HAL open science

Simulating individual-based models of bacterial chemotaxis with asymptotic variance reduction

Mathias Rousset, Giovanni Samaey

To cite this version:

Mathias Rousset, Giovanni Samaey. Simulating individual-based models of bacterial chemotaxis with asymptotic variance reduction. Mathematical Models and Methods in Applied Sciences, 2013, 23 (12), pp.2155 - 2191. 10.1142/S0218202513500292 . hal-00643324

HAL Id: hal-00643324

https://inria.hal.science/hal-00643324v1

Submitted on 22 Nov 2011

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.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Simulating individual-based models of bacterial chemotaxis with asymptotic variance reduction

Mathias Rousset* and Giovanni Samaey†

We discuss variance reduced simulations for an individual-based model of chemotaxis of bacteria with internal dynamics. The variance reduction is achieved via a coupling of this model with a simpler process in which the internal dynamics has been replaced by a direct gradient sensing of the chemoattractants concentrations. In the companion paper [26], we have rigorously shown, using a pathwise probabilistic technique, that both processes converge towards the same advection-diffusion process in the diffusive asymptotics. In this work, a direct coupling is achieved between paths of individual bacteria simulated by both models, by using the same sets of random numbers in both simulations. This coupling is used to construct a hybrid scheme with reduced variance. We first compute a deterministic solution of the kinetic density description of the direct gradient sensing model; the deviations due to the presence of internal dynamics are then evaluated via the coupled individual-based simulations. We show that the resulting variance reduction is *asymptotic*, in the sense that, in the diffusive asymptotics, the difference between the two processes has a variance which vanishes according to the small parameter.

MSC: 35Q80, 92B05, 65C35.

Key words: asymptotic variance reduction, bacterial chemotaxis, velocity-jump process.

1. Introduction

The motion of flagellated bacteria can be modeled as a sequence of run phases, during which a bacterium moves in a straight line at constant speed. Between two run phases, the bacterium changes direction in a tumble phase, which takes much less time than the run phase and acts as a reorientation. To bias movement towards regions with high concentration of chemoattractant, the bacterium adjusts its turning rate by increasing, resp. decreasing, the probability of tumbling when moving in an unfavorable, resp. favorable, direction [3,29]. Since many species are unable to sense chemoattractant gradients reliably due to their small size, this adjustment is often done via an intracellular mechanism that allows the bacterium to retain information on the history of the chemoattractant concentrations along its path [5]. The resulting model, which will be called the "internal state" or "fine-scale" model in this text, can be formulated as a velocity-jump process, combined with an ordinary differential equation (ODE) that describes the evolution of an internal state that incorporates this memory effect [12].

*SIMPAF, INRIA Lille - Nord Europe, Lille, France, (mathias.rousset@inria.fr).
†Department of Computer Science, K.U. Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium, (giovanni.samaey@cs.kuleuven.be).

The probability density distribution of the velocity-jump process evolves according to a kinetic equation, in which the internal variables appear as additional dimensions. A direct deterministic simulation of this equation is therefore prohibitively expensive, and one needs to resort to a stochastic particle method. Unfortunately, a direct fine-scale simulation using stochastic particles presents a large statistical variance, even in the diffusive asymptotic regime, where the behavior of the bacterial density is known explicitly to satisfy a Keller-Segel advection-diffusion equation. Consequently, it is difficult to assess accurately how the solutions of the fine-scale model differ from their advection-diffusion limit in intermediate regimes. We refer to [16,17,18] for numerous historical references on the Keller-Segel equation, to [11,12] for formal derivations (based on moment closures) of convergence of the the velocity-jump process with internal state to the Keller-Segel equation, and to the companion paper [26] for a proof of this convergence based on probabilistic arguments.

In this paper, we propose and analyze a numerical method to simulate individual-based models for chemotaxis of bacteria with internal dynamics with reduced variance. The variance reduction is based on a coupling technique (control variate): the main idea is to simultaneously simulate, using the same random numbers, a simpler, "coarse" process where the internal dynamics is replaced by a direct "gradient sensing" mechanism (see [2,23,25] for references on such gradient sensing models). The probability density of the latter satisfies a kinetic equation without the additional dimensions of the internal state, and converges to a similar advection-diffusion limit, see e.g. [6,15,24,26].

More precisely, we consider the coupling of two velocity-jump processes with exactly the same advection-diffusion limit (for both processes, a proof of the latter convergence has been obtained in the companion paper [26]). The first one is the process with internal dynamics described by the system of equations (2.4), with jump rates satisfying (2.3) and internal dynamics satisfying the two main assumptions (2.7) and (2.9) as well as other technical assumptions contained in Section 2.2. The second one is the process with direct gradient sensing (2.12), with jump rates satisfying (2.14) with (2.15). The same random numbers are used for the two processes in the definition of both the jump times and the velocity directions. The main contributions of the present paper are then twofold :

- From a numerical point of view, we couple two systems of many particles consisting of different realizations of the fine-scale (internal state) and coarse (gradient sensing) processes, simulated with similar discretization schemes (see Section 3.1) and the same random numbers. Additionally, for the coarse model, also the continuum description is simulated on a spatial grid. Evolution of the fine-scale model is then evaluated on the grid at hand by adding the difference between the two coupled particle descriptions to the evolution of the coarse continuum description. This can be seen as using the coarse process as a control variate. As coupling is lost over time, a regular reinitialization of the two coupled particle systems is included in

Model	Numerics
Internal state model–Individual description	Monte-Carlo method
Gradient sensing model–Individual description	Monte-Carlo method (coupled)
Gradient sensing model–Density description	Grid method

Table 1. Models and numerical methods used in the variance reduction algorithm.

the algorithm. The asymptotic variance reduction method can be depicted using Table 1 (see also Figure 1 in Section 3).

- From an analysis point of view, we rigorously prove, using probabilistic arguments, L^p-bounds on the position difference at diffusive times of both processes in terms of appropriate powers of the scaling parameter (Theorem 4.7). This analysis shows that the numerical variance reduction obtained by the method is *asymptotic* (the remaining statistical variance vanishes with the small scaling parameter), and provides upper bounds on the rate of the latter convergence. We argue in Section 4.3 that this bound should be sharp in relevant regimes.

The idea of asymptotic variance reduction is a general and very recent idea in scientific computing that appears when using hybrid Monte-Carlo/PDE (partial differential equation) methods. While the use of control variates is fairly common in Monte Carlo simulation, the only explicit attempt to develop asymptotic variance reduction in hybrid methods, up to our knowledge, can be found in [7] and related papers in the context of the Boltzmann equation, and is based on an importance sampling approach (see also [4] and related papers).

This paper is organized as follows. In Section 2, we briefly review the mathematical models that we will consider, and summarize the results on their advection-diffusion limit that were obtained in the companion paper [26]. In Section 3, we proceed to describe the coupled numerical method in detail. Section 4 is dedicated to a detailed analysis of the variance reduction of the coupled method. The main result is the following: up to a time shift perturbation, the variance of the difference between the two coupled process scales according to the small parameter of the diffusive asymptotics. We proceed with numerical illustrations in Sections 5 and 6, and conclude in Section 7 with an outlook to future research.

2. Particle-based models for bacterial chemotaxis

In this paper, we directly present the models of interest in a nondimensional form, incorporating the appropriate space and time scales. For a discussion on the chosen scaling, we refer to the companion paper [26].

2.1. *Model with internal state*

We consider bacteria that are sensitive to the concentration of m chemoattractants $(\rho_i(x))_{i=1}^m$, with $\rho_i(x) \geq 0$ for $x \in \mathbb{R}^d$. While we do not consider time dependence of chemoattractant via production or consumption by the bacteria, a generalization to this situation is straightforward, at least for the definition of the models and the numerical method. Bacteria move with a constant speed v (run), and change direction at random instances in time (tumble), in an attempt to move towards regions with high chemoattractant concentrations. As in [11], we describe this behavior by a velocity-jump process driven by some internal state $y \in \mathbb{Y} \subset \mathbb{R}^n$ of each individual bacterium. The internal state models the memory of the bacterium and is subject to an evolution mechanism attracted by a function $\psi : \mathbb{R}^m \to \mathbb{R}^n$ of the chemoattractants concentrations,

$$S(x) := \psi(\rho_1(x), \ldots, \rho_m(x)) \in \mathbb{R}^n,$$

where x is the present position of the bacterium. S is assumed to be smooth with bounded derivatives up to order 2. We refer to Section 2.2 in [26] for some typical choices; in the concrete example in this paper, we choose $m = n = 1$, and $S(x) = \rho_1(x)$. By convention, the gradient of $S(x) : \mathbb{R}^d \to \mathbb{R}^n$ is a matrix with dimension

$$\nabla S(x) \in \mathbb{R}^{n \times d}.$$

We then denote the evolution of each individual bacterium position by $t \mapsto X_t$, with normalized velocity

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = \epsilon V_t, \qquad V_t \in \mathbb{V} = \mathbb{S}^{d-1},$$

with $\mathbb{S}^{d-1}$ the unit sphere in $\mathbb{R}^d$. Hence, V_t represents the direction and the parameter ϵ represents the size of the velocity. The evolution of the internal state is denoted by $t \mapsto Y_t$. The internal state adapts to the local chemoattractant concentration through an ordinary differential equation (ODE),

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = F_\epsilon(Y_t, S(X_t)), \tag{2.1}$$

which is required to have a unique fixed point $y^* = S(x^*)$ for every fixed value $x^* \in \mathbb{R}^d$. We also introduce the deviations from equilibrium $z = S(x) - y$. The evolution of these deviations is denoted as

$$t \mapsto Z_t = S(X_t) - Y_t.$$

The velocity of each bacterium is switched at random jump times $(T_n)_{n \geq 1}$ that are generated via a Poisson process with a time dependent rate given by $\lambda(Z_t)$, where $z \mapsto \lambda(z)$ is a smooth function satisfying

$$0 < \lambda_{\min} \leq \lambda(z) \leq \lambda_{\max}, \tag{2.2}$$

as well as,

$$\lambda(z) = \lambda_0 - b^T z + c_\lambda \mathcal{O}\left(|z|^k\right), \tag{2.3}$$

with $b \in \mathbb{R}^n$, $k \geq 2$, and $c_\lambda > 0$ is used to keep track of the non-linearity in the analysis. The new velocity at time T_n is generated at random according to a centered probability distribution $\mathcal{M}(dv)$ with $\int v\,\mathcal{M}(dv) = 0$, typically

$$\mathcal{M}(dv) = \sigma_{\mathbb{S}^{d-1}}(dv),$$

where $\sigma_{\mathbb{S}^{d-1}}$ is the uniform distribution on the unit sphere.

The resulting fine-scale stochastic evolution of a bacterium is then described by a left continuous with right limits (lcrl) process

$$t \mapsto (X_t, V_t, Y_t),$$

which satisfies the following differential velocity-jump equation :

$$\begin{cases} \dfrac{\mathrm{d}X_t}{\mathrm{d}t} = \epsilon V_t, \\ \dfrac{\mathrm{d}Y_t}{\mathrm{d}t} = F_\epsilon(Y_t, S(X_t)), \\ \displaystyle\int_{T_n}^{T_{n+1}} \lambda(Z_t)\mathrm{d}t = \theta_{n+1}, \qquad \text{with } Z_t := S(X_t) - Y_t, \\ V_t = \mathcal{V}_n \quad \text{for } t \in [T_n, T_{n+1}] \ , \end{cases} \tag{2.4}$$

with initial condition $X_0, \mathcal{V}_0 \in \mathbb{R}^d$, $Y_0 \in \mathbb{R}^n$ and $T_0 = 0$. In (2.4), $(\theta_n)_{n\geq 1}$ denote i.i.d. random variables with normalized exponential distribution, and $(\mathcal{V}_n)_{n\geq 1}$ denote i.i.d. random variables with distribution $\mathcal{M}(dv)$.

Example 2.1. For concreteness, we provide a specific example, adapted from [11], which will also be used later on to illustrate our results numerically. We consider $m = 1$, i.e., there is only one chemoattractant $S(x)$ and the internal dynamics (2.1) reduces to a scalar equation

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = \frac{S(x) - y}{\tau} = \frac{z}{\tau}. \tag{2.5}$$

For the turning rate $z \mapsto \lambda(z)$, we choose the following nonlinear strictly decreasing smooth function

$$\lambda(z) = 2\lambda_0 \left(\frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{\pi}{2\lambda_0} z \right) \right). \tag{2.6}$$

2.2. *Asymptotic regimes and linearizations*

In the adimensional models (2.4)-(2.12), a small parameter ϵ is introduced, which can be interpreted as the ratio of the typical time of random change of velocity direction to the typical time associated with chemoattractant variations as seen by the bacteria. In the present section, we present the main scaling assumptions on the internal dynamics that are necessary to consider the asymptotic regime $\epsilon \ll 1$, as well as the corresponding notation. Most of these assumptions were introduced in the companion paper [26], to which we refer for more details and motivation.

Remark 2.1 (Notation). Throughout the text, the Landau symbol $\mathcal{O}$ denotes a *deterministic* and globally Lipschitz function satisfying $\mathcal{O}(0)=0$. Its precise value *may vary* from line to line, may depend on all parameters of the model, but its Lipschitz constant is uniform in ϵ. In the same way, we will denote generically by $C>0$ a deterministic constant that may depend on all the parameters of the model except ϵ.

