HAL
open science

Single-Trajectory Characterization of Active Swimmers in a Flow

Gaspard Junot, Eric Clément, Harold Auradou, Reinaldo García-García

To cite this version:

Gaspard Junot, Eric Clément, Harold Auradou, Reinaldo García-García. Single-Trajectory Characterization of Active Swimmers in a Flow. Physical Review E , 2021, 103 (3), pp.032608. 10.1103/PhysRevE.103.032608 . hal-03267100

HAL Id: hal-03267100

https://hal.sorbonne-universite.fr/hal-03267100v1

Submitted on 22 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Single-Trajectory Characterization of Active Swimmers in a Flow.

Gaspard Junot,[1] Eric Clément,[1,2] Harold Auradou,[3] and Reinaldo García-García[1,*]

[1] *Laboratoire PMMH-ESPCI Paris, PSL Research University, Sorbonne Université and Denis Diderot, 7, quai Saint-Bernard, Paris, France*

[2] *Institut Universitaire de France (IUF)*

[3] *Université Paris-Saclay, CNRS, FAST, 91405, Orsay, France*

We develop a maximum likelihood method to infer relevant physical properties of elongated active particles. Using individual trajectories of advected swimmers as input, we are able to accurately determine their rotational diffusion coefficients and an effective measure of their aspect ratio, also providing reliable estimators for the uncertainties of such quantities. We validate our theoretical construction using numerically generated active trajectories upon no-flow, simple shear, and Poiseuille flow, with excellent results. Being designed to rely on single-particle data, our method eases applications in experimental conditions where swimmers exhibit a strong morphological diversity. We briefly discuss some of such ongoing experimental applications, specifically, in the characterization of swimming *E. coli* in a flow.

I. INTRODUCTION

To monitor passive probes, living cells and microorganisms moving in their environment, modern instrumentation is now offering possibilities for long-time tracking and trajectory reconstruction, with excellent spatial and temporal resolution [1]. The collected data sets contain in principle relevant informations on the internal and external processes, of either mechanical or biological origin. However, inferring them is still a challenging task which has motivated the recent development of adapted analytical tools [2–7].

Tracking colloids close to thermal equilibrium in various media was proven quite fruitful as it allowed the development of a quantitative micro-rheology instrumentation [8]. This technique was further extended to investigate complex fluids [9, 10], and dynamical processes inside living cells [1, 11, 12]. In recent years, much effort was dedicated to the study of passive colloidal particles as they diffuse in active suspensions composed of self-propelling entities [13–15]. Several studies focus on the contribution of the active environment to the diffusion of tracers [16–18], while others aim at generalizing the notion of thermal bath and to characterize the thermodynamic fluctuations associated with the passive particle in complex environments [19–22]. However, there has been much less focus on understanding how an active particle gets itself affected by the properties of its the environment.

In this work, we develop a maximum likelihood (ML) approach to study elongated active swimmers using raw data of swimming tracks. The general aim is to extract from a stochastic data set, the physical parameters that characterize the particle dynamics as well as the source of rotational noise. In general, such data encompass the active transport contribution resulting from the interaction between the flow and the swimmer, and the stochastic component associated with rotational noise. Although our framework is general, we focus here on three particular situations, (*i*) a free swimmer, (*ii*) a swimmer in a shear flow and, finally, (*iii*) a swimmer in a Poiseuille flow. We take into account inherent limitations encountered in experiments, such as the finite duration of the tracks and their discrete sampling. The method is first tested against numerical simulations, hence showing an excellent agreement between prescribed parameter values and our estimators.

Our tool has been conceived to study *E. coli* mutants (smooth swimmers) in a parabolic flow, however, it can be generalized to different experimental situations by adapting the underlying stochastic process to the system at hand. For instance, many studies focused on the behavior of synthetic microswimmers in dynamic environments [23], such as externally imposed chemical gradients [24], or flows [25]. Our method can also be used to extract key parameters from trajectories of synthetic microswimmers in such situations.

Smooth swimmers were recently considered in Ref. [26]. It was shown that a deterministic model, consisting of an advected swimming ellipsoid, fairly reproduces bacterial trajectories at short times. However, at longer times, a stochastic component comes into play as a multiplicative noise leading to the stochastic exploration of the phase space. Using the ML method, we are able to extract the effective rotational diffusivity from the swimming trajectories, which can be compared quantitatively with the Brownian diffusion of an ellipsoid in a viscous fluid.

ML approaches have been applied earlier to the characterization of either passive or active tracers. For instance, in Ref. [27] a combination of Bayesian and ML analysis was used to infer the stochastic model that fits best a given single-passive-particle track. In Ref. [28] similar methods led to a noninvasive protocol to infer molecular chemotactic responses from bacterial trajectories. Regarding those studies, the main differences with our method are: (*i*) it disentangles the deterministic (advective) and random (diffusive) components of single bacterial trajectories to better understand how active swimmers interact with different flow profiles, (*ii*) the input of our ML procedure are raw dynamical trajectories (i.e.,

the set of positions and orientation vectors of the particle as function of time).

The paper is organized as follows. In Sec. II we introduce the theoretical model and briefly discuss the general philosophy of our ML method. In Sec. III we sketch the steps to build the log-likelihood and derive our ML estimators as well as their uncertainties to leading order in the number of sampling points. In Sec. IV we use numerical simulations to validate our method in three configurations, no-flow, simple shear flow, and Poiseuille flow. A preliminary application of our method in the analysis of experimental tracking data is discussed in Sec. V, and general conclusions are given in Sec. VI. All technical details regarding our calculations and experiments are left to the Appendix.

II. PRELIMINARIES

A. Stochastic dynamics of the active Betherton-Jeffery model

In Ref. [26], it was shown that rotational diffusion considerably affects the trajectories of mutant *E. coli* in a flow. This fact represents a strong motivation to consider the effects of rotational noise in the dynamics of the orientation vector of smooth swimmers. We focus on the model studied in Ref. [26] in presence of noise, which we refer to as the stochastic active Betherton-Jefferey (SABJ) model from now on. The model describes the behavior of an ellipsoid swimming at a constant speed in a flow. Its dynamics read:

$$\dot{\mathbf{r}} = v_0\mathbf{p} + \mathbf{v}^F(\mathbf{r}), \tag{1}$$

$$\dot{\mathbf{p}} = \left(\mathbb{1} - \mathbf{p}\otimes\mathbf{p}\right)\left[\beta\mathbb{E}(\mathbf{r}) + \mathbb{\Omega}(\mathbf{r})\right]\mathbf{p} - 2D_R\mathbf{p} + \sqrt{2D_R}\mathbf{p}\wedge\boldsymbol{\xi}_R. \tag{2}$$

Above, v_0 is the self-propulsion velocity of the particle, $\mathbf{v}^F$ is the local flow velocity, the symbol $\mathbb{1}$ denotes the identity matrix, and $\otimes$ is used to denote tensor products. The number $\beta = (r^2-1)/(r^2+1)$ is the Betherton parameter, which represents a measure of the geometrical asymmetry of the swimmer (r is the aspect-ratio of the ellipsoidal particle). The components of the tensors $\mathbb{E}$ and $\mathbb{\Omega}$ are given as follows, $E_{ij}(\mathbf{r}) = (\partial_{x_i}v_j^F + \partial_{x_j}v_i^F)/2$ and $\Omega_{ij}(\mathbf{r}) = (\partial_{x_i}v_j^F - \partial_{x_j}v_i^F)/2$.

In Eq. (2), rotational diffusion is encoded in the diffusion coefficient D_R and the Gaussian white noise $\boldsymbol{\xi}_R$ which has zero mean and variance $\langle\boldsymbol{\xi}_R(t)\otimes\boldsymbol{\xi}_R(t')\rangle = \mathbb{1}\delta(t-t')$. In this work, Eq. (2) is interpreted in the Ito sense. The Ito term $-2D_R\mathbf{p}$ is thus needed to guarantee the conservation of the norm of $\mathbf{p}$. Note that, although one could also introduce a translational diffusion term in Eq. (1), such contribution can be assumed negligible in typical experimental situations. Consider, for instance, an *E. coli* bacterium which is few microns length, swimming at a typical speed of $25\,\mu\mathrm{m\,s^{-1}}$. The distance over which it diffuses in 1 s is $\approx 0.1\,\mu$m to be compared to the $25\,\mu$m travelled due to its activity.

