. In this case, further assume that the measurement time T is chosen small such that $e^{T|V|C_K} < 2$.*

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 GD algorithm (2.6) *to $K_\star$ cannot be guaranteed.*

The two scenarios describe two different qualities of convergence:

(1) strong convergence in the L^1 norm $\|\mu_{1,m} - \mu_2\|_{L^1} \to 0$, and
(2) weak convergence in L^1, a distributional property that requires a test function $h \in L^\infty$ to observe convergence $\int \mu_{1,m}(x)h(x)\,\mathrm{d}x \to \int \mu_2(x)h(x)\,\mathrm{d}x$.

The two different settings require different proof techniques, and these technical constraints are the main reason for (2) to require additional assumptions on T.

Remark 3.4. In the case of pointwise measurement, the test functions μ_i are Dirac delta measures. Since $\delta_{x_i} \notin L^1(\mathbb{R})$, the case is not covered in our setting. However, a small modification to the proof lets us handle the situation where μ_i is a mollification of Dirac. If the mollification parameters are independent of i, having $x_{1,m} \to x_2$ puts us back in the setting of the theorem, and the result still holds.

The two theorems, to be proved in detail in sections 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:

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

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 of the parameter reconstruction. 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 3.5, 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. Propositions 3.10 and 3.11 consider the two different notions of closeness, each of which entails its own strategy to control the vanishing information gain from the first measurement.

3.1. Local well-posedness of the optimization problem. Generally speaking, it would not be easy to characterize the full landscape of the loss function and thus it is hard to prescribe conditions for obtaining global convergence. However, suppose the data is prepared well enough 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 local well-posedness.

PROPOSITION 3.5. *Let Assumption* 2.1 *hold. Assume the Hessian $H_K\mathcal{C}(K_\star)$ is positive definite at $K_\star$ and that there exists a neighborhood $\mathcal{U}_{K_\star}$ of $K_\star$ in which the Hessian of the measurements is uniformly bounded in the Frobenius norm, i.e., for all $l = 1,\dots,L$ and $K \in \mathcal{U}_{K_\star}$ one has $\|H_K M_l(K)(v,v')\|_F \le C_{H_K M}$. Then there exists a (bounded) neighborhood $U \subset \mathcal{U}_{K_\star}$ of $K_\star$ in the L^∞ norm, where $H_K\mathcal{C}(K)$ is positive definite for all $K \in U$. Moreover, the minimal eigenvalues $\lambda_{\min}(H_K\mathcal{C})$ satisfy*

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

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 $C_{H_K M}$, but not on K. 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 a strategy will be spelled out in detail in the proof. Now Theorem 3.2 is immediate.

Proof for Theorem 3.2. By Proposition 3.5, there exists a neighborhood 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) [19, Prop. 3.1]. Convergence of GD follows from strong convexity of $\mathcal{C}$ in U. ∎

Now we give the proof for Proposition 3.5. It mostly relies on the matrix perturbation theory [29, Cor. 6.3.8] and continuity of (1.1) with respect to the parameter K.

Proof for Proposition 3.5. According to the matrix perturbation theory, the minimal eigenvalue is continuous with respect to a perturbation to the matrix; we have

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

where we used the Hessian form (3.1), triangle inequality, and submultiplicativity 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 3.6, 3.8, and 3.9 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

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

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

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

As can be seen from the proof, Proposition 3.5 strongly relies on the boundedness of the terms in (3.3). We present the estimates below.

LEMMA 3.6. *Let Assumption* 2.1 *hold; 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 SM2.1 to the difference equation for $\bar{f} := f_{K_\star} - f_K$,

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

with initial condition 0 and source term $h = \mathcal{K}_{(K_\star-K)}(f_{K_\star}) \in L^1((0,T); L^\infty(\mathbb{R}\times V)\cap L^1(\mathbb{R}; L^\infty(V)))$ by the regularity (SM2.1) of $f_{K_\star}$. This leads to

$$
\begin{aligned}
\operatorname*{ess\,sup}_{v,x} |\bar{f}|(x,t,v) &\le \int_0^t e^{2|V|C_K(t-s)} \|\mathcal{K}_{(K_\star-K)}(f_{K_\star})(s)\|_{L^\infty(\mathbb{R}\times V)\cap L^1(\mathbb{R};L^\infty(V))}\,\mathrm{d}s \\
&\le 2|V|\|K_\star - K\|_\infty e^{2|V|C_K t} C_\phi t,
\end{aligned} \tag{3.5}
$$