We assume that the ODE (2.1) driving the internal state is well approximated by a near equilibrium evolution equation in following sense :

Assumption 2.2. We have

$$F_\epsilon(y,s)=-{\tau_\epsilon}^{-1}(y-s)+\epsilon^{1-\delta}c_F\mathcal{O}(|s-y|^2), \tag{2.7}$$

where $\delta>0$ and $\tau_\epsilon\in\mathbb{R}^{n\times n}$ is an invertible constant matrix.

The constant c_F is used to keep track of the dependance on the non-linearity of F_ϵ in the analysis. For technical reasons that are specific to the analysis of the time shift in the coupling (Lemma 4.6), we need the following assumption on the scale of the non-linearity (this assumption was not necessary in the companion paper [26]):

Assumption 2.3. Assume that $|y-s|=\mathcal{O}(\epsilon^\delta)$, then there is a $\gamma>0$ such that:

$$|(y-s)-\tau_\epsilon F_\epsilon(y,s)|=c_F\mathcal{O}(\epsilon^{1+\gamma}). \tag{2.8}$$

We now formulate the necessary assumptions on the scale of the matrix τ_ϵ :

Assumption 2.4. There is a constant $C>0$ such that for any $t\geq 0$, one has

$$\left\|\mathrm{e}^{-t{\tau_\epsilon}^{-1}}\right\|\leq C\mathrm{e}^{-t\epsilon^{1-\delta}/C}. \tag{2.9}$$

Assumption 2.5. There is a constant $C>0$ such that

$$\sup_{t\geq 0}\left\|t{\tau_\epsilon}^{-1}\mathrm{e}^{-t{\tau_\epsilon}^{-1}}\right\|\leq C\epsilon^{-1}.$$

Assumption 2.4 is used to ensure exponential convergence of the linear ODE with time scale at least of order $\epsilon^{1-\delta}$. Assumption 2.5 is necessary for technical reasons in the proof of Lemma 4.5 (which is given in [26]). Note that if ${\tau_\epsilon}^{-1}$ is symmetric and strictly positive with a lower bound of order $\mathcal{O}(\epsilon^{1-\delta})$ on the spectrum, then Assumption 2.4 and Assumption 2.5 are satisfied. This implies that the slowest time scale of the ODE (2.1) is at worse of order $\epsilon^{1-\delta}$. Indeed, integration over time in Assumption 2.4 implies $\|\tau_\epsilon\|\leq C\epsilon^{\delta-1}$. Moreover, we assume that the solution of the ODE driving the internal state in (2.4) satisfies the following long time behavior :

Assumption 2.6. Consider a path $t\mapsto S_t$ such that $\sup_{t\in[0,+\infty]}\left|\frac{dS_t}{dt}\right|=\mathcal{O}(\epsilon)$, and assume $|Y_0-S_0|=\mathcal{O}(\epsilon^\delta)$. Then the solution of

$$\frac{dY_t}{dt}=F_\epsilon(Y_t,s_t),$$

satisfies

$$\sup_{t\in[0,+\infty]} |Y_t - S_t| = \mathcal{O}\left(\epsilon^{\delta}\right).$$

Assumption 2.6 is motivated by the case of linear ODEs; indeed, in that case, Assumption 2.6 is a consequence of Assumption 2.4 (see Section 3.3 in [26]). Recalling that $Z_t = S_t - Y_t$, we will assume in the remainder of the paper that the solution of (2.4) satisfies,

$$|Z_0| = \mathcal{O}(\epsilon^{\delta}),$$

as well as

$$\left|{\tau_\epsilon}^{-1} Z_0\right| = \mathcal{O}(1).$$

Then, under Assumption 2.2, Assumption 2.4, Assumption 2.5, and Assumption 2.6, we obtain the bounds (see Section 3.3 in [26])

$$\sup_{t\in[0,+\infty]} \left|{\tau_\epsilon}^{-1} \mathrm{e}^{-{\tau_\epsilon}^{-1} t}(S_t - Y_t)\right| = \mathcal{O}(1), \tag{2.10}$$

$$\sup_{t\in[0,+\infty]} (|Z_t|) = \mathcal{O}(\epsilon^{\delta}),$$

as well as

$$\sup_{t\in[0,+\infty]} \left(\left|{\tau_\epsilon}^{-1} \mathrm{e}^{-{\tau_\epsilon}^{-1} t} Z_t\right|\right) = \mathcal{O}(1).$$

Finally, we will make use of the following notation. Denoting $c_S = \left\|\nabla^2 S\right\|_\infty$, with ∇^2 the Hessian, the error terms due to non-linearities of the problem will be handled through

$$\begin{cases} \mathrm{nl}(\epsilon) := c_F \epsilon^{\delta} + c_S \epsilon + c_\lambda \epsilon^{k\delta - 1} = o(1), \\ \mathrm{nl}_2(\epsilon) := c_F(\epsilon^{\delta} + \epsilon^{\gamma}) + c_S(\epsilon + \epsilon^{\delta}) + c_\lambda \epsilon^{k\delta - 1} = o(1). \end{cases} \tag{2.11}$$

Note that $\mathrm{nl}(\epsilon) \le \mathrm{nl}_2(\epsilon)$ and $\mathrm{nl}(\epsilon) \sim \mathrm{nl}_2(\epsilon)$ if $\delta \le 1$ and $\delta \le \gamma$. Recall that $k \ge 2$ is defined in (2.3).

Note that all the assumptions of the present section hold in the following case: (i) the ODE is linear (F_ϵ is linear), (ii) the associated matrix τ_ϵ is symmetric positive and satisfies

$$\tau_\epsilon^{-1} \ge C\epsilon^{1-\delta},$$

for some $\delta > 0$, (iii) the initial internal state is close to equilibrium in the sense that ${\tau_\epsilon}^{-1} Z_0 = \mathcal{O}(1)$. The assumptions of the present section can be seen as technical generalizations to non-linear ODEs with non-symmetric linear part.

2.3. *Model with direct gradient sensing (control process)*

We now turn to a simplified, "coarse" model, in which the internal process (2.1), and the corresponding state variables, are eliminated. Instead, the turning rate depends directly on the chemoattractant gradient. This process will be called the *control process* since it will be used in Section 3 as a control variate to perform variance reduced simulations of (2.4).

The control process is a Markov process in position-velocity variables

$$t \mapsto (X_t^c, V_t^c),$$

which evolves according to the following differential velocity-jump equations :

$$\begin{cases} \dfrac{\mathrm{d}X_t^c}{\mathrm{d}t} = \epsilon V_t^c, \\ \displaystyle\int_{T_n^c}^{T_{n+1}^c} \lambda_\epsilon^c(X_t^c, V_t^c)\mathrm{d}t = \theta_{n+1}, \\ V_t^c = \mathcal{V}_n \quad \text{for } t \in [T_n^c, T_{n+1}^c] \ , \end{cases} \tag{2.12}$$

with initial condition $X_0, \mathcal{V}_0 \in \mathbb{R}^d$. In (2.12), $(\theta_n)_{n\geq 1}$ denote i.i.d. random variables with normalized exponential distribution, and $(\mathcal{V}_n)_{n\geq 1}$ denote i.i.d. random variables with distribution $\mathcal{M}(dv)$. The turning rate of the control process is assumed to satisfy

$$0 < \lambda_{\min} \leq \lambda_\epsilon^c(x,v) \leq \lambda_{\max}, \tag{2.13}$$

as well as

$$\lambda_\epsilon^c(x,v) := \lambda_0 - \epsilon\, A_\epsilon^T(x)v + \mathcal{O}(\epsilon^2). \tag{2.14}$$

Typically, $\lambda_\epsilon^c(x,v)$ is a function of $\nabla S(x)$, so that the model (2.14) may describe a large bacterium that is able to directly sense chemoattractant gradients. When $m = n = 1$, and the turning rate (2.14) is proportional to $\nabla S(x)v \in \mathbb{R}$, it can be interpreted as follows: the rate at which a bacterium will change its velocity direction depends on the alignment of the velocity with the gradient of the chemoattractant concentration $\nabla S(x)$, resulting in a transport towards areas with higher chemoattractant concentrations. In the companion paper [26], it is shown that, when the vector field $A_\epsilon(x) \in \mathbb{R}^d$ is given by the formula

$$A_\epsilon(x) = b^T \frac{\tau_\epsilon}{\mathrm{Id} + \lambda_0 \tau_\epsilon} \nabla S(x), \tag{2.15}$$

asymptotic consistency with the internal state model is obtained in the limit when $\epsilon \to 0$ (see Section 2.4).

2.4. *Kinetic formulation and advection-diffusion limits*

We now turn to the kinetic description of the probability distribution of the processes introduced above. The probability distribution density of the fine-scale process with internal state at time t with respect to the measure $\mathrm{d}x\mathcal{M}(\mathrm{d}v)\mathrm{d}y$ is denoted as $p(x,v,y,t)$, suppressing the dependence on ϵ for notational convenience,

and evolves according to the Kolomogorov forward evolution equation (or master equation). In the present context, the latter is the following kinetic equation

$$\partial_t p + \epsilon v \cdot \nabla_x p + \mathrm{div}_y (F_\epsilon(x,y)p) = \lambda(S(x)-y)(R(p)-p), \tag{2.16}$$

where

$$R(p) := \int_{v \in \mathbb{S}^{d-1}} p(\cdot, v, \cdot)\, \mathcal{M}(dv)$$

is the operator integrating velocities with respect to $\mathcal{M}$, and $F(x,y)$ and $\lambda(z)$ are defined as in Section 2.1. Similarly, the distribution density of the control process is denoted as $p^c(x,v,t)$, and evolves according to the kinetic/master equation:

$$\partial_t p^c + \epsilon v \cdot \nabla_x p^c = (R(\lambda_\epsilon^c p) - \lambda_\epsilon^c p), \tag{2.17}$$

with λ_ϵ^c defined as in (2.14). The reader is referred to [13] for the derivation of master equations associated to Markov jump processes.

In the companion paper [26], we show, using probabilistic arguments, that, in the limit $\epsilon \to 0$, both the equation for the control process (2.17) and the equation for the process with internal state (2.16) converge to an advection-diffusion limit on diffusive time scales. Convergence is to be understood *pathwise*, in the sense of convergence of probability distribution on paths endowed with the uniform topology.

Denote diffusive times by

$$\bar{t} := t\epsilon^2,$$

and the processes considered on diffusive time scales as

$$X_{\bar{t}}^\epsilon := X_{\bar{t}/\epsilon^2}, \qquad X_{\bar{t}}^{c,\epsilon} := X_{\bar{t}/\epsilon^2}^c.$$

For the control process, we then have the following theorem :

Theorem 2.1. *For $\epsilon \to 0$, the process $\bar{t} \mapsto X_{\bar{t}}^{c,\epsilon}$, solution of* (2.12)*, converges towards an advection-diffusion process, satisfying the stochastic differential equation (SDE)*

$$\mathrm{d}X_{\bar{t}}^{c,0} = \left(\frac{DA_0(X_{\bar{t}}^{c,0})}{\lambda_0} \mathrm{d}\bar{t} + \left(\frac{2D}{\lambda_0}\right)^{1/2} \mathrm{d}W_{\bar{t}} \right), \tag{2.18}$$

where $\bar{t} \mapsto W_{\bar{t}}$ is a standard Brownian motion, the parameters λ_0 and $A_0 := \lim_{\epsilon \to 0} A_\epsilon$ originate from the turning rate (2.14)*, and the diffusion matrix is given by the covariance of the Maxwellian distribution :*

$$D = \int_{\mathbb{S}^{d-1}} v \otimes v\, \mathcal{M}(dv) \in \mathbb{R}^{d \times d}. \tag{2.19}$$

In particular, this result implies that, at the level of the Kolomogorov/master evolution equation, the evolution of the position bacterial density,

$$n^{c,\epsilon}(x,\bar{t}) := n^c(x,\bar{t}/\epsilon^2) := \int_{\mathbb{V}} p^c(x,v,\bar{t}/\epsilon^2)\, \mathcal{M}(\mathrm{d}v) \tag{2.20}$$

converges to the advection-diffusion equation

$$\partial_{\bar{t}} n^{c,0} = \frac{1}{\lambda_0} \mathrm{div}_x \left(D \nabla_x n^{c,0} - D A_0(x) n^{c,0} \right) \tag{2.21}$$

on diffusive time scales as $\epsilon \to 0$.

In the same way, a standard probabilistic diffusion approximation argument can be used to derive the pathwise diffusive limit of the process with internal state (2.4) :

Theorem 2.2. *For $\epsilon \to 0$, the process $\bar{t} \mapsto X^{\epsilon}_{\bar{t}}$, solution of* (2.4)*, converges towards an advection-diffusion process, satisfying the stochastic differential equation (SDE)* (2.18)*, where A_0 originates from*

$$A_0(x) = b^T \lim_{\epsilon \to 0} \frac{\tau_\epsilon}{\lambda_0 \tau_\epsilon + \mathrm{Id}} \nabla S(x), \tag{2.22}$$

in which, b, τ_ϵ, and λ_0 were introduced in (2.7)-(2.3) *as parameters of the process with internal state, and $\mathrm{Id} \in \mathbb{R}^{n \times n}$ is the identity matrix. Again, the diffusion matrix D is given by the covariance of the Maxwellian distribution* (2.19)*.*

Also here, convergence needs to be understood in terms of probability distribution on paths endowed with the uniform topology.