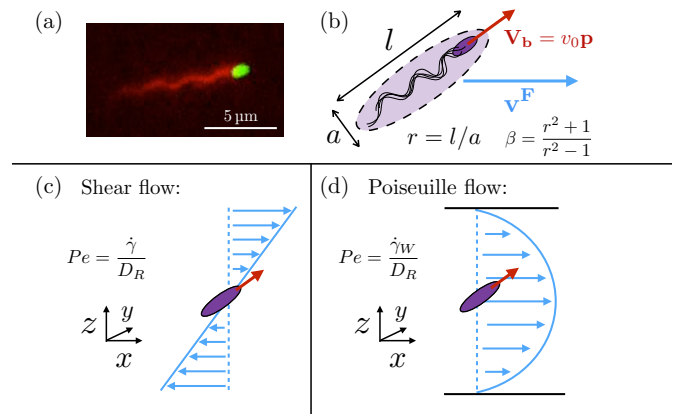

FIG. 1. (a) Picture of an *E. coli* bacterium with the head GFP labelled (green color) and the flagella bundle tagged in red with a specific alexa fluor marker. (b) Sketch of the effective ellipsoid of aspect ratio r modeling the bacterium. (c) Simple shear flow geometry characterized by a constant shear rate $\dot{\gamma}$ and, (d) Poiseuille flow geometry; $\dot{\gamma}_W$ is the maximal shear rate. We also introduce the Péclet number, which is the natural dimensionless form of the inverse of the rotational diffusion coefficient.

In Fig. 1 we illustrate our model system, making explicit the way in which geometrical parameters are defined. It is important to remark that the bacterium is a complex object and that the optical determination of β is non-trivial. In other words, β is an effective parameter and its determination based on imagery alone may not be reliable. However, for an elongated bacterium such as *E. coli* its expected value should be close to one (see the sketch in Fig. 1(b)). In the same figure, we also show how the Péclet number is defined in presence of shear and Poiseuille flows. The Péclet number plays an important role in numerics because it is the natural dimensionless form of the inverse of the rotational diffusion coefficient, see Sec. IV for more details.

An important role is played by the Markov kernel of the process (1)-(2), which we denote by $K(\mathbf{r},\mathbf{p},t|\mathbf{r}',\mathbf{p}',t')$. We focus here on stationary flow profiles in which case the propagtor only depends on $\Delta t = t - t'$ due to time-translation invariance. We thus simplify the notations by writing $K(\mathbf{r},\mathbf{p},t|\mathbf{r}',\mathbf{p}',t') = K(\mathbf{r},\mathbf{p},\Delta t|\mathbf{r}',\mathbf{p}',0) \equiv K(\mathbf{r},\mathbf{p},\Delta t|\mathbf{r}',\mathbf{p}')$. The propagator satisfies the Fokker-Planck equation $\partial_t K = \hat{L}_{FP}K$ with initial condition $K(\mathbf{r},\mathbf{p},0|\mathbf{r}',\mathbf{p}') = \delta(\mathbf{r}-\mathbf{r}')\delta(\mathbf{p}-\mathbf{p}')$. The Fokker-Planck operator associated to these dynamics reads

$$\hat{L}_{FP}(\mathbf{r},\mathbf{p}) = D_R\sum_{i,j}\partial^2_{p_i,p_j}\cdot\tilde{M}_{ij}(\mathbf{p}) - \sum_i\partial_{p_i}\cdot h_i(\mathbf{r},\mathbf{p}) - \sum_i\partial_{x_i}\cdot v_i(\mathbf{r},\mathbf{p}). \tag{3}$$

To write (3), we condensed the translational part of the motion into $\mathbf{v}(\mathbf{r},\mathbf{p}) = v_0\mathbf{p}+\mathbf{v}^F(\mathbf{r})$, and the drift term in Eq. (2) into $\mathbf{h}(\mathbf{r},\mathbf{p}) = (\mathbb{1}-\mathbf{p}\otimes\mathbf{p})(\beta\mathbb{E}(\mathbf{r})+\Omega(\mathbf{r}))\mathbf{p} - 2D_R\mathbf{p}$. Additionally, each element of the matrix $\tilde{\mathbb{M}}(\mathbf{p})$ is a quadratic form in $\mathbf{p}$, $\tilde{M}_{ij}(\mathbf{p}) = \sum_{l,q} T_{ijlq}p_l p_q$, where

$$T_{ijlq} = \sum_k \epsilon_{ilk}\epsilon_{jqk} = \delta_{ij}\delta_{lq} - \delta_{iq}\delta_{jl}, \tag{4}$$

and ϵ_{ijk} denotes the Levi-Civita symbol.

B. Estimating parameters from single trajectories

By studying individual trajectories of bacteria, one could access the statistics of D_R and β over a population. To achieve an accurate measurement of those parameters, we develop an ML method. We start by constructing the log-likelihood characterizing a single-bacterium trajectory; then the relevant physical parameters are determined by maximizing it.

In this paragraph we sketch out the general philosophy of our procedure. In experiments, the trajectory of a bacterium is recorded at regularly spaced time instants $\{t_\alpha\}$, such that $\Delta t_\alpha = t_{\alpha+1} - t_\alpha \equiv f^{-1}$ for each $\alpha = 0, 1, \ldots, N-1$, where f is the sampling frequency. The data has the form of a discrete set of N position and orientation values, $\{\Gamma\}$, where $\Gamma = (\mathbf{r},\mathbf{p})$. Assuming that the dynamics of the swimmer are compatible with Eqs. (1)-(2), the probability of measuring that particular track can be expressed as follows:

$$\mathcal{P}\big[\{\Gamma\}\big] = \prod_{\alpha=0}^{N-1} K(\Gamma(t_{\alpha+1}), f^{-1}|\Gamma(t_\alpha)). \tag{5}$$

Although not made explicit, the probability of the track is parametrized by the relevant physical quantities of the model, i.e., v_0, D_R, and β. One can introduce the log-likelihood as $\mathcal{S} = \ln\mathcal{P}$, or more explicitly

$$\mathcal{S} = \sum_{\alpha=0}^{N-1} \ln K(\Gamma(t_{\alpha+1}), f^{-1}|\Gamma(t_\alpha)), \tag{6}$$

anticipating the importance of the Markov kernel of the SABJ model in this work. By maximizing the log-likelihood with respect to the parameters, we obtain the best estimates compatible with the hypothesis that the data are generated by Eqs. (1)-(2).

The above procedure can be formally justified by the fact that the SABJ model is Markovian, which means that the propagator encodes all the relevant physical information about the process. Markovianity, in conjunction with time-translational invariance, allows us then to interpret one trajectory as an esemble of independent 'events', each consisting on the generation of new values for the increments of the process, with each outcome being fully characterized by the same probabilistic model K.

It is important to emphasize that our method can be used to infer parameters from tracking data using generic models and not only the one we consider here. In particular, our method does not depend on the precise shape of the flow profile as far as it is stationary. It all comes down to computing the corresponding propagator and following the general steps sketched above. In the next section we discuss the case of the SABJ model and derive the ML estimators as well as their corresponding uncertainties.

III. MAXIMUM LIKELIHOOD METHOD

A. Approximate log-likelihood and maximum likelihood estimators

As discussed above, we need access to the propagator K in order to build the log-likelihood. However, solving the Fokker-Planck equation in presence of a generic flow seems hopeless. This implies that, in practice, the likelihood cannot be exactly constructed. Propitiously, a fairly good approximant can be derived by relying on the relatively large value of experimentally accessible sampling frequencies. The scheme developed in this section does not rely upon a direct approximate solution of the Fokker-Planck equation for particular flow profiles, since such approach would hinder the generality of our method. As we discuss below, we start from the expression of the continuous-time path probability representation of the SABJ model in presence of a generic flow profile and then introduce time-discretization in the simplest and more natural way possible.

To begin with, let us first consider the (ideal) limit at which a swimmer trajectory is sampled continuously, which corresponds to $f \to \infty$. In this case the track probability (5) becomes the continuous path probability associated to the process (1)-(2), which can be written in the Martin-Siggia-Rose-de Dominics-Janssen (MSRDJ) representation (see Refs. [29–31] for details) as follows:

$$\mathcal{P}\big[\Gamma(\bullet)\big] = \mathcal{J}\big[\Gamma(\bullet)\big]\int \mathcal{D}\big[\hat{\Gamma}(\bullet)\big]e^{-\hat{\mathcal{A}}[\Gamma(\bullet),\hat{\Gamma}(\bullet)]}, \tag{7}$$

where $\hat{\Gamma} = (\hat{\mathbf{r}},\hat{\mathbf{p}})$ denote the response fields of the MSRDJ formalism, while the action has the form

$$\hat{\mathcal{A}} = \int_0^t \Big\{ \mathrm{i}\big[\dot{\mathbf{r}} - \mathbf{v}\big]^{\mathrm{T}}\hat{\mathbf{r}} + \mathrm{i}\big[\dot{\mathbf{p}} - \mathbf{h}\big]^{\mathrm{T}}\hat{\mathbf{p}} + D_R\hat{\mathbf{p}}^{\mathrm{T}}\tilde{\mathbb{M}}\hat{\mathbf{p}}\Big\}dt'. \tag{8}$$

The vector fields $\mathbf{v}$ and $\mathbf{h}$, as well as the matrix $\tilde{\mathbb{M}}(\mathbf{p})$, were introduced in Sec. II and we have omitted their explicit dependence on $\mathbf{r}$ and $\mathbf{p}$ for compactness. Additionally, the prefactor $\mathcal{J}$ in Eq. (7) is a Jacobian whose precise form depends on the underlying discretization scheme of both, the stochastic dynamics (1)-(2) and the path probability. In the Ito convention (which we use here), one has $\mathcal{J} = 1$ and this factor can be omitted.