where we used the estimate $\|f_{K_\star}(s)\|_{L^\infty(\mathbb{R}\times V)\cap L^1(\mathbb{R};L^\infty(V))} \le e^{2|V|C_K s}C_\phi$ from Lemma SM2.1 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 3.7. *Let Assumption* 2.1 *hold. Denote by* f_K *the mild solution of* (1.1) *and by* $g_l \in C^0([0,T];L^\infty(V;L^1(\mathbb{R})))$ *the mild solution of*

$$
\begin{aligned}
&-\partial_t g_l - v\cdot\nabla g_l = \tilde{\mathcal{K}}(g_l) := \int_V K(x,v',v)(g_l(x,t,v') - g_l(x,t,v))\,\mathrm{d}v', \\
&g_l(t=T,x,v) = -\mu_l(x)\,.
\end{aligned}
\tag{3.6}
$$

Then

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

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 3.8. *Under Assumption* 2.1*, the gradient is uniformly bounded,*

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

Proof. The Frobenius norm is bounded by the entries

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

Representation (3.7) together with (SM2.2) then gives the bound

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

Application of Lemma SM2.3 to $g = g_l$, with $h = 0$ 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 3.9. *In the setting of Theorem* 3.2 *and under Assumption* 2.1*, the gradient difference is uniformly bounded in* $K\in\mathcal{U}_{K_\star}$ *by*

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

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 3.7 and (SM2.2),

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

The first summand can be bounded by (3.5) and (3.9). To estimate the second summand, apply Lemma SM2.3 to $\bar{g} := g_{l,K_\star} - g_{l,K}$ 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 (SM2.4) 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

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

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 3.5.

3.2. Ill-conditioning for close measurements. While the positive Hessian at K_* 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 3.6, 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 $\|\nabla_K M_1(K) - \nabla_K M_2(K)\|_F \approx 0$ when $\mu_1 \approx \mu_2$.

Throughout the derivation, the following formula is important. Recalling (3.7), we have for every $r \in \{1, \ldots, R\}$ 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. Closeness in the strong sense. The following proposition states the loss of strong convexity as $\mu_2 - \mu_{1,m} \to 0$ in $L^1(\mathbb{R})$. In particular, the requirement of Proposition 3.5 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 3.10. *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 the assertion of Theorem 3.3 for L^1 closeness.

Proof of Theorem 3.3. Proposition 3.10 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 $K_\star$. ∎

Proof of Proposition 3.10. As argued above, our goal is to show $\|\nabla_K M_{1,m}(K) - \nabla_K M_2(K)\|_F \to 0$ as $m \to \infty$. Recalling (3.10), we need to show

$$
\frac{\partial M_{1,m}(K)}{\partial K_{r,i}} - \frac{\partial M_2(K)}{\partial K_{r,i}} \xrightarrow{m\to\infty} 0 \quad \text{for all } (r,i) \in \{1, \ldots, R\} \times \{1, 2\}\,, \qquad (3.11)
$$

where $\bar{g}_m := g_{1,m} - g_2$ solves (3.6) with final condition $\bar{g}_m(t = T, x, v) = \mu_2(x) - \mu_{1,m}(x)$. Application of Lemma SM2.3 gives

$$
\|\bar{g}_m(t)\|_{L^\infty(V;L^1(\mathbb{R}))} \le e^{2C_K|V|(T-t)}\|\mu_2 - \mu_{1,m}\|_{L^1(\mathbb{R})} \qquad (3.12)
$$

by independence of $\mu_{1,m}, \mu_2$ with respect to v. Plug the above into (3.10) and estimate f by (SM2.2) to obtain

$$
\begin{aligned}
\left|\frac{\partial(M_{1,m} - M_2)(K)}{\partial K_{r,i}}\right| &\le 2C_\phi \int_0^T e^{2C_K|V|t}\|\bar{g}_m(t)\|_{L^\infty(V;L^1(\mathbb{R}))}\,\mathrm{d}t \\
&\le 2C_\phi e^{2C_K|V|T} T \|\mu_2 - \mu_{1,m}\|_{L^1(\mathbb{R})}.
\end{aligned}
$$

Since every entry (r, i) converges, the gradient difference vanishes: $\|\nabla_K M_{1,m}(K) - \nabla_K M_2(K)\|_F \to 0$ as $m \to \infty$.