Introducing the bacterial density of the process with internal state as

$$n(x,t) = \int_{\mathbb{Y}} \int_{\mathbb{V}} p(x,v,y,t) \mathcal{M}(\mathrm{d}v) \mathrm{d}y, \tag{2.23}$$

this implies that the evolution of n converges to (2.21) on diffusive time scales in the limit of $\epsilon \to 0$.

3. Numerical method

To simulate the process with internal state, solving the kinetic equation (2.16) over diffusive time scales can be cumbersome, due to the additional dimensions associated with the internal state. The alternative is to use to stochastic particles. However, a particle-based simulation of equation (2.16) is subject to a large statistical variance of the order $\mathcal{O}(N^{-1/2})$, where N is the number of simulated particles. The asymptotic analysis shows that the position bacterial density approaches an advection-diffusion limit (2.21) when $\epsilon \to 0$. Consequently, to accurately assess the deviations of the process with internal state (2.4) as compared to its advection-diffusion limit (for small but non-vanishing (intermediate) values of ϵ), the required number of particles needs to increase substantially with decreasing ϵ, which may become prohibitive from a computational point of view.

In this section, we therefore propose a hybrid method, based on the principle of control variates, that couples the process with internal dynamics with the control process, which is simulated simultaneously using a grid-based method. We first describe the variance reduction technique (Section 3.1). The analysis in Section 4 will reveal that the variance reduction is *asymptotic*, in the sense that the variance vanishes in the diffusion limit. To ensure this asymptotic variance reduction during

actual simulations, one needs to ensure that the time discretization preserves the diffusion limits of the time-continuous process. An appropriate time discretization is discussed in Section 3.2.

3.1. *Coupling and asymptotic variance reduction*

The proposed variance reduction technique is based on the introduction of a control variate that exploits a coupling between the process with internal state and the control process. While the idea of control variates for Monte Carlo simulation is already well known, see e.g. [21] and references therein, the coupling that is proposed here is particular (we call it *asymptotic*), since the difference of the coupled processes, and hence the variance, vanishes with an estimable rate in the diffusion limit $\epsilon \to 0$, as will be shown in Section 4.

3.1.1. *The control variates*

Let us first assume that we are able to compute the exact solution of the kinetic equation for the control process, (2.17), with infinite precision in space and time.

The algorithm of asymptotic variance reduction is based on a coupling between an ensemble of realizations evolving according to the process with internal state (2.4), denoted as

$$\left\{X_t^i, V_t^i, Y_t^i\right\}_{i=1}^N,$$

and an ensemble of realization of the control process (2.12), denoted as

$$\left\{X_t^{i,c}, V_t^{i,c}\right\}_{i=1}^N.$$

We denote the empirical measure of the particles with internal state in position-velocity space as

$$\mu_{\bar{t}}^N(x,v) = \frac{1}{N}\sum_{i=1}^N \delta_{X^i_{\bar{t}/\epsilon^2}, V^i_{\bar{t}/\epsilon^2}},$$

and, correspondingly, the empirical measure of the control particles as

$$\mu_{\bar{t}}^{c,N} = \frac{1}{N}\sum_{i=1}^N \delta_{X^{c,i}_{\bar{t}/\epsilon^2}, V^{c,i}_{\bar{t}/\epsilon^2}}.$$

A coupling between the two ensembles is obtained by ensuring that both simulations use *the same random numbers* $(\theta_n)_{n\geq 1}$ and $(\mathcal{V}_n)_{n\geq 0}$, which results in a strong correlation between (X_t^i, V_t^i) and $(X_t^{i,c}, V_t^{c,i})$ for each realization. Simultaneously, the kinetic equation for the control process (2.17) is also solved using a deterministic method (which, for now, is assumed to be exact). We formally denote the corresponding semi-group evolution as

$$\mathrm{e}^{\bar{t}L^c}, \qquad \text{with } L^c(p^c) = -\epsilon v \cdot \nabla_x p^c + (R(\lambda_\epsilon^c p^c) - \lambda_\epsilon^c p^c).$$

Besides the two particle measures $\mu_{\bar{t}}^N$ and $\mu_{\bar{t}}^{c,N}$, we denote by $\overline{\mu}_{\bar{t}}^N$ the variance reduced measure, which will be defined by the algorithm below. Since, with increasing diffusive time, the variance of the algorithm increases due to a loss of coupling between the particles with internal state and the control particles, the variance reduced algorithm will also make use of a reinitialization time step $\overline{\delta t}_{ri}$, which is defined on the diffusive time scale. The corresponding time instances are denoted as $\bar{t}_n = n\overline{\delta t}_{ri}$ on the diffusive time scale, or equivalently, on the original time scale as $t_n = n\overline{\delta t}_{ri}/\epsilon^2$.

Starting from an initial probability measure μ_0 at time $t=0$, we sample μ_0 to obtain the ensemble $\left\{X_t^i, V_t^i, Y_t^i\right\}_{i=1}^N$, corresponding to μ_0^N, and then set $\mu_0^{c,N} := \mu_0^N$, i.e., $X_0^{i,c} = X_0^i$ and $V_0^{i,c} = V_0^i$ for all $i = 1,\ldots,N$. Furthermore, we set the variance reduced estimator as $\overline{\mu}_0^N := \mu_0 = \mathbb{E}(\mu_0^N)$. We then use the following algorithm to advance from $\bar{t}_n$ to $\bar{t}_{n+1}$, (see also Figure (1)) :

Algorithm 3.1. At time t_n, we have that the particle measure $\mu_{\bar{t}_n}^{c,N} = \mu_{\bar{t}_n}^N$, and the variance reduced measure is given by $\overline{\mu}_{\bar{t}_n}^N$. To advance from time $\bar{t}_n$ to $\bar{t}_{n+1}$, we perform the following steps :

- Evolve the particles $\left\{X_t^i, V_t^i, Y_t^i\right\}_{i=1}^N$ from t_n to t_{n+1}, according to (2.4),
- Evolve the particles $\left\{X_t^{i,c}, V_t^{i,c}\right\}_{i=1}^N$ according to (2.12), using the same random numbers as for the process with internal state,
- Compute the variance reduced evolution

$$\overline{\mu}_{\bar{t}_{n+1}}^N = \overline{\mu}_{\bar{t}_n}^N \mathrm{e}^{\overline{\delta t}_{ri}/\epsilon^2 L^c} + \left(\mu_{\bar{t}_{n+1}}^N - \mu_{\bar{t}_{n+1}^-}^{c,N}\right). \tag{3.1}$$

- Reinitialize the control particles by setting

$$X_{t_{n+1}}^{i,c} = X_{t_{n+1}}^i, \qquad V_{t_{n+1}}^{i,c} = V_{t_{n+1}}^i, \qquad i = 1,\ldots,N,$$

  i.e., we set the state of the control particles to be identical to the state of the particles with internal state.

In formula (3.1), we use the symbol $\bar{t}_{n+1}^-$ to emphasize that the involved particle positions and velocities are those obtained *before* the reinitialization. An easy computation shows that the algorithm is unbiased in the sense that for any $n \geq 0$,

$$\mathbb{E}\left(\overline{\mu}_{\bar{t}_n}^N\right) = \mathbb{E}\left(\mu_{\bar{t}_n}^N\right),$$

since the particles with internal dynamics are unaffected by the reinitialization, and, additionally,

$$\mathbb{E}\mu_{\bar{t}_{n+1}}^{c,N} = \mathbb{E}\overline{\mu}_{\bar{t}_n}^N \mathrm{e}^{\overline{\delta t}_{ri}/\epsilon^2 L^c}.$$

Moreover, the variance is controlled by the coupling between the two processes. Indeed, using the independence of the random numbers between two steps of Algorithm 3.1, and introducing φ as a position and velocity dependent test function),

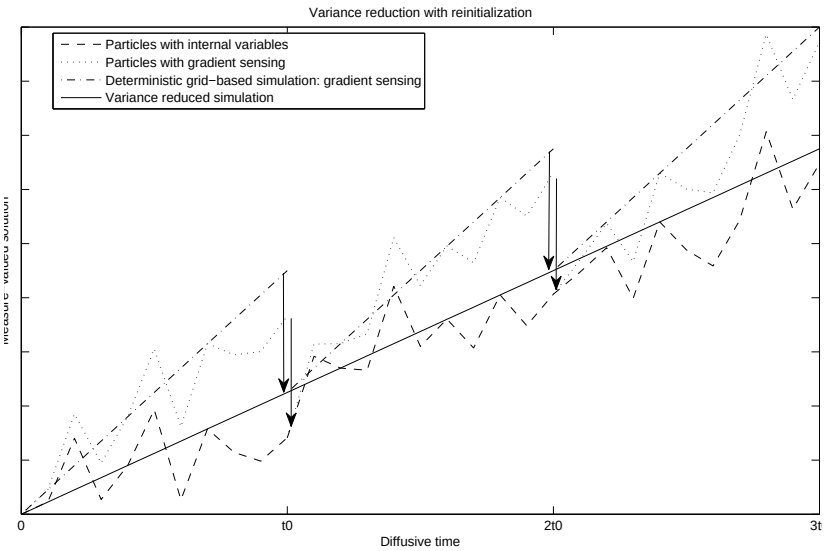

Fig. 1. A schematic description of Algorithm 3.1. The dashed line represent the evolution of N bacteria with internal state. The dotted line represent the coupled evolution of N bacteria with gradient sensing, subject to regular reinitializations. The dashed-dotted line is computed according to a deterministic method simulating the density of the model with gradient sensing, and subject to regular reinitializations. The solid line is the variance reduced simulation of the internal state dynamics, and is computed by adding the difference between the particle computation with internal state, and the particle simulation with gradient sensing to the deterministic gradient sensing simulation. At each reinitialization step, the two simulations (deterministic and particles) of the gradient sensing dynamics are reinitialized to the values of their internal state simulation counterpart (as represented by the arrows).

we get (according to the main result of Theorem 4.7)

$$\begin{aligned}\mathrm{var}(\overline{\mu}^N_{\bar{t}_n}(\varphi)) &= \sum_{k=1}^{n} \mathbb{E}\left(\mu^N(\varphi) - \mu^{c,N}_{\bar{t}_k^-}(\varphi)\right)^2 \\ &\leq \sum_{k=1}^{n} \frac{\|\nabla\varphi\|_\infty}{N} \mathbb{E}\left(\left|X_{\bar{t}_k/\epsilon^2} - X^c_{\bar{t}_k/\epsilon^2}\right|^2\right),\end{aligned}$$

and thus

$$\mathrm{var}(\overline{\mu}^N_{\bar{t}_n}(\varphi)) \leq Cn\frac{\epsilon+\epsilon^\delta+\mathrm{nl}_2(\epsilon)}{N}, \tag{3.2}$$

where in the last line, C is independant of n, ϵ, and N.

In some generic situations (see Section 4.3), we can argue that the statistical error in Algorithm 3.1 coming from the coupling is "sharp" with respect to the order in ϵ. This means that the difference between the probability distribution of the model with internal state and the probability distribution of the model with gradient sensing is of the same order. This would imply that, with the asymptotic variance reduction technique, one is able to reliably assess the true deviation of the process with internal variables from the control process using a number of particles N that is independent of ϵ.

3.1.2. *Time-space discretization of the kinetic equation*

For both processes, the probability density distribution of the position and velocity can then be computed via binning in a histogram, or via standard kernel density estimation [27,28] using a kernel K_h, in which the chosen bandwidth h can be based on the grid used for the deterministic simulation and on the data, for instance using the Silverman heuristic [28]. This yields

$$\hat{p}_N(x,v,t) = \frac{1}{Nh}\sum_{i=1}^N K_h(x-X^i_t, v-V^i_t), \tag{3.3}$$

and similarly

$$\hat{p}^c_N(x,v,t) = \frac{1}{Nh}\sum_{i=1}^N K_h(x-X^{c,i}_t, v-V^{c,i}_t). \tag{3.4}$$

The important point is that the solution $p^c(x,v,t)$ for the control process may be accurately approximated with a deterministic grid-based method, which ensures that

$$\left|p^c(x,v,\bar{t}/\epsilon^2) - p^c_{\mathrm{grid}}(x,v,\bar{t}/\epsilon^2)\right| = \mathcal{O}(\delta x^l) + \mathcal{O}(\delta t^k),$$

for some integers $k,l \geq 1$.

Remark 3.1. In this text, we will perform the simulations in one space dimension using a third-order upwind-biased scheme, and perform time integration using the standard fourth order Runge–Kutta method, i.e., $l=3$ and $k=4$.

Then, the unbiased nature of the variance-reduced estimator is conserved up to $\mathcal{O}(\delta x^l)+\mathcal{O}(\delta t^k)+\mathcal{O}(h)$ discretization errors.

3.2. *Time discretization of velocity-jump processes*

When simulating equation (2.4), a time discretization error originates from the fact that the equation for the evolution of the internal state, and hence the evolution of the fluctuations Z_t, is discretized in time. This results in an approximation of the jump times $(T_n)_{n\geq 1}$, and hence of X_t. To retain the diffusion limits of the time-continuous process, special care is needed. We now briefly recall the time discretization procedure that was proposed and analyzed in the companion paper [26]. For ease of exposition, we consider the scalar equation (2.5) for the internal state; generalization to nonlinear systems of equations is briefly discussed in [26].