Imagine now that the sampling frequency is large enough but finite. Then, to first order in f^{-1}, one can

approximate the time integral involved in the expression of the continuous-time dynamical action (8) by a discrete sum (with a prescribed discretization scheme, i.e., Ito), which is no more than approximating a Riemann integral by one of its associated Darboux sums when the time step is small enough. After integrating over the response variables in (7), one gets for the (now discrete) path probability

$$\mathcal{P}\big[\{\Gamma\}\big] = \prod_{\alpha=0}^{N-1} \exp\bigg(-\frac{1}{f}\mathcal{L}(\Gamma(t_{\alpha+1}),\Gamma(t_\alpha))\bigg). \tag{9}$$

The precise form of the function $\mathcal{L}$ is rather involved and we prefer not to discuss it here. In Appendix A we develop the full computation. The point we do wish to highlight is that a direct comparison between Eqs. (5) and (9) explicitly illustrates the nature of our approximation for the propagator K between two events sampled within a very small time interval. In the same vein, taking the logarithm in (9) and comparing the result with Eq. (6) leads to our approximation for the log-likelihood:

$$\mathcal{S} = -\frac{1}{f}\sum_{\alpha=0}^{N-1}\mathcal{L}(\Gamma(t_{\alpha+1}),\Gamma(t_\alpha)). \tag{10}$$

Following this program, we derive an explicit expression for the log-likelihood of the SABJ model in Appendix A. The result reads:

$$\mathcal{S}(\tilde{\mathcal{N}}_R,\beta) = N\ln\tilde{\mathcal{N}}_R - \frac{\tilde{\mathcal{N}}_R}{4}\big(A_R - 2B_R\beta + C_R\beta^2\big), \tag{11}$$

where $\tilde{\mathcal{N}}_R = D_R^{-1}f$ is the inverse of the rotational diffusion coefficient adimensionalized by the sampling frequency. The dimensionless constants A_R, B_R, and C_R are extensive in the number of sample points N of the trajectory for $N \gg 1$ (see Eq. (A13)), and depend on the registered position and orientation vectors along the track.

Although the computation leading to (11) is tedious, the final result exhibits a very simple dependence on both, D_R and β. Maximizing $\mathcal{S}$ is now trivial. Provided that $\tilde{\mathcal{N}}_R \neq 0$, we obtain the following expressions

$$\beta = \frac{B_R}{C_R}, \tag{12}$$

$$\tilde{\mathcal{N}}_R = \frac{4NC_R}{A_RC_R - B_R^2}. \tag{13}$$

For consistency, these results must be positive numbers. We discuss below on the accuracy of such estimates.

B. Note on convergence and uncertainties

From a methodological point of view, it is crucial to be able to quantify the quality of the estimates (12) and (13). A first important question is whether our approximation for the log-likelihood is compatible with experimental data or not. In other words, we need to provide a quantitative meaning to the statement "f is sufficiently large". A good way to answer this question is to have a closer look to a consistency condition derived in Appendix A, namely (see Eq. (A11) and the discussion following it)

$$\mathbf{p}(t_\alpha)\cdot\Delta\mathbf{p}(t_\alpha) + 2D_Rf^{-1}|\mathbf{p}(t_\alpha)|^2 = 0, \tag{14}$$

where $\Delta\mathbf{p}(t_\alpha)$ is the increment of the orientation vector in a time step, $\Delta\mathbf{p}(t_\alpha) = \mathbf{p}(t_{\alpha+1}) - \mathbf{p}(t_\alpha)$. This condition is linked to the conservation of the norm of $\mathbf{p}$. Indeed, when $f\to\infty$, Eq. (14) becomes $\mathbf{p}(t)\cdot d\mathbf{p}(t) = 0$. However, an inconsistency may appear if f is not large enough. To understand this, let us consider the scalar product $\mathbf{p}(t_\alpha)\cdot\mathbf{p}(t_{\alpha+1})$ for vectors with constant norm, $|\mathbf{p}(t_\alpha)| = |\mathbf{p}(t_{\alpha+1})| = 1$ for all α, as is the case for the orientation vector. From one side, we have $\mathbf{p}(t_\alpha)\cdot\mathbf{p}(t_{\alpha+1}) = \cos(\phi_\alpha)$, where ϕ_α is the angle between $\mathbf{p}(t_\alpha)$ and $\mathbf{p}(t_{\alpha+1})$. On the other hand, we can write $\mathbf{p}(t_\alpha)\cdot\mathbf{p}(t_{\alpha+1}) = \mathbf{p}(t_\alpha)\cdot(\mathbf{p}(t_\alpha) + \Delta\mathbf{p}(t_\alpha)) \equiv 1 + \mathbf{p}(t_\alpha)\cdot\Delta\mathbf{p}(t_\alpha)$. Using then the condition (14), we finally get $\mathbf{p}(t_\alpha)\cdot\mathbf{p}(t_{\alpha+1}) = 1 - 2D_Rf^{-1}$, which implies $\cos(\phi_\alpha) = 1 - 2D_Rf^{-1}$. It is now clear that our approximation breaks down when $f < D_R$, since in that case, one would have $|\cos(\phi_\alpha)| > 1$. In other words, our approximation is not reliable when the rotational diffusion coefficient to be estimated is larger than the sampling frequency.

To assess the efficiency of the method, we estimate the speed at which the computed parameters converge toward their exact values as a function of the length of the sample, N. Let us denote by $\boldsymbol{\theta} = (\beta,\tilde{\mathcal{N}}_R)$ the estimated two-dimensional parameter vector, and by $\Delta\boldsymbol{\theta}$ the estimation error. We also write $\boldsymbol{\theta}^\star$ to denote the vector of the true values of the parameters. If the assumption that the data is generated by the model (1)-(2) is statistically correct, then one can assert that, up to order $N^{-1/2}$, the estimated parameters are unbiased. More precisely, the error vector $\sqrt{N}\Delta\boldsymbol{\theta}$ converges in distribution to a normal vector of zero mean and variance matrix $\mathbb{F}^{-1}(\boldsymbol{\theta}^\star)$, where $\mathbb{F}(\boldsymbol{\theta}^\star)$ is the Fisher information matrix [32], which can be approximately determined from the log-likelihood (for large N) as $F_{ij}(\theta^\star) = -N^{-1}\partial^2_{\theta_i,\theta_j}S\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}$. Beyond order $N^{-1/2}$, the estimates (12) and (13) are known to exhibit a bias of order $O(1/N)$, but we can neglect such contributions if N is large enough. From this calculation, we estimate the following error bars (we use the notation $e(\theta_i) = \sqrt{\langle\theta_i^2\rangle}$):

$$e(\beta) \sim \frac{2\sqrt{N}}{\sqrt{\tilde{\mathcal{N}}_R^\star[2NC_R - \tilde{\mathcal{N}}_R^\star(B_R - \beta^\star C_R)^2]}}, \tag{15}$$

$$e(\tilde{\mathcal{N}}_R) \sim \frac{\sqrt{2C_R}\,\tilde{\mathcal{N}}_R^\star}{\sqrt{2NC_R - \tilde{\mathcal{N}}_R^\star(B_R - \beta^\star C_R)^2}}. \tag{16}$$

Interestingly, there are also cross-correlations between the errors, meaning that β and $\tilde{\mathcal{N}}_R$ cannot be determined