We utilize this fact to show the degeneracy of the Hessian. Note

$$H_K\mathcal{C}(K_\star) = \underbrace{\left[\sum_{l=3}^{2R} \nabla M_l \otimes \nabla M_l + 2\nabla M_2 \otimes \nabla M_2\right]}_{A} + \underbrace{\left[\nabla M_{1,m} \otimes \nabla M_{1,m} - \nabla M_2 \otimes \nabla M_2\right]}_{B^{(m)}}.$$

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

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

3.2.2. Closeness in the weak sense. We now study the more general scenario of Theorem 3.3 where $\mu_{1,m}$ converges weakly in L^1. Because we no longer obtain smallness of $\bar{g}$ in the strong sense as in (3.12), the additional assumption of small measurement time is required to provide smallness in a weak sense.

PROPOSITION 3.11. *Let $\mu_{1,m} \rightharpoonup \mu_2$ weakly in L^1. Consider a small neighborhood of $K_\star$ and let Assumption* 2.1 *hold. Additionally, let the measurement time T be chosen such that $(e^{T|V|C_K} - 1) < 1$. Then*

$$\nabla_K M_{1,m}(K) \to \nabla_K M_2(K) \qquad \text{as } m \to \infty \text{ in the standard Euclidean norm.}$$

This proposition explains the breakdown of well-posedness presented in Theorem 3.3 for weakly convergent measurement test functions. Since the proof for the theorem is rather similar to that of the first scenario, we omit it 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 $\bar{g}_m$, 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 $\bar{g}_m^{(>N)}$ of $\bar{g}_m$ represent the difference of $\bar{g}_m$ and its singular parts and evolve from the integral operator on the right-hand side of (2.8), which exhibits a diffusive effect. Smallness is obtained by adjusting the cutoff 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.

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

$$\bar{g}_m = \sum_{n=0}^{N} \bar{g}_m^{(n)} + \bar{g}_m^{(>N)}, \tag{3.13}$$

where the regularity of $\bar{g}_m^{(n)}$ increases with n. Here, we define $\bar{g}_m^{(0)}$ as the solution to

$$\begin{aligned} -\partial_t \bar{g}_m^{(0)} - v \cdot \nabla_x \bar{g}_m^{(0)} &= -\sigma \bar{g}_m^{(0)}\,, \\ \bar{g}_m^{(0)}(t = T, x, v) &= \mu_2(x) - \mu_{1,m}(x)\,, \end{aligned}$$

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

$$-\partial_t \bar{g}_m^{(n)} - v\cdot\nabla_x \bar{g}_m^{(n)} = -\sigma\bar{g}_m^{(n)} + \tilde{\mathcal{L}}(\bar{g}_m^{(n-1)})\,,\qquad \bar{g}_m^{(n)}(t=T,x,v) = 0\,, \tag{3.14}$$

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

$$-\partial_t \bar{g}_m^{(>N)} - v\cdot\nabla_x \bar{g}_m^{(>N)} = -\sigma\bar{g}_m^{(>N)} + \tilde{\mathcal{L}}(\bar{g}_m^{(N)} + \bar{g}_m^{(>N)})\,,\qquad \bar{g}_m^{(>N)}(t=T,x,v) = 0\,. \tag{3.15}$$

It is a straightforward calculation that

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

We are to show, in the two lemmas below, that both terms are small when $\mu_{1,m} \rightharpoonup \mu_2$.

Lemma 3.12. *Let the assumptions of Proposition* 3.11 *be satisfied. For any $\varepsilon > 0$, and any $n \in \mathbb{N}_0$, there exists a $\tilde{m}_n(\varepsilon) \in \mathbb{N}$ such that*

$$\left|\int_0^T\int_{I_r} f'\bar{g}_m^{(n)}\,\mathrm{d}x\,\mathrm{d}t\right| \le \varepsilon \quad \textit{if} \quad m \ge \tilde{m}_n(\varepsilon)\,. \tag{3.17}$$

The remainder can be bounded similarly.

Lemma 3.13. *Under the assumptions of Proposition* 3.11*, one has*

$$\left|\int_0^T\int_{I_r} f'\bar{g}_m^{(>N)}\,\mathrm{d}x\,\mathrm{d}t\right| \le T^2|V|C_K C_\phi e^{2|V|C_K T}(e^{C_K|V|T}-1)^N C_\mu,$$

which becomes arbitrarily small for large N.

The proof for Lemma 3.12 exploits the smallness of $\bar{g}_m^{(n)}$ in a weak sense which is inherited from the final condition, whereas Lemma 3.13 is based on the smallness of the higher regularity components of $\bar{g}_m$ in the small time regime where tumbling is not so frequent. Since it is not essential to the core of the paper, we leave the details to section SM3 in the supplementary materials. The application of the two lemmas gives Proposition 3.11.