Consider first the linear turning rate (2.3). In that case, we define a numerical solution $(X_t^{\delta t}, V_t^{\delta t}, Y_t^{\delta t})$ as follows. Between jumps, we discretize the simulation in steps of size δt and denote by $(X_{n,k}^{\delta t}, Z_{n,k}^{\delta t})$ the solution at $t_{n,k} = T_n^{\delta t} + k\delta t$. The numerical solution for $t \in [t_{n,k}, t_{n,k+1}]$ is given by

$$\begin{cases} X_t^{\delta t} = X_{n,k}^{\delta t} + \epsilon \mathcal{V}_n \left(t - t_{n,k}\right) \\ Z_t^{\delta t} = \exp(-(t-t_{n,k})\tau_\epsilon{}^{-1}) Z_{n,k}^{\delta t} + \epsilon\tau_\epsilon \left(\mathrm{Id} - \exp(-(t-t_{n,k})\tau_\epsilon{}^{-1})\right) \nabla S(X_{n,k}^{\delta t}) \mathcal{V}_n. \end{cases} \tag{3.5}$$

We denote by $K \geq 0$ the integer such that the simulated jump time $T_{n+1}^{\delta t} \in [t_{n,K}, t_{n,K+1}]$. To find $T_{n+1}^{\delta t}$, we first approximate the integral $\int_{T_n^{\delta t}}^{T_{n+1}^{\delta t}} \lambda(Z_t^{\delta t}) \mathrm{d}t$ using

$$\int_{T_n^{\delta t}}^{T_{n+1}^{\delta t}} \lambda(Z_t^{\delta t}) \mathrm{d}t = \sum_{k=0}^{K-1} \int_{t_{n,k}}^{t_{n,k+1}} \lambda(Z_t^{\delta t}) \mathrm{d}t + \int_{t_{n,K}}^{T_{n+1}^{\delta t}} \lambda(Z_t^{\delta t}) \mathrm{d}t, \tag{3.6}$$

and then compute :

$$\begin{aligned} \int_{t_{n,k}}^{t_{n,k+1}} \lambda(Z_t^{\delta t}) \mathrm{d}t = \lambda_0 \delta t - b^T \left(\mathrm{Id} - \mathrm{e}^{-\delta t \tau_\epsilon{}^{-1}}\right) \tau_\epsilon Z_{T_n}^{\delta t} \\ - \epsilon b^T \left(\delta t \tau_\epsilon - (\mathrm{Id} - \mathrm{e}^{-\delta t \tau_\epsilon{}^{-1}}) \tau_\epsilon{}^2\right) \nabla S(X_{n,k}^{\delta t}) \mathcal{V}_n. \end{aligned} \tag{3.7}$$

The jump time $T_{n+1}^{\delta t}$ can then be computed as the solution of

$$\int_{t_{n,K}}^{T_{n+1}^{\delta t}} \lambda(Z_t^{\delta t}) \mathrm{d}t = \theta_{n+1} - \sum_{k=0}^{K-1} \int_{t_{n,k}}^{t_{n,k+1}} \lambda(Z_t^{\delta t}) \mathrm{d}t, \tag{3.8}$$

using a Newton procedure. It is shown in [26] that the results on the diffusion limit (as outlined in Section 2.4) are not affected by the discretization.

We now consider a general nonlinear turning rate. We again discretize in time to obtain the time-discrete solution (3.5). The jump time $T_{n+1}^{\delta t}$ is now computed by linearizing (2.6) in each time step,

$$\lambda^{\delta t}\left(Z_t^{\delta t}, Z_{n,k}^{\delta t}\right) = \lambda(Z_{n,k}^{\delta t}) + \frac{\mathrm{d}\lambda(Z_{n,k}^{\delta t})}{\mathrm{d}z} \left(Z_t^{\delta t} - Z_{n,k}^{\delta t}\right). \tag{3.9}$$

One then replaces equation (3.6) by

$$\int_{T_n^{\delta t}}^{T_{n+1}^{\delta t}} \lambda^{\delta t}(Z_t^{\delta t})\mathrm{d}t = \sum_{k=0}^{K-1} \int_{t_{n,k}}^{t_{n,k+1}} \lambda^{\delta t}(Z_t^{\delta t}, Z_{n,k}^{\delta t})\mathrm{d}t + \int_{t_{n,K}}^{T_{n+1}^{\delta t}} \lambda^{\delta t}(Z_t^{\delta t}, Z_{n,K}^{\delta t})\mathrm{d}t, \quad (3.10)$$

and proceeds in the same way as for the linear case. Also in this case, the diffusive limit is recovered in an exact fashion for the time-discretized process.

Remark 3.2 (Discretization of the control process). For the control process, there is no internal state. Hence, the only time dependence of turning rate λ^c is due to the spatial variation of $\nabla S(x)$, which can be treated by discretizing the integral of λ in the same way as above.

4. Asymptotic variance reduction of the coupling

In this section, we show how the difference between the two coupled processes on diffusive time scales $\bar{t} = t/\epsilon^2$ behaves in the limit of $\epsilon \to 0$. We first recall some notation from the companion paper [26] (Section 4.1), after which we state and prove the main theorem (Section 4.2).

4.1. *Notations and asymptotic estimates*

The main theorem in Section 4.2 relies on following auxiliary definitions and lemmas that were given in the companion paper [26]:

Definition 4.1. We denote by $m : \mathbb{R} \to \mathbb{R}^{n\times n}$ the function

$$m(t) := t\tau_\epsilon - \left(\mathrm{Id} - \mathrm{e}^{-t{\tau_\epsilon}^{-1}}\right){\tau_\epsilon}^2, \quad (4.1)$$

whose derivative is given by

$$m'(t) = \tau_\epsilon \left(\mathrm{Id} - \mathrm{e}^{-t{\tau_\epsilon}^{-1}}\right).$$

This function satisfies the following lemmas :

Lemma 4.2. *Let θ be an exponential random variable of mean 1. Then:*

$$\mathbb{E}\left(m\left(\frac{\theta}{\lambda_0}\right)\right) = \frac{1}{\lambda_0}\frac{\tau_\epsilon}{\mathrm{Id} + \lambda_0\tau_\epsilon} = A_\epsilon(x), \quad (4.2)$$

see equation (2.15).

Lemma 4.3. *For all $t \in \mathbb{R}$, we have $\|m'(t)\| \le t$, as well as $\|m(t)\| \le t^2/2$.*

The difference between jump times is denoted as

$$\Delta T_{n+1}^c := T_{n+1}^c - T_n^c, \qquad \Delta T_{n+1} := T_{n+1} - T_n.$$

The proofs of the asymptotic variance reduction make use of the following asymptotic expansions of the jump time differences of both processes :

Lemma 4.4. *The difference between two jump times of the control process satisfies*

$$\Delta T^c_{n+1} = \frac{\theta_{n+1}}{\lambda_0} + \epsilon \frac{\theta_{n+1}}{\lambda_0^2} A^T_\epsilon(X^c_{T^c_n})\mathcal{V}_n + \theta_{n+1}\mathcal{O}(\epsilon^2). \tag{4.3}$$

and

Lemma 4.5. *The jump time variations of the process with internal state can be written in the following form* ($\mathrm{nl}(\epsilon)$ *being defined by* (2.11)) *:*

$$\Delta T_{n+1} = \Delta T^0_{n+1} + \epsilon \Delta T^1_{n+1} + (\theta^6_{n+1} + \theta_{n+1})\mathcal{O}\left(\epsilon^2 + \epsilon \mathrm{nl}(\epsilon)\right), \tag{4.4}$$

where

$$\begin{aligned}\Delta T^0_{n+1} &= \frac{\theta_{n+1}}{\lambda_0} + \frac{b^T}{\lambda_0} m'(\Delta T^0_{n+1}) Z_{T_n} \\ &= \frac{\theta_{n+1}}{\lambda_0} + \theta_{n+1}\mathcal{O}(\epsilon^\delta)\end{aligned} \tag{4.5}$$

and, correspondingly,

$$\begin{aligned}\Delta T^1_{n+1} &= \frac{1}{\lambda_0 - b^T \mathrm{e}^{-\Delta T^0_{n+1}\tau^{-1}} Z_{T_n}} b^T m(\Delta T^0_{n+1}) \nabla S(X_{T_n})\mathcal{V}_n \\ &= \frac{1}{\lambda_0} b^T m\left(\frac{\theta_{n+1}}{\lambda_0}\right) \nabla S(X_{T_n})\mathcal{V}_n + \theta_{n+1}\mathcal{O}(\epsilon^\delta).\end{aligned} \tag{4.6}$$

It is also useful to recall that according to (2.2)-(2.13), the following hold :

$$|\Delta T_{n+1}| \le C\theta_{n+1}, \qquad \left|\Delta T^c_{n+1}\right| \le C\theta_{n+1}.$$

Finally, we will also need a different expansion of the jump times of the process with internal state :

Lemma 4.6. *When using* $\mathrm{nl}_2(\epsilon)$ *as defined by* (2.11)*, the jump time variations of the process with internal state can be written in the following form :*

$$\begin{aligned}\Delta T_{n+1} &= \frac{\theta_{n+1}}{\lambda_0} - \frac{b^T \tau_\epsilon}{\lambda_0}(Z_{T_{n+1}} - Z_{T_n}) + \epsilon \Delta T_0 b^T \tau_\epsilon \nabla S(X_{T_n})\mathcal{V}_n \\ &+ \left(\theta^3_{n+1} + \theta_{n+1}\right)\mathcal{O}\left(\epsilon \mathrm{nl}_2(\epsilon) + \epsilon^{1+\delta}\right).\end{aligned} \tag{4.7}$$

Proof. The proof is based on Duhamel integration of the ODE (2.4) on $[T_n, T_{n+1}]$, as in the proof of Lemma 4.7 of the companion paper [26]. Following equation (4.24) in [26], we get for $t \in [T_n, T_{n+1}]$:

$$Z_t = \mathrm{e}^{-(t-T_n)\tau_\epsilon^{-1}} Z_{T_n} + \epsilon\left(\mathrm{Id} - \mathrm{e}^{-(t-T_n)\tau_\epsilon^{-1}}\right)\tau_\epsilon \nabla S(X_{T_n})\mathcal{V}_n + (\theta^2_{n+1} + \theta_{n+1})\mathcal{O}\left(c_S\epsilon^2 + c_F\epsilon^{1+\delta}\right). \tag{4.8}$$

Multiplying both sides by τ_ϵ, recalling that Assumption 2.4 implies $\|\tau_\epsilon\| = \mathcal{O}(\epsilon^{\delta-1})$, and using Assumption 2.3, we get for $t \in [T_n, T_{n+1}]$:

$$\tau_\epsilon Z_t = \tau_\epsilon \mathrm{e}^{-(t-T_n)\tau_\epsilon^{-1}} Z_{T_n} + \epsilon\left(\mathrm{Id} - \mathrm{e}^{-(t-T_n)\tau_\epsilon^{-1}}\right){\tau_\epsilon}^2 \nabla S(X_{T_n})\mathcal{V}_n + (\theta^2_{n+1} + \theta_{n+1})\mathcal{O}\left(c_S\epsilon^{1+\delta} + c_F\epsilon^{1+\gamma}\right). \tag{4.9}$$

Moreover, integrating (4.8) on $t\in[T_n,T_{n+1}]$ yields

$$\int_{T_n}^{T_{n+1}} Z_t\mathrm{d}t=\left(\mathrm{Id}-\mathrm{e}^{-(T_{n+1}-T_n)\tau_\epsilon{}^{-1}}\right)\tau_\epsilon(Z_{T_n}-\epsilon\tau_\epsilon\nabla S(X_{T_n})\mathcal{V}_n)$$
$$+\epsilon(T_{n+1}-T_n)\tau_\epsilon\nabla S(X_{T_n})\mathcal{V}_n+(\theta_{n+1}^3+\theta_{n+1}^2)\mathcal{O}\left(c_S\epsilon^2+c_F\epsilon^{1+\delta}\right). \tag{4.10}$$

Evaluating equation (4.9) at $t=T_{n+1}$, add adding equation (4.10), we get

$$\int_{T_n}^{T_{n+1}} Z_t\mathrm{d}t+\tau_\epsilon Z_{T_{n+1}}=\tau_\epsilon Z_{T_n}+\epsilon(T_{n+1}-T_n)\tau_\epsilon\nabla S(X_{T_n})\mathcal{V}_n$$
$$+\left(\theta_{n+1}^3+\theta_{n+1}\right)\mathcal{O}\left(c_S(\epsilon^{1+\delta}+\epsilon^2)+c_F(\epsilon^{1+\delta}+\epsilon^{1+\gamma})\right).$$

The estimates in Lemma 4.8 from [26] (i.e., $\left|\Delta T_{n+1}-\Delta T_{n+1}^0\right|=(\theta_{n+1}^4+\theta_{n+1})\mathcal{O}(\epsilon)$) yields

$$\int_{T_n}^{T_{n+1}} Z_t\mathrm{d}t+\tau_\epsilon Z_{T_{n+1}}=\tau_\epsilon Z_{T_n}+\epsilon\Delta T_{n+1}^0\tau_\epsilon\nabla S(X_{T_n})\mathcal{V}_n$$
$$+\left(\theta_{n+1}^3+\theta_{n+1}\right)\mathcal{O}\left(c_S(\epsilon^{1+\delta}+\epsilon^2)+(c_F+1)\epsilon^{1+\delta}+c_F\epsilon^{1+\gamma})\right).$$