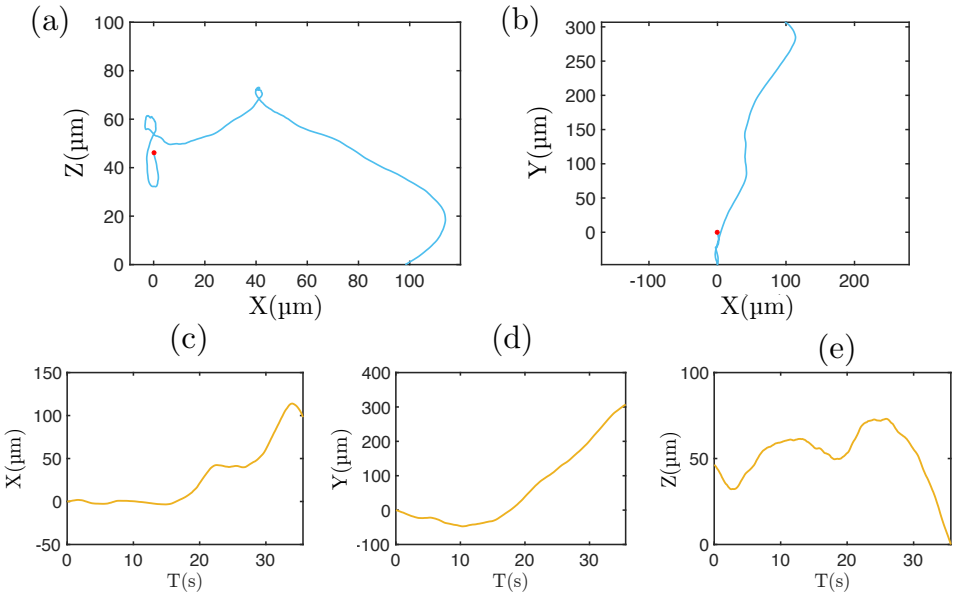

FIG. 2. Simulated track of a bacterium swimming at constant speed $v_0 = 25\,\mu\mathrm{m\,s}^{-1}$ in a Poiseuille flow. The height of the channel is $h = 100\,\mu\mathrm{m}$. The maximal flow velocity is $U_m = 25\,\mu\mathrm{m\,s}^{-1}$ along the X direction, which sets $Pe = 100$. The geometric parameter $\beta = 0.9$. (a) Projection of the trajectory in the shear plane z-x. (b) Projection of the trajectory in the bottom of the channel i.e., the plane y-x. The red dot indicates the starting point of the trajectory. (c-e) x,y,z coordinates as function of time.

independently with arbitrarily high accuracy. Explicitly, we have

$$\langle \Delta\beta\Delta\tilde{\mathcal{N}}_R\rangle \sim \frac{2\tilde{\mathcal{N}}_R^\star(B_R-\beta^\star C_R)}{2NC_R-\tilde{\mathcal{N}}_R^\star(B_R-\beta^\star C_R)^2}. \tag{17}$$

As a consistency check, note that as A_R, B_R and C_R are extensive for $N \gg 1$, one has $e(\boldsymbol{\theta}) \sim N^{-1/2}$, as it should be.

IV. NUMERICAL VALIDATION OF THE METHOD

In the previous section, we derived ML estimators for the parameters of the SABJ model. To test theses expressions and their accuracy, we apply the ML method on simulated bacterial tracks generated using Eqs. (1)-(2), with an Euler-Maruyama scheme and with a given set of parameters β^* and D_R^*. Then, we apply the ML method (Eq. (12) and Eq. (13)) on these trajectories to find the estimates β and D_R. Finally, we compare our estimators with the input parameters of the simulations. We study three different cases, namely: (i) a free swimmer (in which case the parameter β plays no role), (ii) a simple shear flow (Fig.1(c)) and (iii) a Poiseuille flow (Fig.1(d)). In Fig. 2 we illustrate a simulated trajectory of a swimmer in a Poiseulle flow.

A. Free swimmer

In this case, the orientation of the bacteria follows a diffusion process in the 2-dimensional unit sphere:

$$\dot{\mathbf{p}} = -2\mathbf{p} + \sqrt{2}\mathbf{p}\wedge\boldsymbol{\xi}_R. \tag{18}$$

We adimensionalized time using the rotational diffusion coefficient D_R, and simulated trajectories with different duration T. The (dimensionless) time interval between two sampling events is fixed as $\Delta t = 10^{-6}$. To test the convergence of the method, we compute the diffusion coefficient D_R^{ML}, using Eq (13), for trajectories of different duration. In Fig. 3, we illustrate the convergence of the estimated value D_R^{ML} towards the prescribed value $D_R^* = 1$ as the length of the trajectory increases. For instance, we find a value of D_R^{ML} very close to 1, i.e $D_R^{ML} = 0.9994$, at $T = 1$.

B. Shear flow

In this case we simulate Eqs. (1) and (2) with a flow profile of the form $\mathbf{v}^F(\mathbf{r}) = \dot{\gamma} z\mathbf{e_x}$, where $\dot{\gamma}$ is the shear rate. As before, simulations are performed using dimensionless variables. Time is adimensionalized by the inverse of the shear rate, so the relevant dimensionless, diffusion-related quantity to be estimated in this case is is the Péclet number, $Pe = \dot{\gamma}/D_R$. Note that the Péclet number can be directly estimated using Eq. (13), since it

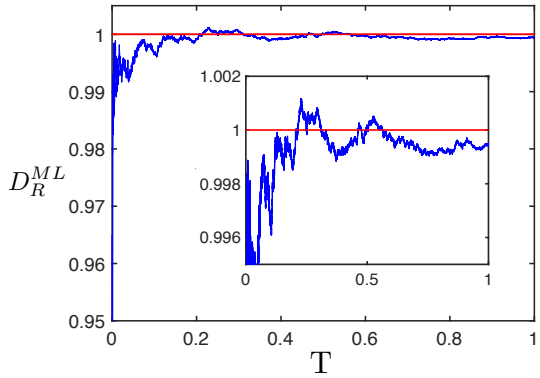

FIG. 3. Convergence of the estimated rotational diffusion coefficient D_R^{ML} in the free swimmer case. Inset: zoom close to the region $D_R^{ML} = 1$.

is just a rescaling of $\tilde{\mathcal{N}}_R$ by the known quantities $\dot{\gamma}$ and f, $Pe = (\dot{\gamma}/f)\tilde{\mathcal{N}}_R$.

We simulate trajectories for a vector $(\beta^*, Pe^*) = (0.9, 100)$. To test the convergence of the ML method, we generate a set of trajectories of different duration T with randomly chosen initial positions and orientations. In Fig. 4, we present the results of shear flow simulations. For a given trajectory length T, the estimated value of the parameters (β^{ML} in Fig. 4(a) and Pe^{ML} in Fig. 4(b)) are scattered around Pe^* and β^*. This scattering is linked to the randomness encoded in the noise and the initial orientations. The average value of the estimations over all trajectories of the same length (dark blue points) is very close to the input value of the simulations. On average, even for short trajectories, the ML method gives a good estimation of the parameters. For instance, for a trajectory of duration $T = 10$, the averaged estimated parameter differ only by 1% from the input values. As T increases, the scattering of the estimators decreases and our estimators become accurate for individual trajectories. Moreover, the standard deviation of the estimated parameters is bounded by the uncertainties computed using the expressions Eq. (15) and Eq. (16) and decays as $N^{-1/2}$.

C. Poiseuille flow

We now test the method on a Poiseuille flow profile. Here we consider a flow confined between two infinite plates separated by a vertical distance h, so that the velocity profile is given by $\mathbf{v}^F(\mathbf{r}) = \frac{4U_m}{h} z\left(1 - \frac{z}{h}\right)\mathbf{e_x}$. U_m denotes the maximal flow velocity at the middle of the Poiseille flow ($z = h/2$). We also work with dimensionless variables in this case, so that lengths and time are adimensionlalized using h and the maximum shear rate $\dot{\gamma}_W = \frac{4U_m}{h}$, respectively. In addition to β, the relevant dimensionless parameter is again the Péclet number, $Pe = \frac{\dot{\gamma}_W}{D_R}$. We simulate trajectories for a vector: $(\beta^*, Pe^*) = (0.9, 100)$ and different durations T with the additional constraint that the simulation ends when the particle hits the surfaces (located in z=0 and z=1). For instance, for the parameters used in the Fig.5, it is difficult to have trajectories longer than T=50.

For a given T, the estimators are scattered around the input values. As T increases the scattering decreases. The estimator Pe^{ML} remains very good, both on average (less than 2% of error) and for individual trajectories (less than 10% of error), even for trajectories with durations as short as $T = 1$. The estimator β^{ML} is not as good as in the shear flow, but we still have an average estimated value within a 10% of the input one. The standard deviation of the estimated values are still bounded by the expressions Eq. (15) and Eq. (16) and still decrease as $N^{-1/2}$. Our estimators can be further improved by computing errors up to $O(N^{-1})$ to compensate for the shorter tracks. This will be discussed in more detail elsewhere.