Proof of Proposition 3.11. 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 $m \ge \max_{n=0,\dots,N}\{\tilde{m}_n(\frac{\varepsilon}{4(N+1)})\}$. Then with the triangle inequality and Lemmas 3.12 and 3.13, we obtain from (3.16)

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

4. Experimental design. We now provide an explicit experimental setup that ensures well-posedness with a minimal number of measurements $L = 2R$. Recalling that Proposition 3.5 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 experimental domain $I = [a_0, a_R)$ into R intervals $I = \cup_{r=1}^{R} I_r$ 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 $K_r(v, v')$ take on the form of* (2.1). *The design is as follows:*

- *initial condition $\phi(x, v) = \sum_{r=1}^{R} \phi_r(x)$ is a sum of R positive functions ϕ_r that are compactly supported in $a_{r-1/2} + [-d, d]$ with $d < \min(\frac{a_r - a_{r-1}}{4})$, symmetric and monotonously decreasing in $|x - a_{r-1/2}|$ (for instance, a centered Gaussian with a cutoff tail);*
- *measurement test functions $\mu_{l_i^r} = \bar{C}_\mu \mathbb{1}_{[x_{l_i^r} - d_\mu, x_{l_i^r} + d_\mu]}$, $i = 1, 2$, for some $\bar{C}_\mu > 0$, centered around $x_{l_i^r} := a_{r-1/2} + (-1)^i T$ with $d_\mu \le d$;*
- *measurement time 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. This places the experiment in a regime where tumbling does not occur too frequently. Indeed, when the system tumbles too often, one only obtains an accumulation of influence of many tumbling events, making it difficult to separate out each parameter. The same strategy was deployed in our theoretical paper [27]. However, the specific bound for T may not be optimal. In the proof of Theorem 4.4, only crude estimates were deployed. It is possible to relax the bound for T.

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

Under this design, we have the following proposition.

PROPOSITION 4.3. *Design* 4.1 *decouples the reconstruction of K_r. To be more specific, recall* (2.2)

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

The Hessian $H_K\mathcal{C}$ has a block diagonal structure where 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), its solution $f = \sum_{s=1}^{R} f_s$ decomposes into solutions f_s of (1.1) with initial conditions ϕ_s. By construction of T and the constant speed of propagation $|v| = 1$, the spatial supports of the f_s are fully contained in I_s for all $t \in [0, T], v \in V$. As such, only f_r carries information about K_r, and no information for other K_s with $s \neq r$. Because f_r is only measured by measurements $M_{l_i^r}$, $i = 1, 2$, only the gradients of these measurements can attain a nonzero value corresponding to the partial derivative with respect to K_r. $\square$

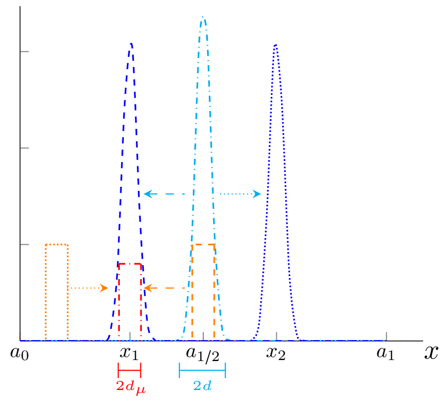

FIG. 1. *Motion of the ballistic parts* $f^{(0)}(t = 0, v)$ *(cyan, dashdotted) to* $f^{(0)}(t = T, v = +1)$ *(blue, dotted) and* $f^{(0)}(t = T, v = -1)$ *(blue, dashed) and* $g_1^{(0)}(t = 0, v = +1)$ *(orange, dotted) and* $g_1^{(0)}(t = 0, v = -1)$ *(orange, dashed) to* $g_1^{(0)}(t = T, v)$ *(red, dashdotted); compare also* (4.5). *(Color figures are available online.)*

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 neighborhood of* $K_\star$*, then Design* 4.1 *generates a locally well-posed optimization problem* (2.5).

The proof is layed out in subsection 4.1. As it relies on a perturbative argument, only local reconstructability can be proven. However, numerical experiments—such as those displayed in Figures 2 and 4—exhibited reconstructability in a wide range of parameters values, pointing toward global reconstructability with this design.