Finally, plugging the last equation in the estimate (2.3) yields the result. □

4.2. *Analysis of variance of the coupling*

The present section is devoted to the proof of the following theorem:

Theorem 4.7. *Assume the assumptions of Section 2.2 hold, and that $k\delta>1$ where $k\geq 2$ is defined in* (2.3)*. Then the difference between the process with internal state and the coupling process satisfies :*

$$\mathbb{E}\left(\left(X_{\bar{t}/\epsilon^2}-X^c_{\bar{t}/\epsilon^2}\right)^p\right)^{1/p}=\mathcal{O}(\epsilon+\epsilon^\delta+\mathrm{nl}_2(\epsilon)),\qquad p\geq 1, \tag{4.11}$$

where $\mathrm{nl}_2(\epsilon)$ *is defined in* (2.11)*.*

The proof of Theorem 4.7 relies on a number of steps that will be detailed by a series of lemmas. We will make use of the following random sequences, defined as the position of both processes after n jumps,

$$\Xi_n:=X_{T_n}=X_0+\epsilon\sum_{m=0}^{n-1}\Delta T_{m+1}\mathcal{V}_m,\qquad n\geq 0, \tag{4.12}$$

as well as

$$\Xi_n^c:=X^c_{T_n^c}=X_0+\epsilon\sum_{m=0}^{n-1}\Delta T^c_{m+1}\mathcal{V}_m,\qquad n\geq 0. \tag{4.13}$$

We will also make use of the following random integers :

Definition 4.8. The random integers $N\geq 1$ and $N^c\geq 1$ are uniquely defined by:

$$T_N\leq\bar{t}/\epsilon^2<T_{N+1},\qquad T^c_{N^c}\leq\bar{t}/\epsilon^2<T^c_{N^c+1}. \tag{4.14}$$

The first lemma bounds the difference at time $\bar{t}/\epsilon^2$ between the two processes $X_{\bar{t}/\epsilon^2}$ and $X^c_{\bar{t}/\epsilon^2}$ by expressing it in terms of differences of positions and jump times of both processes after the same random number of jumps, Ξ_n and Ξ^c_n :

Lemma 4.9. *The difference between the rescaled process with internal state $X^\epsilon_{\bar{t}} := X_{\bar{t}/\epsilon^2}$ and the rescaled coupling process $X^{c,\epsilon}_{\bar{t}} := X^c_{\bar{t}/\epsilon^2}$ satisfies:*

$$\left|X^\epsilon_{\bar{t}} - X^{\epsilon,c}_{\bar{t}}\right| \leq |\Xi_N - \Xi^c_N| + \epsilon|T_N - T^c_N| + (\theta_{N+1} + \theta_{N_c+1})\mathcal{O}(\epsilon) \tag{4.15}$$

$$\leq |\Xi_{N^c} - \Xi^c_{N^c}| + \epsilon|T_{N^c} - T^c_{N^c}| + (\theta_{N+1} + \theta_{N_c+1})\mathcal{O}(\epsilon) \tag{4.16}$$

Only one of the two estimates (4.15)-(4.16) is necessary in the remainder of the proof, but we detail both to highlight the symmetry.

Proof. By definition, we have

$$\begin{cases} X^\epsilon_{\bar{t}} = \Xi_N + \epsilon(\bar{t}/\epsilon^2 - T_N)\mathcal{V}_N, \\ X^{c,\epsilon}_{\bar{t}} = \Xi^c_{N^c} + \epsilon(\bar{t}/\epsilon^2 - T^c_{N^c})\mathcal{V}_N, \end{cases}$$

so that by Definition 4.8 of N and N^c, and by realizing that $\left|\bar{t}/\epsilon^2 - T^c_{N^c}\right| \leq C\theta_{N_c+1}$ and $\left|\bar{t}/\epsilon^2 - T_N\right| \leq C\theta_{N+1}$, we get

$$\begin{cases} \left|X^\epsilon_{\bar{t}} - X^{c,\epsilon}_{\bar{t}}\right| \leq |\Xi_N - \Xi^c_N| + \underbrace{|\Xi^c_{N^c} - \Xi^c_N|}_{(a)} + (\theta_{N+1} + \theta_{N_c+1})\mathcal{O}(\epsilon), \\ \left|X^\epsilon_{\bar{t}} - X^{c,\epsilon}_{\bar{t}}\right| \leq |\Xi_{N^c} - \Xi^c_{N^c}| + \underbrace{|\Xi_{N^c} - \Xi_N|}_{(b)} + (\theta_{N+1} + \theta_{N_c+1})\mathcal{O}(\epsilon). \end{cases} \tag{4.17}$$

To analyze (a), we consider three cases:

$N = N_c$. Then $(a) = 0$.

$N > N_c$. Then

$$\begin{aligned} \frac{(a)}{\epsilon} &= \left|\sum_{n=N_c}^{N-1} \Delta T^c_{n+1}\mathcal{V}_n\right| \\ &\leq T^c_N - T^c_{N_c} = \Delta T^c_{N^c+1} + T^c_N - T^c_{N^c+1} \\ &\leq T^c_N - T_N + C\theta_{N^c+1}, \end{aligned}$$

where in the last line we have used that, $\Delta T^c_{N^c+1} \leq C\theta_{N^c+1}$, and, by definition, $T_N \leq \bar{t}/\epsilon^2 < T^c_{N^c+1}$.

$N_c > N$. Then

$$\begin{aligned} \frac{(a)}{\epsilon} &= \left|\sum_{n=N}^{N_c-1} \Delta T^c_{n+1}\mathcal{V}_n\right| \\ &\leq T^c_{N_c} - T^c_N \\ &\leq T_{N+1} - T^c_N = \Delta T_{N+1} + T_N - T^c_N \leq T_N - T^c_N + C\theta_{N+1}, \end{aligned}$$

where in the last line we have used that by definition $T^c_{N^c} \leq \bar{t}/\epsilon^2 \leq T_{N+1}$.

This proves (4.15). By symmetry, we get:

$$\frac{(b)}{\epsilon} \leq |T_{N^c} - T^c_{N^c}| + C\left(\theta_{N+1} + \theta_{N^c+1}\right),$$

which yields (4.16). □

In the next lemma, we estimate the supremum of the difference of the position of both processes after n jumps for $n \in [0, n_\epsilon]$.

Lemma 4.10. *Let $n_\epsilon \geq 1$ be a deterministic integer verifying $n_\epsilon = \mathcal{O}(\epsilon^{-2})$. Then*

$$\mathbb{E}\left(\sup_{0\leq n\leq n_\epsilon} |\Xi_n - \Xi^c_n|^p\right)^{1/p} = \mathcal{O}(\epsilon + \epsilon^\delta + \mathrm{nl}(\epsilon)). \tag{4.18}$$

Proof. First, using the estimates of jump times (4.3)-(4.4), we can decompose the differences of positions of both processes as follows :

$$\begin{aligned}\Xi_n - \Xi^c_n &= \sum_{m=0}^{n-1} \epsilon(\Delta T_{m+1} - \Delta T^c_{m+1})\mathcal{V}_m \\ &= \sum_{m=0}^{n-1} \alpha_{m+1} + d_{m+1} + r_{m+1},\end{aligned}$$

where, by definition,

$$\begin{cases} \alpha_{m+1} := \epsilon^2 \frac{\mathcal{V}_m}{\lambda_0^2}\left(A_\epsilon(\Xi_m)^T - A_\epsilon(\Xi^c_m)^T\right)\mathcal{V}_m, \\ d_{m+1} := \epsilon^2 \frac{\mathcal{V}_m}{\lambda_0^2}\left(1-\theta_{m+1}\right) A_\epsilon(\Xi^c_m)^T \mathcal{V}_m \\ \qquad\qquad + \epsilon^2 \frac{\mathcal{V}_m}{\lambda_0^2}\left(\lambda_0 b^T m\left(\frac{\theta_{m+1}}{\lambda_0}\right)\nabla S(\Xi_m) - A_\epsilon(\Xi_m)^T\right)\mathcal{V}_m \\ \qquad\qquad + \epsilon b^T m'(\Delta T^0_{m+1}) Z_{T_m}\mathcal{V}_m, \\ r_{m+1} := \left(\theta^6_{m+1} + \theta_{n+1}\right)\mathcal{O}(\epsilon^3 + \epsilon^2 \mathrm{nl}(\epsilon)). \end{cases}$$

Since A_ϵ is Lipschitz, we have that $\alpha_{m+1} \leq C\epsilon^2 |\Xi_m - \Xi^c_m|$, and we can write

$$|\Xi_n - \Xi^c_n| \leq C\epsilon^2 \sum_{m=0}^{n-1} |\Xi_m - \Xi^c_m| + \sup_{0\leq l\leq n_\epsilon}\left|\sum_{m=1}^{l} d_m\right| + \sum_{m=1}^{n_\epsilon} |r_m|,$$

for $n \leq n_\epsilon$, which yields

$$|\Xi_n - \Xi^c_n| \leq \left(\sup_{0\leq l\leq n_\epsilon}\left|\sum_{m=1}^{l} d_m\right| + \sum_{m=1}^{n_\epsilon} |r_m|\right)\left(1 + C\epsilon^2 + \cdots + (C\epsilon^2)^{n_\epsilon}\right),. \tag{4.19}$$

Note that

$$\left(1 + C\epsilon^2 + \cdots + (C\epsilon^2)^{n_\epsilon}\right) \leq C.$$

Now, we can remark that the discrete time process $(M_l)_{l\geq 0} := \left(\sum_{m=1}^{l} d_m\right)_{l\geq 0}$ is a martingale with respect to the filtration $\mathcal{F}_l = \sigma\left((\theta_m, \mathcal{V}_{m-1}) | 1 \leq m \leq l\right)$ with $M_0 = 0$,

so that we can apply the Burkholder-Davies-Gundy upper bound (which is a simple consequence of Doob maximal inequality here, see [1]), for some $p \geq 1$:

$$\mathbb{E}\left(\sup_{0\leq l\leq n_\epsilon} |M_l|^p\right) \leq C_p n_\epsilon^{p/2-1} \sum_{l=1}^{n_\epsilon} \mathbb{E}(|M_l - M_{l-1}|^p). \tag{4.20}$$

In the present case, since $|Z_{T_n}| = \mathcal{O}(\epsilon^\delta)$:

$$\mathbb{E}(|d_l|^p) \leq C_p \epsilon^{p(1+\delta)}.$$

Moreover,

$$\begin{aligned}\mathbb{E}\left(\left|\sum_{m=1}^{n_\epsilon} |r_m|\right|^p\right) &\leq n_\epsilon^{p-1} \sum_{m=1}^{n_\epsilon} \mathbb{E}|r_m|^p \\ &\leq C(\epsilon + \mathrm{nl}(\epsilon))^p,\end{aligned}$$

and the result follows from (4.19). □

In the same way, we can estimate the difference of the n_ϵ-th jump time of both processes :

Lemma 4.11. *Let $n_\epsilon \geq 1$ a deterministic integer. Then*

$$\epsilon \mathbb{E}\left(\sup_{0\leq n\leq n_\epsilon} |T_n - T_n^c|^p\right)^{1/p} \leq C_p\left(\epsilon^\delta + \mathrm{nl}_2(\epsilon)\right) \tag{4.21}$$

Proof. First, using the estimates of jump times (4.3)-(4.7), we can decompose the differences of jump times of both processes as follows :

$$\begin{aligned}T_n - T_n^c &= \sum_{m=0}^{n-1} \Delta T_{m+1} - \Delta T_{m+1}^c \\ &= \sum_{m=0}^{n-1} \delta_{m+1} + d_{m+1} + r_{m+1},\end{aligned}$$

where, by definition,

$$\begin{cases}\delta_{m+1} := -\frac{b^T \tau_\epsilon}{\lambda_0}(Z_{T_{m+1}} - Z_{T_m}), \\ d_{m+1} := \epsilon \Delta T_0 b^T \tau_\epsilon \nabla S(X_{T_n}) \mathcal{V}_n \\ \qquad\qquad -\epsilon \frac{\theta_{n+1}}{\lambda_0^2} A_\epsilon^T (X_{T_n^c}^c) \mathcal{V}_n \\ r_{m+1} := \left(\theta_{m+1}^3 + \theta_{n+1}\right) \epsilon \mathcal{O}(\epsilon^\delta + \mathrm{nl}_2(\epsilon)).\end{cases}$$

We can write for $n \leq n_\epsilon$:

$$|T_n - T_n^c| \leq \left|\frac{b^T \tau_\epsilon}{\lambda_0}(Z_{T_n} - Z_{T_0})\right| + \sup_{0\leq l\leq n_\epsilon} \left|\sum_{m=1}^{l} d_m\right| + \sum_{m=1}^{n_\epsilon} |r_m|.$$