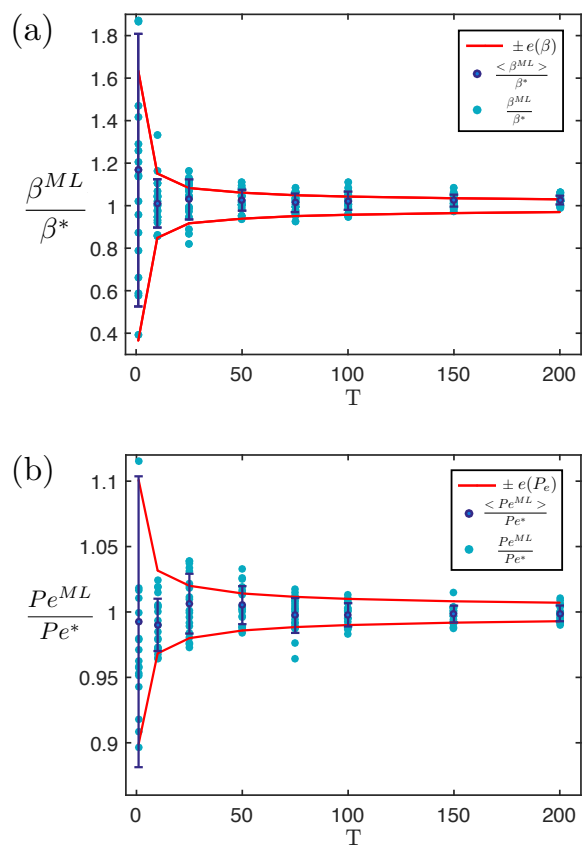

FIG. 4. Convergence of the estimated (a) β^{ML} and (b) Péclet Pe^{ML} for simulated trajectories in shear flow. For each duration T, 20 trajectories were simulated. Each cyan point corresponds to the estimation of one trajectory, the dark blue points are the average of these estimations over the 20 trajectories. The red curves represent the uncertainties computed using the expressions Eq. (15) (resp. Eq. (16)).

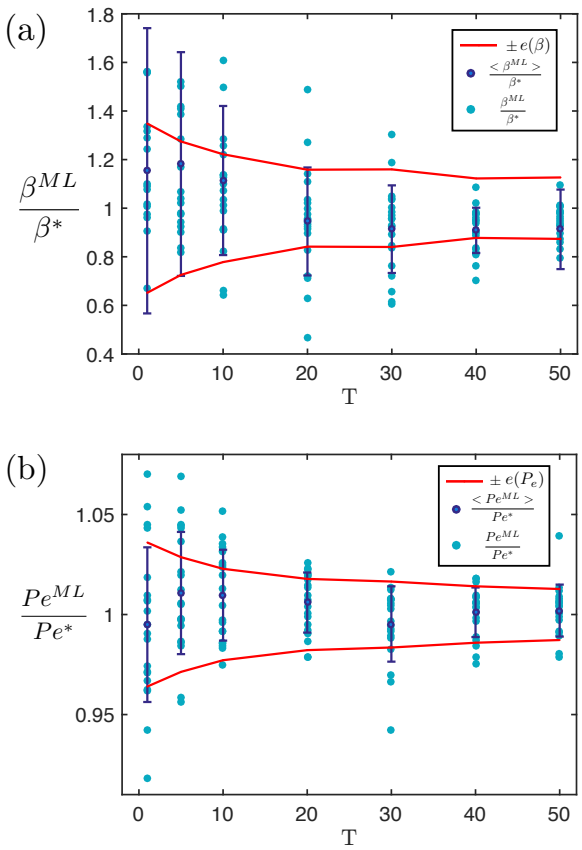

FIG. 5. Convergence of the estimated (a) β^{ML} and (b) Péclet Pe^{ML}. For each duration T, 20 trajectories were simulated. Each cyan point corresponds to the estimation of one trajectory, the dark blue points are the average of these estimations over the 20 trajectories. The red curves represent the uncertainties computed using the expressions Eq. (15) (resp. Eq. (16)).

V. EXPERIMENTAL DETERMINATION OF THE PARAMETERS

In this section, we give an example of the applicability of the ML method on experimental bacterial trajectories. A set of Lagrangian 3D swimming trajectories of mutant bacteria with inhibited tumbling (smooth swimmer δ CheY) was recorded at a frequency $f_{ac} = 80\,\mathrm{Hz}$ for freely swimming bacteria and $f_{ac} = 100\,\mathrm{Hz}$ for bacteria in flow. Further details on the experimental setup are given in the Appendix.

We first consider the free swimmer case. The absence of flow makes the determination of the rotational diffusion coefficient easier and standard methods can be applied. For each trajectory, we computed the correlation function of the orientation vector $\mathbf{p}$ as a function of time. It decays exponentially with a characteristic time scale $\tau = (2D_R)^{-1}$, which provides a first estimation of the rotational diffusion coefficient for each bacterium.

To be able to use the ML method, the time interval between two samplings, Δt has to fulfill some conditions. As discussed in Sec. III B, it has to be small enough so that our approximations are reliable. On the other hand, if Δt is too small, the change in the bacterium orientation between any instants t and $t+\Delta t$ will be dominated by noise coming from the measurement and not by real physical contributions. The optimal value of f_{ac} can be estimated from the theoretical value of the Brownian diffusion coefficient obtained by Perrin [33, 34]:

$$D_B = \frac{3k_B T ln(2l/a)}{\pi\eta l^3} \tag{19}$$

where l is the length of the ellipsoid and a its width, while η denotes de viscosity of the fluid at a temperature T. For an elongated objet of length $l = 10\,\mu\mathrm{m}$ and width $a = 1\,\mu\mathrm{m}$ (typical bacterium dimensions), it gives $D_B \approx 0.012\,\mathrm{s}^{-1}$. One should then have $f \gg D_B$; accordingly, we resampled our trajectories at a frequency $1/f = \Delta t = 1/3\,\mathrm{s}$.

By varying the point from where the resampling starts, we obtained from one trajectory $\Gamma(1/f_{ac})$, a set of trajectories $\Gamma_i(1/f)$, $i = 1, ..., N$, $N = f_{ac}/f$. We applied the ML method on each resampled trajectory Γ_i to obtain N values of the estimated D_R that we averaged to get D_R^{ML}; the ML estimation of the rotational diffusion coefficient of the trajectory. We then applied this method on each of the trajectories. Fig. 6(a) displays a scatter plot where D_R is in the vertical axis, while the horizontal axis corresponds to D_R^{ML}. One can immediately see that both methods give similar results and that the ML estimation works remarkably well.

In our second experimental test, we injected a suspension of smooth swimmers in a microchannel of rectangular cross section (height $h = 100\,\mu\mathrm{m}$ and width $w = 600\,\mu\mathrm{m}$). Dozens of 3D bacterial trajectories were recorded at different flow rates. We focused on bulk trajectories that are at least $10\,\mu\mathrm{m}$ away from the top and the bottom walls and that last at least $10\Delta t$ ($\Delta t = 1/3\,\mathrm{s}$). The ML method is then used on those trajectories to determine the Péclet number. Fig. 6(b) displays the results of the estimated Péclet number Pe^{ML} on experimental bacterial trajectories. One can see that Pe^{ML} increases with $\dot{\gamma}_W$, as expected for a constant rotational diffusion coefficient. By a linear fit on the average values of Pe^{ML}, we get the rotational diffusion coefficient of the bacteria which is the inverse of the slope. It then gives $D^{ML} = 0.032 \pm 0.016\,\mathrm{s}^{-1}$ which is consistent with $< D_R^{ML} >= 0.029\,\mathrm{s}^{-1}$ measured in absence of flow, as well as with values reported in presence of flow [26].

VI. DISCUSSION

We have developed an ML method suitable to infer parameters from individual stochastic trajectories. We were able to estimate the parameters of the SABJ model

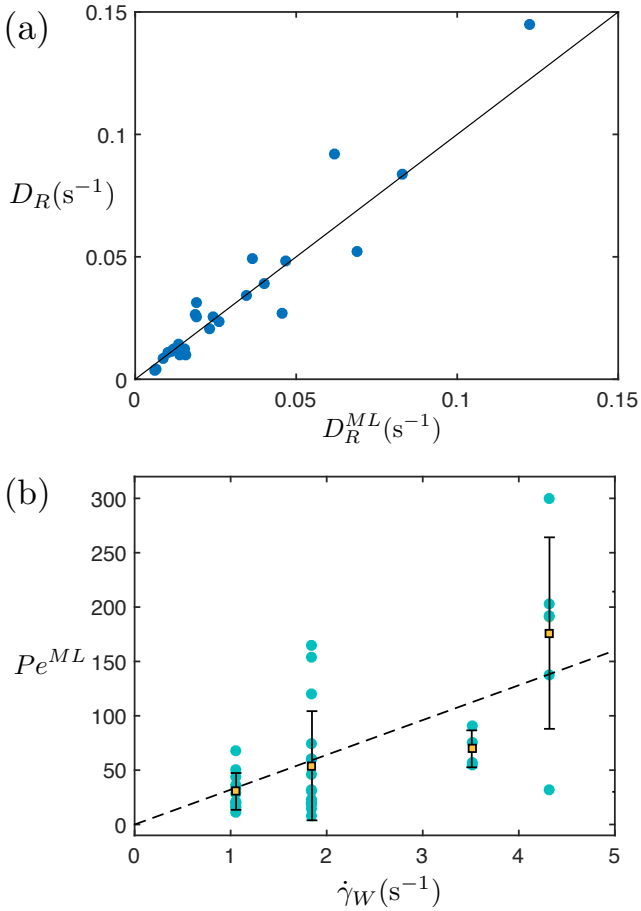

FIG. 6. (a) Comparison between the rotational diffusion coefficients obtained by the orientation decorrelation method (D_R) and with the ML method (D_R^{ML}). Each point corresponds to a track. The black line corresponds to the equation $y = x$. Average values are $< D_R >= 0.031\,\mathrm{s}^{-1}$, with standard deviation $\sigma_R = 0.031\,\mathrm{s}^{-1}$ and $< D_R^{ML} >= 0.029\,\mathrm{s}^{-1}$ with standar deviation $\sigma_R^{ML} = 0.027\,\mathrm{s}^{-1}$. (b) Experimental determination of Pe^{ML}. Squares are the average value of all the Pe^{ML} at a given $\dot{\gamma}_W$; the error bars represent the corresponding standard deviation at a given $\dot{\gamma}_W$. The black dashed line corresponds to a linear fit over the averaged estimated values: $< Pe^{ML} >= \dot{\gamma}_W/D^{ML}$, with $D^{ML} = 0.032 \pm 0.016\,\mathrm{s}^{-1}$.