Remark 4.5. Design 4.1 shares similarities with the theoretical reconstruction setting in [27]: local reconstructions are based on measurements close to the considered location and small time information is needed. In particular, the pointwise reconstruction of K in [26] relies on a sequence of experiments where the measurement time asymptotically vanishes 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 a 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, since it can only provide information on σ. Because there exist only two directions $V = \{\pm 1\}$ in one dimension, there is a one-to-one correspondence of σ and K and ballistic data is sufficient for the reconstruction.

4.1. Proof of Theorem 4.4. According to Theorem 3.2, it is sufficient to show $H_K \mathcal{C}(K_\star) > 0$. As the Hessian attains a block diagonal structure (Proposition 4.3), we 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 4.6. *Let Assumption* 2.1 *hold. Then Design* 4.1 *produces a positive definite Hessian block* $H_{K_1}\mathcal{C}(K_\star)$.

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

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

hold 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 4.1.

Lemma 4.7. *In the setting of Proposition* 4.6, *for* $(v,v') = (+1,-1)$, *the ballistic part*

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

satisfies

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

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

At the same time, the remainder term is small.

Lemma 4.8. *In the setting of Proposition* 4.6, *the remaining scattering term*

$$S := \int_0^T \int_{I_1} f(v')(g_1(v') - g_1(v))\,\mathrm{d}x\,\mathrm{d}t - \int_0^T \int_{I_1} f^{(0)}(v')(g_1^{(0)}(v') - g_1^{(0)}(v))\,\mathrm{d}x\,\mathrm{d}t$$

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 4.6 is a corollary of Lemmas 4.7 and 4.8.

Proof of Proposition 4.6. By the bounds obtained in Lemmas 4.7 and 4.8, one has

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

where the estimate $e^{-TC_K|V|} \leq 1 - TC_K|V|$ was used to derive the last line. This holds due to smallness of $TC_K|V| < 0.09(1-\delta) < 1$ by construction (4.1)–(4.2), which also provides positivity of the emerging term. 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 4.6.

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 neighborhood of $K_\star$, Theorem 3.2 proves local well-posedness of the inverse problem. ◻

The proofs for Lemmas 4.7 and 4.8 are rather technical and we leave them to section SM4 in the supplementary materials. 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 span $T \approx T$ 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 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 [20] 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 that 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 locally 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, Thm. 3], can be observed.

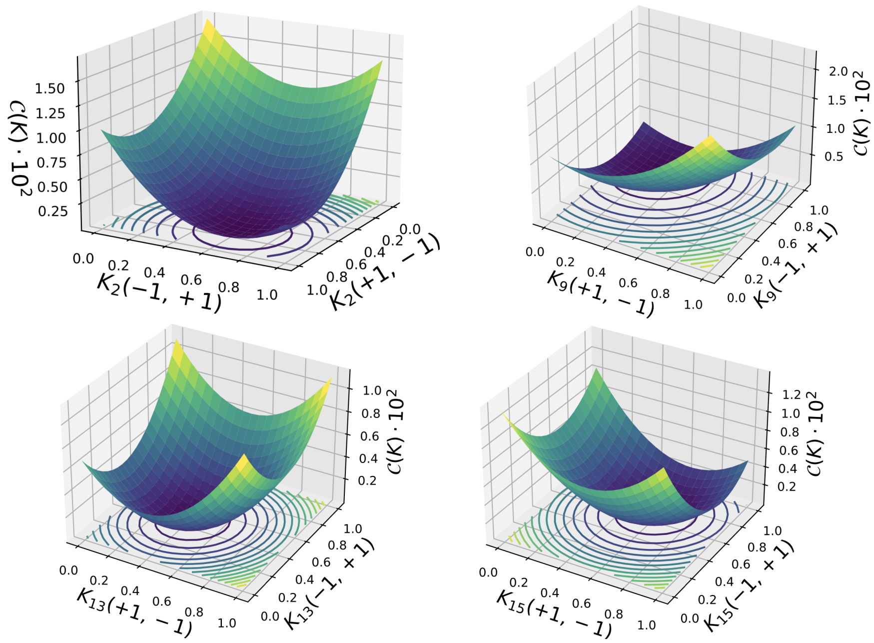

FIG. 2. *(Marginal) loss functions $\mathcal{C}(K)$ for $R=20$: For a fixed $r \in \{2,9,13,15\}$, we plot $\mathcal{C}$ as a function of K_r with all $K_{s\neq r}$ set to be the ground-truth $(K_\star)_s$.*