Now, we can remark that the discrete time process $(M_l)_{l\geq 0} := \left(\sum_{m=1}^{l} d_m\right)_{l\geq 0}$ is a martingale with respect to the filtration $\mathcal{F}_l = \sigma((\theta_m, \mathcal{V}_{m-1}) | 1 \leq m \leq l)$ with $M_0 = 0$. Using the Burkholder-Davies-Gundy inequality (4.20), together with:

$$\mathbb{E}(|m_m|^p) \leq C\epsilon^{p\delta},$$

yields

$$\mathbb{E}\left(\sup_{0\leq l\leq n_\epsilon}\left|\sum_{m=1}^{l} d_m\right|^p\right)^{1/p} \leq C\epsilon^{\delta-1}.$$

Using:

$$\left|\frac{b^T\tau_\epsilon}{\lambda_0}(Z_{T_n} - Z_{T_0})\right| = \mathcal{O}(\epsilon^{2\delta-1}) < \mathcal{O}(\epsilon^{\delta-1}),$$

and plugging in the rest term, we finally get (4.21). □

We can now conclude with the proof of Theorem 4.7 :

Proof. [Proof of Theorem 4.7] We consider a given integer $n_\epsilon \geq 1$. We decompose the difference between both processes as

$$\left|X_{\bar{t}}^{\epsilon} - X_{\bar{t}}^{c,\epsilon}\right| = \left|X_{\bar{t}}^{\epsilon} - X_{\bar{t}}^{c,\epsilon}\right| \mathbf{1}_{N\leq n_\epsilon} + \left|X_{\bar{t}}^{\epsilon} - X_{\bar{t}}^{c,\epsilon}\right| \mathbf{1}_{N>n_\epsilon}, \tag{4.22}$$

and use (4.15) to write

$$\left|X_{\bar{t}}^{\epsilon} - X_{\bar{t}}^{c,\epsilon}\right| \mathbf{1}_{N\leq n_\epsilon} \leq \sup_{0\leq n\leq n_\epsilon} |\Xi_n - \Xi_n^c| + \epsilon \sup_{0\leq n\leq n_\epsilon} |T_n - T_n^c| + (\theta_{N+1} + \theta_{N_c+1})\mathcal{O}(\epsilon). \tag{4.23}$$

Using now the estimates (4.18) and (4.21) in (4.23), we get

$$\mathbb{E}\left(\left|X_{\bar{t}}^{\epsilon} - X_{\bar{t}}^{c,\epsilon}\right|^p \mathbf{1}_{N\leq n_\epsilon}\right)^{1/p} \leq C(\epsilon + \epsilon^\delta + \mathrm{nl}_2(\epsilon)) \tag{4.24}$$

$$+ C\epsilon\mathbb{E}(\theta_N^p + \theta_{N^c}^p)^{1/p}$$

$$= \mathcal{O}(\epsilon + \epsilon^\delta + \mathrm{nl}_2(\epsilon)), \tag{4.25}$$

where in the last line we have used a technical lemma (Lemma Appendix A.1).

It remains to control the probability of the event $\{N > n_\epsilon\}$. To this purpose, we consider $t \mapsto N_t \in \mathbb{N}$ the Poisson process uniquely defined by:

$$\sum_{n=1}^{N_t} \theta_n \leq t < \sum_{n=1}^{N_t+1} \theta_n.$$

By definition of N and N_c, there is a constant C such that $N_{C\bar{t}/\epsilon^2} \geq N$ and $N_{C\bar{t}/\epsilon^2} \geq N^c$, so that we can write:

$$\left|X_{\bar{t}}^{\epsilon} - X_{\bar{t}}^{c,\epsilon}\right| \leq \epsilon C\left(\sum_{n=1}^{N_{C\bar{t}/\epsilon^2}+1} \theta_n\right) \mathbf{1}_{N_{C\bar{t}/\epsilon^2}>n_\epsilon}. \tag{4.26}$$

Let us choose $n_\epsilon \geq 4eC\bar{t}/\epsilon^2$, a choice that satisfies $n_\epsilon = \mathcal{O}(\epsilon^{-2})$. A second technical lemma (Lemma Appendix A.2) implies that

$$\mathbb{E}\left(\left|X_{\bar{t}}^{\epsilon} - X_{\bar{t}}^{c,\epsilon}\right|^p \mathbf{1}_{N>n_\epsilon}\right) \leq \epsilon C_p 2^{-1/(C\epsilon^2)}, \tag{4.27}$$

so that the contribution of the latter to the final estimate (4.11) is negligible, since it is exponentially small with respect to ϵ^{-2}. The proof is complete. □

4.3. *Discussion about the sharpness of the coupling*

One may ask about the "sharpness" of the precise estimates of the coupling, given mainly by Theorem 4.7; (see also (4.18) (position shift), and (4.21) (time shift)). Let us discuss the case where $c_F = 0$, $\delta \in [1/k, 1/(k-1)]$. In this case the dominant error term in the jump times expansion of the internal state model in Lemma 4.8 in [26] is due to the non-linearity of the turning rate: $c_\lambda \epsilon^{k\delta-1}$. Since this term is due to the internal state mechanism, it can be conjectured that the difference between the density of the internal state model and gradient sensing model will be at least of this order. On the other hand, the coupling estimate in Theorem 4.7 is controlled by the same term: $c_\lambda \epsilon^{k\delta-1}$, suggesting the sharpness of the latter.

5. Numerical illustration

In this section, we demonstrate the validity of the analysis above. For the numerical experiments, we restrict ourselves to one space dimension. For the process with internal state, we use internal dynamics that are given by the scalar cartoon model (2.5). In all experiments, τ is chosen independently of ϵ so that $\delta = 1$. The corresponding turning rate is given by (2.6) with $\lambda_0 = 1$. For the control process, we choose the linear turning rate (2.14), with parameters (2.22) such that the two processes have the same diffusion limit; in particular, $b = 1$. The physical domain is $x \in [0, 20]$, and we use reflecting boundary conditions, *i.e.* the bacterial velocity is reversed when $x = 0$ or $x = 20$. We fix a scalar bimodal chemoattractant concentration field

$$S(x) = \alpha\left(\exp\left(-\beta(x-\xi)^2\right) + \exp\left(-\beta(x-\eta)^2\right)\right), \tag{5.1}$$

in which $\alpha = 5$, $\beta = 1$, $\xi = 7.5$ and $\eta = 12.5$.

Single bacterium as a function of time. In a first experiment, we simulate a single bacterium evolving according to the fine-scale model (2.4), in which we set the parameters to $\epsilon = 0.2$ and $\tau = 1$. We compare this evolution to that of a bacterium that satisfies the corresponding control process (2.12). Both simulations are performed using the same random numbers starting from the initial condition $X_0 = X_0^c = 8$ and $V_0 = V_0^c = +1$; the bacterium with internal state starts from $Y_0 = S(X_0)$. The time step $\delta t = 0.1$. The results are shown in figure 2. We see a very good coupling initially, which degrades over time. Note the time shift in the short time picture. In the long time picture, coupling is completely lost at some point.

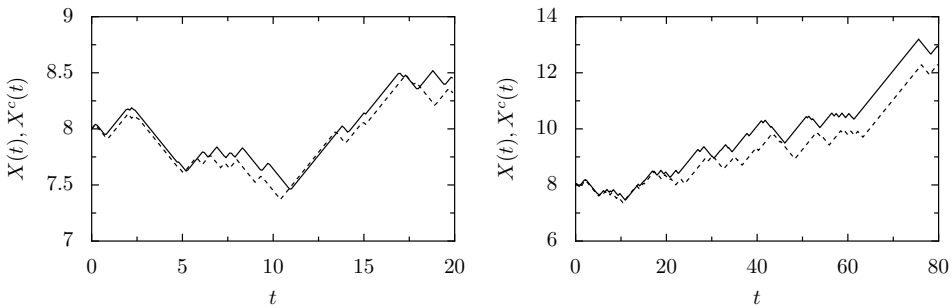

Fig. 2. Evolution of a single bacterium evolving according to the fine-scale process (2.4) (solid line) and the corresponding control process (2.12) (dashed) on short (left) and long (right) time-scales.

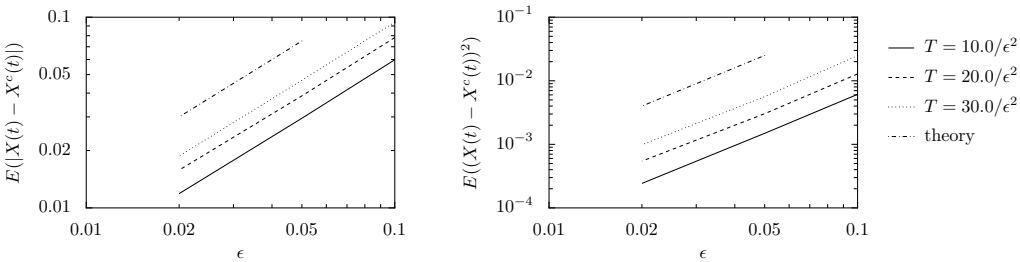

Fig. 3. Empirical mean (left) and variance (right) of the difference between the fine-scale process (2.4) and the corresponding control process (2.12) as a function of ϵ for different values of the reporting time T. The theoretical slope is indicated with a dashdotted line. The sample size is $N = 10000$.

Expectation and variance as a function of ϵ and $\bar{t}$. Next, we repeat the experiment using $N = 10000$ particles and compute the empirical mean and variance of the coupling, i.e.,

$$E(|X_{\bar{t}} - X^c_{\bar{t}}|) = \frac{1}{N}\sum_i^N \left(\left|X^i_{\bar{t}} - X^{i,c}_{\bar{t}}\right|\right), \text{ resp., } E\left((X_{\bar{t}} - X^c_{\bar{t}})^2\right) = \frac{1}{N}\sum_i^N \left(X^i_{\bar{t}} - X^{i,c}_{\bar{t}}\right)^2.$$

As fine-scale parameters, we choose $\tau = 1$, $\lambda_0 = b = 1$ and several values of ϵ. The chemoattractant concentration is again given as (5.1), now with $\alpha = \beta = 1$, $\xi = 7.5$ and $\eta = 12.5$. The time interval for the computation is $t \in [0, 30/\epsilon^2]$. Figure 3 shows the dependence in ϵ of the coupling, by plotting the empirical mean and variance defined above as a function of ϵ for different values of the reporting time T. The results shown in figure 3 are in clear accordance with the theoretical slope predicted by the asymptotic analysis. Figure 4 shows the mean and variance of the coupling difference as a function of time. The time dependence of the variance has not been analyzed mathematically in Section 4; the specific behaviour viewed in figure 4 and is probably due to: (i) sufficiently short diffusive times; (ii) the specific double-well

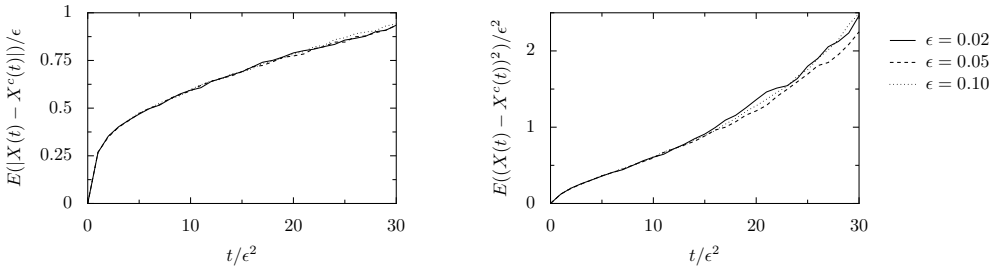

Fig. 4. Evolution of the empirical mean (left) and variance (right) of the difference between the fine-scale process (2.4) and the corresponding control process (2.12) as a function of $\bar{t}=t/\epsilon^2$ for different values of ϵ. The sample size is $N=10000$.

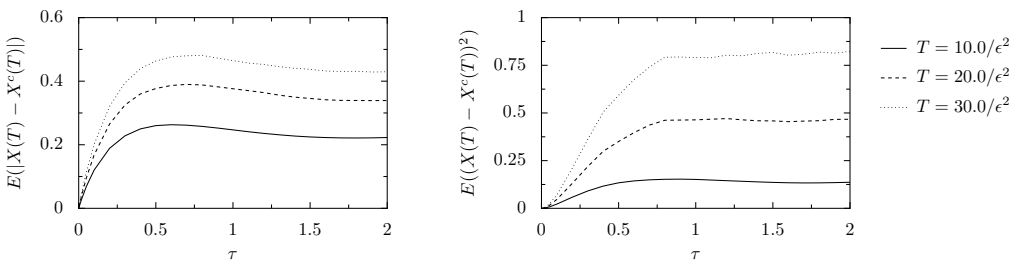

Fig. 5. Empirical mean (left) and variance (right) of the difference between the fine-scale process (2.4) and the corresponding control process (2.12) as a function of τ for different values of the reporting time T. The sample size is $N=10000$.

form of the chemoattractant potential.

Expectation and variance as a function of τ. Finally, in a last experiment, we illustrate the dependence on τ. To this end, we again simulate $N=10000$ particles choosing $\lambda_0=1$, $\epsilon=0.1$ and $X_0=7.5$, for different values of τ. The results in figure 5 show that the variance quickly increases with τ until it reaches a plateau for $\tau>1$.