accurately using both, numerical simulations and experiments. Preliminary results obtained from the analysis of experimental tracks are encouraging. One of the outcomes of our numerical study is that uncertainties fit well an $O(N^{1/2})$ decay even for relatively short trajectories. As an important remark, note such estimates can in principle be further improved in a systematic way by considering errors up to $O(N^{-1})$.

Our estimators for the rotational diffusion coefficient are much more scattered in experiments than in simulations. In particular, standard deviations in Figs. 6(a) and (b) are large. However, it is important to keep in mind that each point in Fig. 6 represents a different bacterium, which implies that one should be careful not to conflate biological variability with a lack of accuracy of the method. In support of the hypothesis of individual variability, note that the shape of a bacterium affects the value of D_R, which depends sensitively on the effective value of the ellipsoid long axis l. Such l^{-3} dependence may naturally lead to a rather broad distribution of coefficient values within a population, which is indeed what was found. For instance, for an effective ellipsoid of length $l = 5$ to $10\,\mu\mathrm{m}$ and width $a = 1\,\mu\mathrm{m}$, the estimation of D_R using Eq. (19), yields $D_R = 1.25$ to $5 \times 10^{-3}\mathrm{s}^{-1}$. We thus conclude that there is an intrinsic variability of D_R within a bacterial population, which is precisely why one needs a method like the one developed in this work to analyze individual trajectories.

The tool we have devised here is general in a natural way and can be applied to any Markovian and time-translationally invariant stochastic process. In the context of swimming bacteria, one can study variations of parameters due to external conditions and physical effects that are not considered in the SABJ model. For instance, shear may shorten the flagella bundle leading to higher effective D_R (i.e lower Pe). Modifying the SABJ model as to account for rheotactic effects [35, 36], and addapting our ML method appropriately, one could study rheotactic drift [37] in detail, analyze possible changes in bundle formation due to shear, and probe the flexibility of the hook at the base of the flagellum and its interplay with external flow. These and other physical questions will be addressed in the immediate future.

ACKNOWLEDGMENTS

This work was supported by the ANR grant "BacFlow" ANR-15-CE30-0013 and by Institut Pierre-Gilles de Gennes (Équipement d'Excellence, Investissements d'Avenir program ANR-10-EQPX-34). RGG acknowledges support fron ANR grant ANR-17-CE08-0047-02. E.C. is supported by the Institut Universitaire de France. We thank Carine Douarche for providing us the E.coli CR20 smooth mutant. We are also very thankful to the Edinburgh group of Pr Wilson Poon and in particular to Angela Dawson and Vincent Martinez, for providing the 2-color AD63 E.coli strain we display on Fig.1.

Appendix A: Approximate log-likelihood for the SABJ model

In this secction we go through the full calculation leading to the result (11) for the log-likelihood of the SABJ model. Our starting point is the MSRDJ path-integral representation of the process (1)-(2), which is given by Eqs. (7) and (8) with $\mathcal{J}=1$. As mentioned in the text, one approximates the time integrals by discrete sums in the action when the sampling frequency is large enough. We write

$$\hat{\mathcal{A}}[\{\Gamma,\hat{\Gamma}\}]=\frac{1}{f}\sum_{\alpha}\Big\{\mathrm{i}\big[f\Delta\mathbf{r}_{\alpha}-\mathbf{v}_{\alpha}\big]^{\mathrm{T}}\hat{\mathbf{r}}_{\alpha}+\mathrm{i}\big[f\Delta\mathbf{p}_{\alpha}-\mathbf{h}_{\alpha}\big]^{\mathrm{T}}\hat{\mathbf{p}}_{\alpha}+D_R\hat{\mathbf{p}}_{\alpha}^{\mathrm{T}}\tilde{\mathbb{M}}(\mathbf{p}_{\alpha})\hat{\mathbf{p}}_{\alpha}\Big\}, \tag{A1}$$

where Γ_α ($\hat{\Gamma}_\alpha$) is a shorthand notation for $\Gamma(t_\alpha)$ ($\hat{\Gamma}(t_\alpha)$), while $\Delta\Gamma_\alpha\equiv\Gamma_{\alpha+1}-\Gamma_\alpha$. Additionally, we have introduced the notations $\mathbf{v}_\alpha\equiv\mathbf{v}(\Gamma_\alpha)$ and $\mathbf{h}_\alpha\equiv\mathbf{h}(\Gamma_\alpha)$. With all this, we write

$$\mathcal{P}[\{\mathbf{r},\mathbf{p}\}]=\int\prod_{\alpha}d^3\hat{\mathbf{r}}_{\alpha}d^3\hat{\mathbf{p}}_{\alpha}\exp\Big(-\frac{1}{f}\sum_{\alpha}\Big\{\mathrm{i}\big[f\Delta\mathbf{r}_{\alpha}-\mathbf{v}_{\alpha}\big]^{\mathrm{T}}\hat{\mathbf{r}}_{\alpha}+\mathrm{i}\big[f\Delta\mathbf{p}_{\alpha}-\mathbf{h}_{\alpha}\big]^{\mathrm{T}}\hat{\mathbf{p}}_{\alpha}+D_R\hat{\mathbf{p}}_{\alpha}^{\mathrm{T}}\tilde{\mathbb{M}}(\mathbf{p}_{\alpha})\hat{\mathbf{p}}_{\alpha}\Big\}\Big). \tag{A2}$$

The integrals over $\{\hat{\mathbf{r}}_\alpha\}$ are immediate and can be readily performed to yield

$$\mathcal{P}[\{\mathbf{r},\mathbf{p}\}]=\prod_{\alpha}\delta\big(\Delta\mathbf{r}_{\alpha}-f^{-1}\mathbf{v}_{\alpha}\big)\int\prod_{\alpha}d^3\hat{\mathbf{p}}_{\alpha}\exp\Big(-\frac{1}{f}\sum_{\alpha}\Big\{\mathrm{i}\big[f\Delta\mathbf{p}_{\alpha}-\mathbf{h}_{\alpha}\big]^{\mathrm{T}}\hat{\mathbf{p}}_{\alpha}+D_R\hat{\mathbf{p}}_{\alpha}^{\mathrm{T}}\tilde{\mathbb{M}}(\mathbf{p}_{\alpha})\hat{\mathbf{p}}_{\alpha}\Big\}\Big). \tag{A3}$$

The remaining integrals over $\{\hat{\mathbf{p}}_\alpha\}$ are less trivial because, as can be verified, the matrix $\tilde{\mathbb{M}}$ is singular, i.e., $\det(\tilde{\mathbb{M}})=0$. Such singularity is linked to the fact that the continuous dynamics Eq. (2) imposes a hard constraint on $\mathbf{p}$, i.e., $|\mathbf{p}|^2=1$. This means that at any time instant only two numbers are needed to fully describe the state of the three-dimensional vector $\mathbf{p}$. For instance, by knowing the projection of $\mathbf{p}$ on the z axis, p_z, and the polar angle θ of the projection of $\mathbf{p}$ in the $x-y$ plane, one can write $\mathbf{p}=(\sqrt{1-p_z^2}\cos\theta,\sqrt{1-p_z^2}\sin\theta,p_z)$.