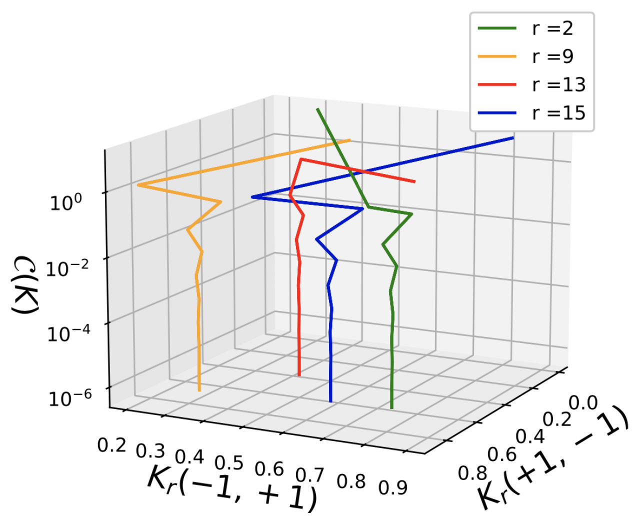

FIG. 3. *Convergence of the parameter values $K_r(v,v')$ from* (2.1) *for $r = 2, 9, 13, 15$ to the ground-truth as the cost function converges.*

The strictly positive definiteness feature persists in a small neighborhood of the optimal solution $K_\star$. This means that by 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 (located at 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,

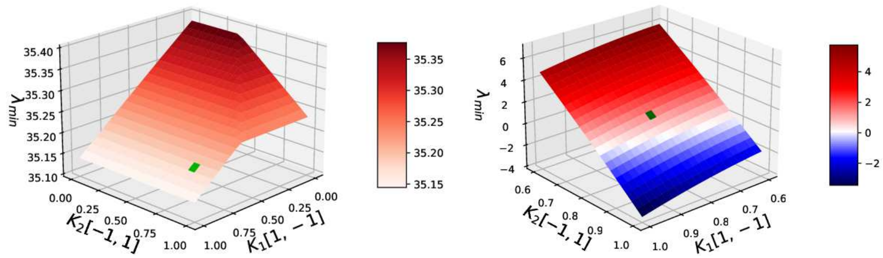

FIG. 4. *Minimal eigenvalues of the Hessian $H_K\mathcal{C}(K)$ around the true parameter $K_\star$ for two experimental designs. We perturb K by changing values in $K_1(1,-1)$ and $K_2(-1,1)$. The ground-truth is marked green in both plots.*

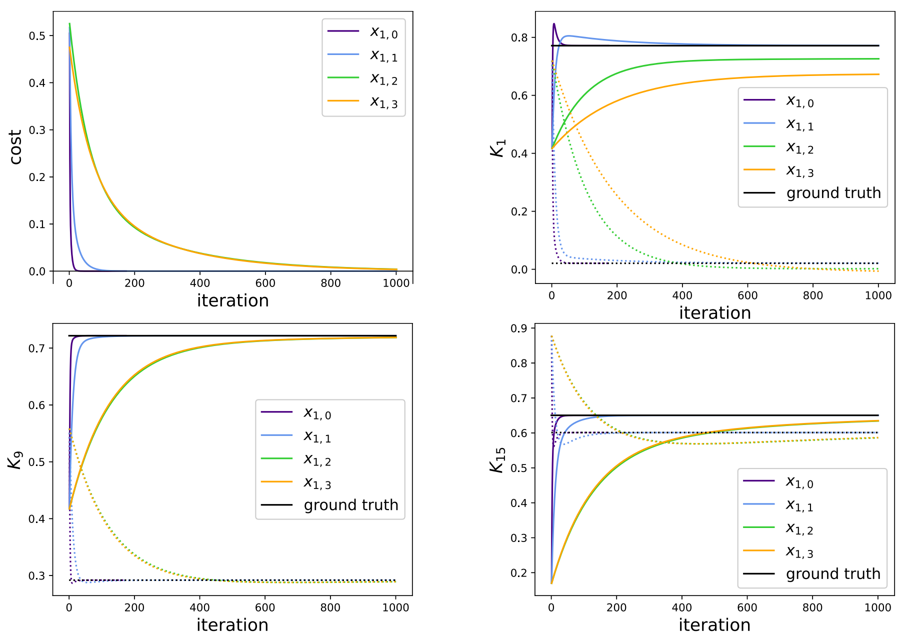

FIG. 5. *Cost function and reconstructions of $K_r(+1,-1)$ (solid lines) and $K_r(-1,+1)$ (dotted lines) for $r = 1, 9, 15$ and $R = 20$ under different measurement locations for x_1 given by $\{x_{1,0} = a_{1/2} - T\,, x_{1,1} = a_{1/2} + \frac{T}{2}\,, x_{1,2} = a_{1/2} + \frac{4}{5}T\,, x_{1,3} = a_{1/2} + T\}$ with $x_{1,3} = x_2$.*

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 run over the entire parameter domain of $[0,1]^2$ and the positive definiteness stays throughout the domain, hinting that the proposed experimental design, Design 4.1, can potentially be globally well-posed. To generate the plots, a simplified setup with $R=2$ was considered.