6. Simulation on diffusive time scales

In this section, we consider a simulation of the density of an ensemble of particles, with and without variance reduction and/or reinitialization, as described in Section 3.1. We again restrict ourselves to one space dimension, with domain $x\in[0,20]$ and periodic boundary conditions. In this case, the kinetic equation corresponding to the control process reduces to the system

$$\begin{cases} \partial_t p_+^c + \epsilon\partial_x p_+^c = -\dfrac{\lambda^c(x,+1)}{2}p_+^c + \dfrac{\lambda^c(x,-1)}{2}p_-^c \\ \partial_t p_-^c - \epsilon\partial_x p_-^c = \dfrac{\lambda^c(x,+1)}{2}p_+^c - \dfrac{\lambda^c(x,-1)}{2}p_-^c \end{cases}. \tag{6.1}$$

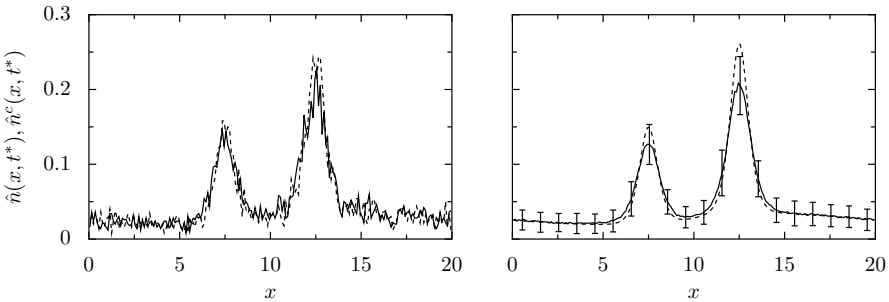

Fig. 6. Bacterial density as a function of space at $t=50/\epsilon^2$ without variance reduction. Left: one realization. Right: mean over 100 realizations and 95% confidence interval. The solid line is the estimated density from a particle simulation using the process with internal state; the dashed line is estimated from a particle simulation using the control process. Both used $N=5000$ particles. The dotted line is the solution of the deterministic PDE (2.20).

of two PDEs, which is straightforward to simulate using finite differences.

We fix the chemoattractant concentration field as (5.1), with parameters $\alpha=2$, $\beta=1$, $\xi=7.5$ and $\eta=12.5$. For the internal dynamics, the same model ((2.5)-(2.6), (2.14)-(2.22)) is used. The parameters are $\epsilon=0.5$, $\lambda_0=1$, $\tau=1$, $\delta t=0.1$.

All simulations are performed with $N=5000$ particles. The initial positions are uniformly distributed in the interval $x\in[13,15]$; the initial velocities are chosen uniformly, i.e., each particle has an equal probability of having an initial velocity of $\pm\epsilon$. The initial condition for the internal variable is chosen to be in local equilibrium, i.e., $Y_0^i=S(X_0^i)$. The initial positions and velocities of the control particles are chosen to be identical.

We discretize the continuum description (6.1) on a mesh with $\Delta x=0.1$ using a third-order upwind-biased scheme, and perform time integration using the standard fourth order Runge–Kutta method with time step $\delta t_{pde}=10^{-1}$. The initial condition is given as

$$p^+(x,0)=p^-(x,0)=\begin{cases}0.25, & x\in[13,15],\\ 0, & \text{otherwise.}\end{cases} \tag{6.2}$$

Simulation without variance reduction. First, we simulate both stochastic processes up to time $\bar{t}=50$ ($t=50/\epsilon^2$) and estimate the density of each of these processes $\hat{n}_N(x,\bar{t})$, resp. $\hat{n}_N^c(x,\bar{t})$, without variance reduction. The density is obtained via binning in a histogram, in which the grid points of the deterministic simulation are the centers of the bins. Figure 6 (left) shows the results for a single realization. We see that, given the fluctuations on the obtained density, it is impossible to conclude on differences between the two models. This observation is confirmed by computing the average density of both processes over 100 realizations. The mean densities are shown in figure 6 (right), which also reveals that the mean density of the control process is within the 95% confidence interval of the process

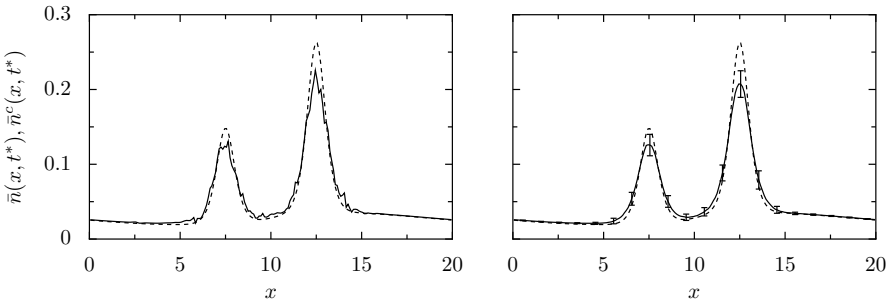

Fig. 7. Bacterial density as a function of space at $t = 50/\epsilon^2$ with variance reduction and reinitialization. Left: variance reduced density estimation of one realization with $N = 5000$ particles (solid) and deterministic solution for the control process (2.20) (dashed). Right: mean over 100 realization and 95% confidence interval (solid) and the deterministic solution for the control process (2.20) (dashed).

with internal state. Both figures also show the density that is computed using the continuum description, which coincides with the mean of the density of the control particles.

Simulation with variance reduction. Next, we compare the variance reduced estimation (3.1) with the density of the control PDE. We reinitialize the control particles after each coarse-scale step, i.e., each k steps of the particle scheme, where $k\delta t = \delta t_{pde}$, (here $k = 1$). The results are shown in figure 7. We see that, using this reinitialization, the difference between the behaviour of the two processes is visually clear from one realization (left figure). Also, the resulting variance is such that the density of the control PDE is no longer within the 95% confidence interval of the variance reduced density estimation (right figure). We see that there is a significant difference between both models: the density corresponding to the control process is more peaked, indicating that bacteria that follow the control process are more sensitive to sudden changes in chemoattractant gradient. This difference can be interpreted from the fact that the bacteria with internal state do not adjust themselves instantaneously to their environment, but instead with a time constant τ.

Simulation with variance reduction but without reinitialization. We also compare the variance reduced estimation (3.1) with the density of the control PDE without performing any reinitialization of the control process to restore the coupling. Figure 8 (right) shows the mean estimation of the density over 200 realizations, as well as the density of the control process. As evidenced by the error bars, simulating a single realization of the process, even with variance reduction, is not able to reliably reveal this difference. This phenomenon is due to the degeneracy of the coupling on long diffusive times. This is also illustrated in figure 8 (left), which

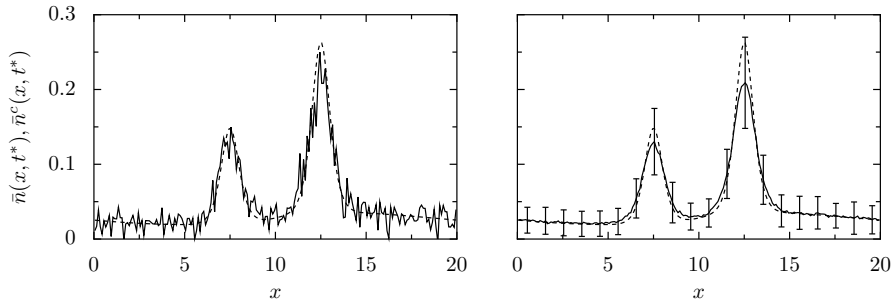

Fig. 8. Bacterial density as a function of space at $t = 50/\epsilon^2$ with variance reduction. Left: variance reduced density estimation of one realization with $N = 5000$ particles (solid) and deterministic solution for the control process (2.20) (dashed). Right: mean over 200 realization and 95% confidence interval (solid) and the deterministic solution for the control process (2.20) (dashed).

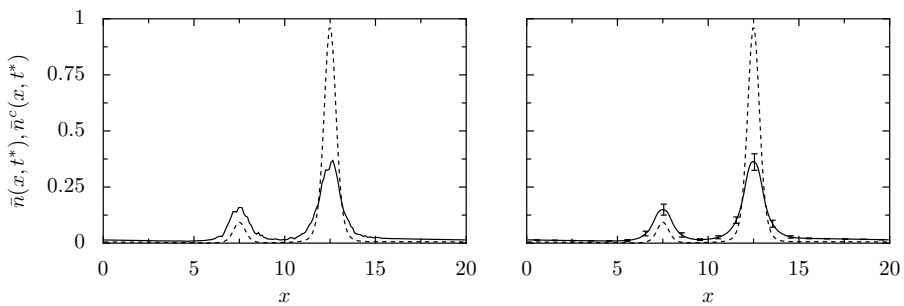

Fig. 9. Bacterial density as a function of space at $t = 50/\epsilon^2$ with variance reduction and reinitialization. Left: variance reduced density estimation of one realization with $N = 5000$ particles (solid) and deterministic solution for the control process (2.20) (dashed). Chemoattractant given by (5.1) with $\alpha = 5$. Right: mean over 100 realization and 95% confidence interval (solid) and the deterministic solution for the control process (2.20) (dashed).

compares the variance reduced density estimation of a single realization with the density of the control PDE. Note that, as predicted by the analysis, the variance is much larger in regions where $\nabla S(x)$ is large.

Finally, we repeat the experiment with a larger chemoattractant gradient. We again choose the chemoattractant field as (5.1), with the same parameters as above, except that we now take $\alpha = 5$. We also use the same discretization parameters as above. The variance reduced estimation (3.1), as well as the density of the control PDE, are shown in figure 9. We see that, although the difference between the process with internal state and the control process becomes much larger, the variance is still very well controlled by the coupling.

7. Conclusions and discussion

In this paper, we studied the simulation of stochastic individual-based models for chemotaxis of bacteria with internal state. We have used a coupling with a simpler, direct gradient sensing model with the same diffusion limit to obtain an "asymptotic" variance reduction. Throughout this work, we have assumed that the computation of the bacterial density using the kinetic description of the direct gradient sensing model can be performed accurately with a grid-based method. Note that in higher spatial dimensions, this may become cumbersome, and the grid-based computation should be carried out at the level of the advection-diffusion description (or any desired moment system) only, using a second level of coupling between the velocity-jump process associated with the kinetic description, and the stochastic differential equation associated with the advection-diffusion limit. This is left for future work.

The current algorithm allows to explore the differences between the fine-scale process with internal state and the simpler, coarse process *using a fixed number of particles, independently of the small parameter* ϵ. However, the computational cost to simulated an individual particle over diffusive time-scales becomes very large for $\epsilon \to 0$. In future work, we will therefore also study the development of truly "asymptotic preserving" schemes in the diffusion asymptotics, in the sense that the computational cost of the simulation of processes is independent of the small parameter ϵ. This will require to deal with two kinds of difficulties: (i) We will need to use an asymptotic preserving method to solve the density evolution of the control model on the grid; and (ii) We will need to extrapolate forward in time the state of the fine-scale simulation.

The first difficulty implies that, instead of solving the full kinetic equation associated with the control process, we only solve its diffusion limit. This may imply the use of a second level of variance reduction, coupling the control velocity jump process and its limiting drift-diffusion process.

The second difficulty, extrapolation in time, is related to the equation-free [19,20] and HMM [8,9] types of methodologies; ideas of this type can be traced back to Erhenfest [10]. One approach is to use a *coarse projective integration* method [14,19]. One then extrapolates the bacterial density over a projective time step on diffusive time scales, after which the projected density needs to be *lifted* to an ensemble of individual bacteria. For such methods, besides the effect of extrapolation on the variance of the obtained results, also the effects of reconstructing the velocities and internal variables have to be systematically studied.

Acknowledgements

The authors thank Radek Erban, Thierry Goudon, Yannis Kevrekidis and Tony Lelièvre for interesting discussions that eventually led to this work. This work was performed during a research stay of GS at SIMPAF (INRIA - Lille). GS warmly thanks the whole SIMPAF team for its hospitality. GS is a Postdoctoral Fellow

of the Research Foundation – Flanders. This work was partially supported by the Research Foundation – Flanders through Research Project G.0130.03 and by the Interuniversity Attraction Poles Programme of the Belgian Science Policy Office through grant IUAP/V/22 (GS). The scientific responsibility rests with its authors.