From this discussion, one can infer that there is a relevant two-dimensional sub-space allowing to fully describe $\mathbf{p}$. One can indeed isolate the singularity, so that a Gaussian integration can still be made in a reduced, two-dimensional submanifold. To show this, we introduce the change of variables $\hat{\mathbf{p}}_\alpha=\mathbb{R}_\alpha\boldsymbol{\omega}_\alpha$, where the matrix $\mathbb{R}_\alpha$ is given as

$$\mathbb{R}_{\alpha}=\begin{bmatrix}p_{x,\alpha} & -p_{z,\alpha} & -\frac{p_{y,\alpha}p_{x,\alpha}^2}{p_{x,\alpha}^2+p_{z,\alpha}^2}\\ p_{y,\alpha} & 0 & p_{x,\alpha}\\ p_{z,\alpha} & p_{x,\alpha} & -\frac{p_{x,\alpha}p_{y,\alpha}p_{z,\alpha}}{p_{x,\alpha}^2+p_{z,\alpha}^2}\end{bmatrix}. \tag{A4}$$

One can check that the determinant of $\mathbb{R}_\alpha$ is given by $\det(\mathbb{R}_\alpha)=-p_{x,\alpha}|\mathbf{p}_\alpha|^2$. Introducing the new vector

$$\boldsymbol{\rho}_{\alpha}=\mathbb{R}_{\alpha}^{\mathrm{T}}[f\Delta\mathbf{p}_{\alpha}-\mathbf{h}_{\alpha}], \tag{A5}$$

we have:

$$\mathcal{P}[\{\mathbf{r},\mathbf{p}\}]=\prod_{\alpha}\delta\big(\Delta\mathbf{r}_{\alpha}-f^{-1}\mathbf{v}_{\alpha}\big)I[\{\mathbf{r},\mathbf{p}\}], \tag{A6}$$

where

$$I[\{\mathbf{r},\mathbf{p}\}]=\int\prod_{\alpha}d^3\boldsymbol{\omega}_{\alpha}\exp\Big(-\frac{1}{f}\sum_{\alpha}\Big\{\mathrm{i}\boldsymbol{\rho}_{\alpha}^{\mathrm{T}}\boldsymbol{\omega}_{\alpha}+D_R\boldsymbol{\omega}_{\alpha}^{\mathrm{T}}\mathbb{\Gamma}_{\alpha}\boldsymbol{\omega}_{\alpha}-f\ln\big(|p_{x,\alpha}||\mathbf{p}_{\alpha}|^2\big)\Big\}\Big), \tag{A7}$$

and the matrix $\mathbb{\Gamma}_\alpha$ has the following form:

$$\mathbb{\Gamma}_{\alpha}=\begin{bmatrix}0 & 0 & 0\\ 0 & (p_{x,\alpha}^2+p_{z,\alpha}^2)|\mathbf{p}_{\alpha}|^2 & 0\\ 0 & 0 & \frac{p_{x,\alpha}^2|\mathbf{p}_{\alpha}|^4}{p_{x,\alpha}^2+p_{z,\alpha}^2}\end{bmatrix}. \tag{A8}$$

The integrals over $\{\boldsymbol{\omega}_\alpha\}$ in (A7) can now be immediately performed to yield

$$I[\{\mathbf{r},\mathbf{p}\}] = \left(\frac{\pi f}{D_R}\right)^N \prod_\alpha \left[\frac{1}{|\mathbf{p}_\alpha|}\delta(f^{-1}\rho_{x,\alpha})\right] \exp\left(-\frac{1}{4D_R f}\sum_\alpha \left\{\frac{\rho_{y,\alpha}^2}{(p_{x,\alpha}^2+p_{z,\alpha}^2)|\mathbf{p}_\alpha|^2} + \frac{\rho_{z,\alpha}^2(p_{x,\alpha}^2+p_{z,\alpha}^2)}{p_{x,\alpha}^2|\mathbf{p}_\alpha|^4}\right\}\right). \tag{A9}$$

This result can be written in a more illustrative way if one calculates $\rho_{x,\alpha}$ explicitly using (A5). One has $f^{-1}\rho_{x,\alpha} = f^{-1}\mathbf{p}_\alpha\cdot[f\Delta\mathbf{p}_\alpha - \mathbf{h}_\alpha] \equiv \mathbf{p}_\alpha\cdot\Delta\mathbf{p}_\alpha + 2D_R f^{-1}|\mathbf{p}_\alpha|^2$. With all this we finally obtain that the probability of a given sequence $\{\mathbf{r},\mathbf{p}\}$, in this approximation, is

$$\mathcal{P}[\{\mathbf{r},\mathbf{p}\}] = \left(\frac{\pi f}{D_R}\right)^N \Sigma[\{\mathbf{r},\mathbf{p}\}] \exp\left(-\frac{1}{4fD_R}\sum_\alpha \left\{\frac{\rho_{y,\alpha}^2}{(p_{x,\alpha}^2+p_{z,\alpha}^2)|\mathbf{p}_\alpha|^2} + \frac{\rho_{z,\alpha}^2(p_{x,\alpha}^2+p_{z,\alpha}^2)}{p_{x,\alpha}^2|\mathbf{p}_\alpha|^4} + 4fD_R\ln|\mathbf{p}_\alpha|\right\}\right), \tag{A10}$$

where $\Sigma[\{\mathbf{r},\mathbf{p}\}]$ is a singular measure inforcing a set of constraints on the trajectories generated by Eqs. (1) and (2):

$$\Sigma[\{\mathbf{r},\mathbf{p}\}] = \prod_\alpha \delta\big(\Delta\mathbf{r}_\alpha - f^{-1}\mathbf{v}_\alpha\big)\prod_\alpha \delta\big(\mathbf{p}_\alpha\cdot\Delta\mathbf{p}_\alpha + 2D_R f^{-1}|\mathbf{p}_\alpha|^2\big). \tag{A11}$$

Let us analyze these constraints in more detail. The condition $\mathbf{p}_\alpha\cdot\Delta\mathbf{p}_\alpha + 2D_R f^{-1}|\mathbf{p}_\alpha|^2 = 0$ is related the conservation of the norm of $|\mathbf{p}|$. Note that, in particular, when $f\to\infty$, it takes the form $\mathbf{p}(t)\cdot d\mathbf{p}(t) = 0$, as one would expect. On the other hand, the condition $\Delta\mathbf{r}_\alpha - f^{-1}\mathbf{v}_\alpha = 0$ provides, in practice, a practical way to determine p_α at each time step. Indeed, the tracking device measures the trajectory of the particle, while p has to be determined indirectly. Writing this condition more explicity as $\Delta\mathbf{r}_\alpha - f^{-1}(v_0\mathbf{p}_\alpha + \mathbf{v}_\alpha^F) = 0$, one has $\mathbf{p}_\alpha = v_0^{-1}(f\Delta\mathbf{r}_\alpha - \mathbf{v}_\alpha^F)$, which is the expression that is used in practice to determine $\mathbf{p}_\alpha$ at each time step. In summary, if we restrict our analysis to the relevant sub-space of the trajectories which are compatible with the constraints Σ (inforcing, in particular, that $|\mathbf{p}_\alpha| = 1$ for all α), we have the following expression for the discretized path probability:

$$\mathcal{P}_\Sigma[\{\mathbf{r},\mathbf{p}\}] = \left(\frac{\pi f}{D_R}\right)^N \exp\left(-\frac{1}{4fD_R}\sum_\alpha \left\{\frac{\rho_{y,\alpha}^2}{p_{x,\alpha}^2+p_{z,\alpha}^2} + \frac{\rho_{z,\alpha}^2(p_{x,\alpha}^2+p_{z,\alpha}^2)}{p_{x,\alpha}^2}\right\}\right). \tag{A12}$$

To proceed further, we introduce the notations $\mathbf{a}_\alpha = f\Delta\mathbf{p}_\alpha - (\mathbb{1}-\mathbf{p}_\alpha\otimes\mathbf{p}_\alpha)\mathbb{\Omega}(\mathbf{r}_\alpha)\mathbf{p}_\alpha$, and $\mathbf{b}_\alpha = (\mathbb{1}-\mathbf{p}_\alpha\otimes\mathbf{p}_\alpha)\mathbb{E}(\mathbf{r}_\alpha)\mathbf{p}_\alpha$, so that we have $f\Delta\mathbf{p}_\alpha - \mathbf{h}_\alpha \equiv \mathbf{a}_\alpha - \beta\mathbf{b}_\alpha + 2D_R\mathbf{p}_\alpha$. We also define

$$\begin{aligned}
S_\alpha &= p_{x,\alpha}^2 + p_{z,\alpha}^2\\
A_\alpha &= a_{z,\alpha}p_{x,\alpha} - a_{x,\alpha}p_{z,\alpha},\\
B_\alpha &= b_{z,\alpha}p_{x,\alpha} - b_{x,\alpha}p_{z,\alpha},\\
C_\alpha &= a_{y,\alpha}S_\alpha - (a_{x,\alpha}p_{x,\alpha} + a_{z,\alpha}p_{z,\alpha})p_{y,\alpha},\\
E_\alpha &= b_{y,\alpha}S_\alpha - (b_{x,\alpha}p_{x,\alpha} + b_{z,\alpha}p_{z,\alpha})p_{y,\alpha},\\
A_R &= f^{-2}\sum_\alpha S_\alpha^{-1}(A_\alpha^2 + C_\alpha^2),\\
B_R &= f^{-2}\sum_\alpha S_\alpha^{-1}(A_\alpha B_\alpha + C_\alpha E_\alpha),\\
C_R &= f^{-2}\sum_\alpha S_\alpha^{-1}(B_\alpha^2 + E_\alpha^2).
\end{aligned} \tag{A13}$$

The log-likelihood can be derived by taking the logarithm of Eq. (A12) after having into account that $|\mathbf{p}_\alpha| = 1$ for all α in any experimental or numerical track, and that Eq. (1) is used to fix the evolution of the vector $\mathbf{p}$. To take the logarithm, we define $\tilde{\mathcal{P}} = \pi^{-N}\mathcal{P}_\Sigma$, with $\mathcal{P}_\Sigma$ given by (A12). It is also convenient to introduce the dimensionless inverse diffusion coeffient $\tilde{\mathcal{N}}_R = D_R^{-1}f$. In terms of the quantities introduced in (A13), we write $S = \ln\tilde{\mathcal{P}}$, which gives Eq. (11).