5.2. Ill-conditioning for close measurement locations. We now provide numerical evidence to reflect the assertion from section 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 fix the initial data as in Design 4.1, but vary the detector positions. To be specific, we assign values to x_1 using $\{x_{1,0} = a_{1/2} - T\,, x_{1,1} = a_{1/2} + \frac{T}{2}\,, x_{1,2} = a_{1/2} + \frac{4}{5}T\,, x_{1,3} = x_2 = a_{1/2} + T\}$. 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 $\eta = \frac{2\lambda_{\min}}{\lambda^2_{\max}}$ 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 panels 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 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, the second panel 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.

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 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 was presented. It is in good agreement with the theoretical predictions.

A natural extension of the results presented in the current paper is the algorithmic performance in higher space dimensions. The theoretical findings seem to apply in a straightforward manner, and we are convinced that an adaptation of Design 4.1 in analogy to Remark 4.5 and [27] could provide well-posedness, 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, 25, 44, 50]. Furthermore, we adopted a first optimize, then discretize approach in this article. Suggested in [4, 24, 37], a first discretize, then optimize framework could 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 nonavoidable. A class of criteria proposed under the Bayesian perspective sheds 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.

REFERENCES

[1] G. Albi, E. Calzola, and G. Dimarco, *A data-driven kinetic model for opinion dynamics with social network contacts*, European J. Appl. Math., (2024), pp. 1–27, https://doi.org/10.1017/S0956792524000068.

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

[3] W. Alt, *Biased random walk models for chemotaxis and related diffusion approximations*, J. Math. Biol., 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 J. Numer. Anal., 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), 123010, https://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 Probl. Imaging, 2 (2008), pp. 23–42, https://doi.org/10.3934/ipi.2008.2.23.

[7] H. Berg, *Random Walks in Biology*, Princeton University Press, Princeton, NJ, 1993, https://doi.org/10.1515/9781400820023.

[8] M. Burger and W. Mühlhuber, *Iterative regularization of parameter identification problems by sequential quadratic programming methods*, Inverse Problems, 18 (2002), pp. 943–969, https://doi.org/10.1088/0266-5611/18/4/301.

[9] J. A. Carrillo, M. Fornasier, J. Rosado, and G. Toscani, *Asymptotic flocking dynamics for the kinetic Cucker–Smale model*, SIAM J. Math. Anal., 42 (2009), pp. 218–236, https://doi.org/10.1137/090757290.

[10] C. Cercignani, *The Boltzmann Equation and Its Applications*, Appl. Math. Sci. 67, Springer New York, 2012, https://link.springer.com/book/10.1007/978-1-4612-1039-9.

[11] K. Chen, Q. Li, and J.-G. Liu, *Online learning in optical tomography: A stochastic approach*, Inverse Problems, 34 (2018), 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 modelsof opinion dynamics*, SIAM J. Appl. Math., 84 (2024), pp. 1096–1115, https://doi.org/10.1137/22M1544415.

[14] B. Davison and J. Sykes, *Neutron Transport Theory*, Internat. Ser. Monogr. Phys., Oxford University Press, Oxford, UK, 1958.

[15] H. Egger, J.-F. Pietschmann, and M. Schlottbom, *Identification of chemotaxis models with volume-filling*, SIAM J. Appl. Math., 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*, Comput. Optim. Appl., 62 (2013), pp. 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 Comput. Biol., 12 (2016), pp. 1–22, https://doi.org/10.1371/journal.pcbi.1004843.

[18] R. Erban and H. G. Othmer, *From individual to collective behavior in bacterial chemotaxis*, SIAM J. Appl. Math., 65 (2004), pp. 361–391, https://doi.org/10.1137/S0036139903433232.

[19] R. Ferrentino and C. Boniello, *On the well-posedness for optimization problems: A theoretical investigation*, Appl. Math., 10 (2019), pp. 19–38, https://doi.org/10.4236/am.2019.101003.

[20] F. Filbet and C. Yang, *Numerical simulations of kinetic models for chemotaxis*, SIAM J. Sci. Comput., 36 (2014), pp. B348–B366, https://doi.org/10.1137/130910208.

[21] K. Fister and M. Mccarthy, *Identification of a chemotactic sensitivity in a coupled system*, Math. Med. Biol., 25 (2008), pp. 215–232, https://doi.org/10.1093/imammb/dqn015.

[22] R. M. Ford and D. A. Lauffenburger, *Measurement of bacterial random motility and chemotaxis coefficients:* II. *Application of single-cell-based mathematical model*, Biotechnol. Bioeng., 37 (1991), pp. 661–672, https://doi.org/10.1002/bit.260370708.

[23] A. Giometto, F. Altermatt, A. Maritan, R. Stocker, and A. Rinaldo, *Generalized receptor law governs phototaxis in the phytoplankton* Euglena gracilis, Proc. Natl. Acad. Sci. USA, 112 (2015), pp. 7045–7050, https://doi.org/10.1073/pnas.1422922112.

[24] M. D. Gunzburger, *Perspectives in Flow Control and Optimization*, SIAM, Philadelphia, 2002, https://doi.org/10.1137/1.9780898718720.

[25] E. Haber, U. M. Ascher, and D. Oldenburg, *On optimization techniques for solving nonlinear inverse problems*, Inverse Problems, 16 (2000), pp. 1263–1280, https://doi.org/10.1088/0266-5611/16/5/309.

[26] K. Hellmuth, C. Klingenberg, Q. Li, and M. Tang, *Multiscale convergence of the inverse problem for chemotaxis in the Bayesian setting*, Computation, 9 (2021), 119, https://doi.org/10.3390/computation9110119.

[27] K. Hellmuth, C. Klingenberg, Q. Li, and M. Tang, *Kinetic chemotaxis tumbling kernel determined from macroscopic quantities*, SIAM J. Math. Anal., 56 (2024), pp. 568–587, https://doi.org/10.1137/22M1499911.

[28] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, Math. Model. 23, Springer, Berlin, 2008, https://doi.org/10.1007/978-1-4020-8839-1.

[29] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 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*, Proc. Natl. Acad. Sci. USA, 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 J. Numer. Anal., 59 (2021), pp. 119–142, https://doi.org/10.1137/20M134383X.

[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), 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 J. Appl. Math., 79 (2019), pp. 2340–2358, https://doi.org/10.1137/18M1207582.

[34] D. Le, *Dynamics of a bio-reactor model with chemotaxis*, J. Math. Anal. Appl., 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*, Proc. Natl. Acad. Sci. USA, 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), 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*, J. Comput. Appl. Math., 362 (2019), pp. 596–613, https://doi.org/10.1016/j.cam.2018.07.028.

[38] S. Motsch and E. Tadmor, *Heterophilious dynamics enhances consensus*, SIAM Rev., 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*, J. Math. Biol., 26 (1988), pp. 263–298, https://doi.org/10.1007/BF00277392.

[40] H. G. Othmer and T. Hillen, *The diffusion limit of transport equations* II*: Chemotaxis equations*, SIAM J. Appl. Math., 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 Comput. Biol., 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/10.1016/j.ifacol.2017.08.1513.

[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), 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*, Commun. Comput. Phys., 8 (2010), pp. 1–50, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7d86877cb31043aad5585b62dfd9e083ac607a83.

[45] G. B. Rybicki and A. P. Lightman, *Radiative Processes in Astrophysics*, Wiley-VCH, Weinheim, 1986, https://doi.org/10.1002/9783527618170.

[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*, Nat. Commun., 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 Comput. Biol., 6 (2010), pp. 1–12, 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*, Proc. Natl. Acad. Sci. USA, 108 (2011), pp. 16235–16240, https://doi.org/10.1073/pnas.1101996108.

[49] M. Seyrich, Z. Alirezaeizanjani, C. Beta, and H. Stark, *Statistical parameter inference of bacterial swimming strategies*, New J. Phys., 20 (2018), 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 Process. Lett., 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 J. Math. Anal., 54 (2022), pp. 2753–2767, https://doi.org/10.1137/21M1417387.

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

[53] A. Tikhonov, *On the stability of the functional optimization problem*, USSR Comput. Math. Math. Phys., 6 (1966), pp. 28–33, 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, Cambridge, UK, 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*, Proc. Natl. Acad. Sci. USA, 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), 014612, https://doi.org/10.1103/PhysRevE.109.014612.