Appendix B: Experimental set-up and protocol

Bacteria used in this work are smooth swimmer mutants of an *E. coli* (strain CR20, Δ-CheY) that almost never tumble and were transformed with a plasmid coding for a yellow fluorescent protein (YFP). Bacteria are grown overnight at 30°C until the early stationary phase. The growth medium is then removed by centrifuging the

culture and removing the supernatant. The bacteria are resuspended in a Motility Buffer (MB) below the very low concentration of 3×10^7 bacteria per mL, in order to visualize one bacterium at a time and to minimize the interactions between bacteria. The suspension is supplemented with L-serine at 0.08g/mL and polyvinyl pyrrolidone (PVP) at 0.005%; L-serine maintains good motility for a few hours and PVP is used to prevent bacteria from sticking to surfaces. The solution is also mixed with Percol (1:1) to avoid bacterial sedimentation. The experiments are performed at a temperature of 25^oC.

The channel is visualized using a home-made Lagrangian tracking microscope [38] here used to track fluorescent swimming bacteria. By a visualization based feedback acting on a mechanical (horizontal) and piezoelectric (vertical) stage, the targeted object is kept close to the center of the visualization field and in focus on an inverted microscope (Zeiss-Observer, Z1 with an objectif C-Apochromat 63x/1.2 W). Images of the tracked objects are acquired at 80 or 100 Hz with a Hamamatsu Orca-flash 4.0 camera. Simultaneously, the three-dimensional positions of the object are recorded.

The measurements take place in a microfluidic channel of rectangular cross-section (height $h = 100\ \mu m$, width $w = 600\ \mu m$), made in Polydimethylsiloxane (PDMS) using standard soft-lithography techniques. Flow is imposed through the channel via a Nemesys syringe pump (dosing unit Low Pressure Syringe Pump neMESYS 290N and base Module BASE 120N). The flow rate varies from 1 to 4.3 nL s^{-1} corresponding respectively to U_m between $(28 \pm 1.9)\ \mu$m/s and $(120 \pm 4.0)\ \mu$m/s and maximal shear rates $\dot{\gamma}_M = 4U_m/h$ between (1.12 ± 0.076) s^{-1} and (4.82 ± 0.16) s^{-1}. We set our region of interest in the center of the channel with respect to its width and consider only trajectories which are at least $200\mu m$ away from the lateral walls.

* reinaldomeister@gmail.com

[1] C. Manzo and M. F. Garcia-Parajo, Rep. Prog. Phys. **78**, 124601 (2015).
[2] R. Sarfati, J. Bawzdziewicz, and E. R. Dufresne, Soft Matter **13**, 2174 (2017).
[3] M. Seyrich, Z. Alirezaeizanjani, C. Beta, and H. Stark, New J. Phys. **20**, 103033 (2018).
[4] L. Pérez García, J. Donlucas Pérez, G. Volpe, A. V. Arzola, and G. Volpe, Nat. Commun. **9**, 5166 (2018).
[5] A. Frishman and P. Ronceray, Phys. Rev. X **10**, 021009 (2020).
[6] Y. Lanoiselée and D. S. Grebenkov, Phys. Rev. E **93**, 052146 (2016).
[7] Y. Lanoiselée and D. S. Grebenkov, Phys. Rev. E **96**, 022144 (2017).
[8] M. L. Gardel, M. T. Valentine, and D. A. Weitz, in *Microscale Diagnostic Techniques*, edited by K. Breuer (Springer, Oxford, 2005) pp. 1–50.
[9] B. Abou and F. m. c. Gallet, Phys. Rev. Lett. **93**, 160603 (2004).
[10] A. M. Puertas and T. Voigtmann, J. Phys. Condens. Matter **26**, 243101 (2014).
[11] D. Wirtz, Annu. Rev. Biophys. **38**, 301 (2009).
[12] P. Struntz and M. Weiss, Phys. Chem. Chem. Phys. **20**, 28910 (2018).
[13] X.-L. Wu and A. Libchaber, Phys. Rev. Lett. **84**, 3017 (2000).
[14] G. Miño, T. E. Mallouk, T. Darnige, M. Hoyos, J. Dauchet, J. Dunstan, R. Soto, Y. Wang, A. Rousselet, and E. Clement, Phys. Rev. Lett. **106**, 048102 (2011).
[15] R. Jeanneret, D. O. Pushkin, V. Kantsler, and M. Polin, Nat. Commun. **7**, 12518 (2016).
[16] D. T. N. Chen, A. W. C. Lau, L. A. Hough, M. F. Islam, M. Goulian, T. C. Lubensky, and A. G. Yodh, Phys. Rev. Lett. **99**, 148302 (2007).
[17] G. L. Miño, J. Dunstan, A. Rousselet, E. Clément, and R. Soto, J. Fluid Mech. **729**, 423 (2013).
[18] T. V. Kasyap, D. L. Koch, and M. Wu, Phys. Fluids **26**, 081901 (2014).
[19] U. Marini Bettolo Marconi and C. Maggi, Soft Matter **11**, 8768 (2015).
[20] E. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. van Wijland, Phys. Rev. Lett. **117**, 038103 (2016).
[21] L. Dabelow, S. Bo, and R. Eichhorn, Phys. Rev. X **9**, 021009 (2019).
[22] E. W. Burkholder and J. F. Brady, J. Chem. Phys. **150**, 184901 (2019).
[23] J. Katuri, K. D. Seo, D. S. Kim, and S. Snchez, Lab Chip **16**, 1101 (2016).
[24] L. Baraban, S. M. Harazim, S. Sanchez, and O. G. Schmidt, Angew. Chem. Int. Ed. **52**, 5552 (2013).
[25] J. Palacci, S. Sacanna, A. Abramian, J. Barral, K. Hanson, A. Y. Grosberg, D. J. Pine, and P. M. Chaikin, Sci. Adv. **1** (2015).
[26] G. Junot, N. Figueroa-Morales, T. Darnige, A. Lindner, R. Soto, H. Auradou, and E. Clément, EPL **126**, 44003 (2019).
[27] S. Thapa, M. A. Lomholt, J. Krog, A. G. Cherstvy, and R. Metzler, Phys. Chem. Chem. Phys. **20**, 29018 (2018).
[28] J.-B. Masson, G. Voisinne, J. Wong-Ng, A. Celani, and M. Vergassola, Proc. Natl. Acad. Sci. U.S.A. **109**, 1802 (2012).
[29] P. C. Martin, E. D. Siggia, and H. A. Rose, Phys. Rev. A **8**, 423 (1973).
[30] C. de Dominics, J. Phys. (Paris) Colloq. **37**, C1 (1976).
[31] H.-K. Janssen, Z. Phys. B **23**, 377 (1976).
[32] W. K. Newey and D. McFadden (Elsevier, 1994) pp. 2111 – 2245.
[33] F. Perrin, J. Phys. Radium **5**, 497 (1934).
[34] F. Perrin, J. Phys. Radium **7**, 1 (1936).
[35] K. Son, J. S. Guasto, and R. Stocker, Nat. Phys. **9**, 494 (2013).
[36] I. Spring, V. A. Martinez, C. Hotz, J. Schwarz-Linek, K. L. Grady, J. M. Nava-Sedeo, T. Vissers, H. M. Singer, M. Rohde, C. Bourquin, H. Hatzikirou, W. C. K. Poon, Y. S. Dufour, and M. Erhardt, PLOS Biology **16**, 1 (2018).
[37] Marcos, H. C. Fu, T. R. Powers, and R. Stocker, Proc. Natl. Acad. Sci. U.S.A. **109**, 4780 (2012).
[38] T. Darnige, N. Figueroa-Morales, P. Bohec, A. Lindner,

and E. Clément, Rev. Sci. Instrum. **88**, 055106 (2017).