References

1. *Brownian Motion and Stochastic Calculus (Graduate Texts in Mathematics)*. Springer, 2nd edition, August 1991.
2. W Alt. Biased random-walk models for chemotaxis and related diffusion approximations. *J Math Biol*, 9(2):147–177, 1980.
3. W Alt. Orientation of cells migrating in a chemotactic gradient. *Adv Appl Probab*, 12(3):566–566, 1980.
4. Lowell L. Baker and Nicolas G. Hadjiconstantinou. Variance reduction for monte carlo solutions of the boltzmann equation. 17(5):051703, 2005.
5. A Bren and M Eisenbach. How signals are heard during bacterial chemotaxis: Protein-protein interactions in sensory signal propagation. *Journal of bacteriology*, 182(24):6865–6873, 2000.
6. FACC Chalub, PA Markowich, B Perthame, and Christian Schmeiser. Kinetic models for chemotaxis and their drift-diffusion limits. *Monatsh Math*, 142(1-2):123–141, 2004.
7. G Dimarco and L Pareschi. Hybrid multiscale methods II. Kinetic equations. *Multiscale Model Sim*, 6(4):1169–1197, 2007.
8. Weinan E and Bjorn Engquist. The heterogeneous multi-scale methods. *Comm. Math. Sci.*, 1(1):87–132, 2003.
9. Weinan E, Bjorn Engquist, Xiantao Li, Weiqing Ren, and Eric Vanden-Eijnden. Heterogeneous multiscale methods: A review. *Commun Comput Phys*, 2(3):367–450, 2007.
10. Paul Ehrenfest and Tatiana Ehrenfest. The conceptual foundations of the statistical approach in mechanics. pages xiv+114, 1990.
11. R Erban and HG Othmer. From individual to collective behavior in bacterial chemotaxis. *SIAM Journal on Applied Mathematics*, 65(2):361–391, 2004.
12. R Erban and HG Othmer. From signal transduction to spatial pattern formation in e-coli: A paradigm for multiscale modeling in biology. *Multiscale Model Sim*, 3(2):362–394, 2005.
13. SN Ethier and TG Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Statistics, 1986.
14. C. W Gear, IG Kevrekidis, and C Theodoropoulos. "coarse" integration/bifurcation analysis via microscopic simulators: micro-galerkin methods. *Computers and Chemical Engineering*, 26:941–963, 2002.
15. T Hillen and HG Othmer. The diffusion limit of transport equations derived from velocity-jump processes. *SIAM Journal on Applied Mathematics*, 61(3):751–775, 2000.
16. D Horstman. From 1970 until present: the Keller–Segel model in chemotaxis and its consequences I. *Jahresber. Deutsh. Math.-Verein*, 105(3):103–165, 2003.
17. D Horstman. From 1970 until present: the Keller–Segel model in chemotaxis and its consequences II. *Jahresber. Deutsh. Math.-Verein*, 106(2):51–69, 2004.
18. E Keller and L Segel. Initiation of slime mold aggregation viewed as an instability. *Journal on Theoretical Biology*, 26:399–415, 1970.
19. IG Kevrekidis, C. W Gear, JM Hyman, PG Kevrekidis, O Runborg, and C Theodoropoulos. Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level tasks. *Comm. Math. Sciences*, 1(4):715–762, 2003.

20. Ioannis G Kevrekidis and Giovanni Samaey. Equation-free multiscale computation: Algorithms and applications. *Annu Rev Phys Chem*, 60:321–344, 2009.
21. P Kloeden and E Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics*. Springer-Verlag (Berlin), 1992.
22. J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.
23. HG Othmer, SR Dunbar, and W Alt. Models of dispersal in biological-systems. *J Math Biol*, 26(3):263–298, 1988.
24. HG Othmer and T Hillen. The diffusion limit of transport equations ii: Chemotaxis equations. *SIAM Journal on Applied Mathematics*, 62(4):1222–1250, 2002.
25. C Patlak. Random walk with persistence and external bias. *B Math Biol*, 15:311–338, 1953.
26. M Rousset and G Samaey. Individual-based models for bacterial chemotaxis in the diffusion limit. *Mathematical Models and Methods in Applied Sciences*, 2010 (submitted).
27. David W. Scott. *Multivariate density estimation: theory, practice, and visualization*. Wiley, 1992.
28. B W. Silverman. *Density estimation for statistics and data analysis*, volume 26 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, 1986.
29. AM Stock. A nonlinear stimulus-response relation in bacterial chemotaxis. *P Natl Acad Sci Usa*, 96(20):10945–10947, 1999.

Appendix A. Technical lemmas, used in proof of Theorem 4.7

Lemma Appendix A.1 (Technical lemma). *Let $(\theta_n)_{n\geq 1}$ a sequence of independent exponential random variables with mean 1, and $(T_n)_{n\geq 1}$ a strictly increasing sequence of random times with $T_0=0$ such that:*

- *For any $m\geq 1$, the sequence $(\theta_n)_{n\geq m+1}$ is independent of the past $(\theta_n,T_n)_{n\leq m}$.*
- *There is a constant C such that for all $n\geq 0$,*

$$\frac{1}{C}\theta_{n+1}\leq T_{n+1}-T_n\leq C\theta_{n+1}. \tag{A.1}$$

Let N_t a random integer such that:

$$T_{N_t}\leq t\leq T_{N_t+1}.$$

Then for any integer $k\geq 0$, there is a constant C_k independent of t such that:

$$\mathbb{E}\left(\theta_{N_t+1}\right)^k\leq C_k.$$

Proof. The result may seem rather intuitive, but the proof is slightly tricky. First, we consider sub-intervals of $[0,t]$ of the form:

$$I_p=[\frac{p-1}{P}t,\frac{p}{P}t],$$

for some $P\geq 1$ and $p=1,\ldots,P$.

Step (i). The first step consists in proving the following estimate:

$$\mathbb{E}\left(\theta_{N_t+1}^k\right) \leq \sum_{p=0}^{P-1} \mathbb{E}\left(\theta^k \mathbf{1}\left(\theta \geq \frac{p}{C}\frac{t}{P}\right)\right) \sup_{p=1\ldots P} \sum_{n=0}^{+\infty} \mathbb{P}(T_n \in I_p) \tag{A.2}$$

Indeed,

$$\begin{aligned}\mathbb{E}\left(\theta_{N_t+1}^k\right) &= \sum_{n=0}^{+\infty} \mathbb{E}\left(\theta_{n+1}^k \mathbf{1}\left(T_n \leq t \leq T_{n+1}\right)\right)\\ &\leq \sum_{n=0}^{+\infty}\sum_{p=1}^{P} \mathbb{E}\left(\theta_{n+1}^k \mathbf{1}\left(T_{n+1}-T_n \geq \frac{P-p}{P}t | T_n \in I_p\right)\right) \mathbb{P}\left(T_n \in I_p\right).\end{aligned}$$

Using the independence of θ_{n+1} with T_n, and the assumption (A.1), it yields,

$$\mathbb{E}\left(\theta_{n+1}^k \mathbf{1}\left(T_{n+1}-T_n \geq \frac{P-p}{P}t | T_n \in I_p\right)\right) \leq \mathbb{E}\left(\theta^k \mathbf{1}_{\left(\theta \geq \frac{P-p}{CP}t\right)}\right).$$

and finally changing the sum index p to $P-p+1$ yields (A.2).

Step (ii). The second step consists in the following decomposition. Let $I=[a,b]$ be a finite interval of $\mathbb{R}^+$, $T_{-1}=-\infty$. Then,

$$\sum_{n=0}^{+\infty} \mathbb{P}(T_n \in I) = \sum_{n=0}^{+\infty}\left(\mathbb{P}(T_n \in I, T_{n-1} \notin I) \sum_{m=0}^{+\infty} \mathbb{P}(T_{n+1},\ldots,T_{n+m} \in I | T_n \in I, T_{n-1} \notin I)\right). \tag{A.3}$$

The key is to write

$$\begin{aligned}\underbrace{\sum_{n=0}^{+\infty} \mathbb{P}(T_n \in I)}_{(a_0)} &= \sum_{n=0}^{+\infty} \mathbb{P}(T_{n-1} \notin I, T_n \in I) + \sum_{n=0}^{+\infty} \mathbb{P}(T_{n-1}, T_n \in I)\\ &= \sum_{n=0}^{+\infty} \mathbb{P}(T_{n-1} \notin I, T_n \in I) + \underbrace{\sum_{n=0}^{+\infty} \mathbb{P}(T_n, T_{n+1} \in I)}_{(a_1)}.\end{aligned}$$

We wish to iterate the decomposition of (a_0) to (a_1), (a_2), etc... For this purpose, remark that

$$\begin{aligned}(a_m) &= \sum_{n=0}^{+\infty} \mathbb{P}(T_n,\ldots,T_{n+m} \in I)\\ &= \sum_{n=0}^{+\infty} \mathbb{P}(T_n \in I)\mathbb{P}(T_{n+1},\ldots,T_{n+m} \in I | T_n \in I)\\ &\leq \left(\sum_{n=0}^{+\infty} \mathbb{P}(T_n \in I)\right) \mathbb{P}(\theta_1+\cdots+\theta_m \leq C(b-a))\\ &\xrightarrow{m\to+\infty} 0,\end{aligned}$$

so that we get in the end

$$(a_0) = \sum_{n=0}^{+\infty} \sum_{m=0}^{+\infty} \mathbb{P}(T_{n-1} \notin I, T_n, \ldots T_{n+m} \in I).$$

Factorizing $\mathbb{P}(T_n \in I, T_{n-1} \notin I)$ with Bayes' formula yields (A.3).

Step (iii). The estimate (A.2) yields

$$\mathbb{E}\left(\theta_{N_t+1}^k\right) \leq \sum_{p=0}^{+\infty} \mathbb{E}\left(\theta^k \mathbf{1}_{\left(\theta \geq \frac{p}{C} \frac{t}{P}\right)}\right) \sum_{m=0}^{+\infty} \mathbb{P}\left(\theta_1 + \cdots + \theta_m \leq \frac{t}{CP}\right). \tag{A.4}$$

Indeed for $I = [a,b]$, using the independence of $(\theta_{n+1}, \ldots, \theta_{n+m})$ from (T_n, T_{n-1}):

$$\sum_{m=0}^{+\infty} \mathbb{P}(T_{n+1}, \ldots, T_{n+m} \in I | T_n \in I, T_{n-1} \notin I) \leq \sum_{m=0}^{+\infty} \mathbb{P}\left(\theta_1 + \cdots + \theta_m \leq C(b-a)\right),$$

and

$$\sum_{n=0}^{+\infty} \mathbb{P}(T_n \in I, T_{n-1} \notin I) = \mathbb{P}(\exists n \geq 0 | T_n \in I) \leq 1.$$

The two last estimates in (A.3) yields:

$$\sup_{p=1\ldots P} \sum_{n=0}^{+\infty} \mathbb{P}(T_n \in I_p) \leq \sum_{m=0}^{+\infty} \mathbb{P}\left(\theta_1 + \cdots + \theta_m \leq \frac{t}{CP}\right),$$

and (A.4) follows. Finally, we choose $P = \lfloor t+1 \rfloor$ so that the right hand side of (A.4) does not depend on t any more. Finally we use the exponential convergence of the terms of the two series in the right hand side of (A.4) (using for instnace large deviation estimates for the second) one, to conclude that the two series indeed converge, and conclude the proof. □

Lemma Appendix A.2 (Technical lemma). *Let $(\theta_n)_{n \geq 0}$ be a sequence of exponential random variables of mean 1, and $t \mapsto N_t \in \mathbb{N}$ the Poisson process uniquely defined by:*

$$\sum_{n=1}^{N_t} \theta_n \leq t < \sum_{n=1}^{N_t+1} \theta_n.$$

Let $n_0 \in \mathbb{N}$ such that $n_0 \geq 4\mathrm{e}t$, $p_1 \geq 1$, and $p_2 \geq 1$. Then there is a constant C_{p_1,p_2} independent of (n_0, t) such that:

$$\mathbb{E}\left(\left|\sum_{n=1}^{N_t+1} \theta_n^{p_1}\right|^{p_2} \mathbf{1}_{N_t > n_0}\right) \leq C_{p_1,p_2} 2^{-n_0}.$$

Proof.

Step (i).
Let us condition on the different possible values of $N_t > n_0$:

$$\mathbb{E}\left(\left|\sum_{n=1}^{N_t+1}\theta_n^{p_1}\right|^{p_2}\mathbf{1}_{N_t>n_0}\right) = \sum_{m=n_0+1}^{+\infty}\mathbb{E}\left(\left|\sum_{n=1}^{N_t+1}\theta_n^{p_1}\right|^{p_2}\Big|N_t=m\right)\mathbb{P}(N_t=m)$$
$$= \sum_{m=n_0+1}^{+\infty}\mathbb{E}\left(\left|\sum_{n=1}^{N_t+1}\theta_n^{p_1}\right|^{p_2}\Big|N_t=m\right)\mathrm{e}^{-t}\frac{t^m}{m!}.$$

Step (ii).
Let us denote $(\mathbb{P}_{t,m},\mathbb{E}_{t,m}) = (\mathbb{P}(.|N_t=m),\mathbb{E}(.|N_t=m))$ the probability/expectation conditionally on the event $\{N_t=m\}$. Conditionally on $\{N_t=m\}$, the sequence $(\theta_1,\ldots,\theta_{m+1})$ has the same distribution as $(tU_{(1)},tU_{(2)}-tU_{(1)},\ldots,tU_{(m)}-tU_{(m-1)},t-tU_{(m)}+\theta)$, where $(U_{(1)}<\cdots<U_{(m)})$ is the order statistics of m independent random variables $(U_1,\ldots,U_m)$ uniformly distributed on the interval $[0,1]$, and θ is an independent exponential random variable of mean 1 (see Section 2.4 of [22] for classical properties of Poisson processes). Thus denoting $U_{(0)}=0$ we can write:

$$\mathbb{E}\left(\left|\sum_{n=1}^{N_t+1}\theta_n^{p_1}\right|^{p_2}\mathbf{1}_{N_t>n_0}\right) =$$
$$\sum_{m=n_0+1}^{+\infty}\mathbb{E}_{t,m}\left(\left|\left(t-tU_{(m)}+\theta\right)^{p_1}+t^{p_1}\sum_{n=1}^{m}\left(U_{(n)}-U_{(n-1)}\right)^{p_1}\right|^{p_2}\right)\mathrm{e}^{-t}\frac{t^m}{m!}. \quad \text{(A.5)}$$

Using the inequality $(a+b)^p \le C_p a^p + b^b$ for any $a,b>0$, Jensen inequality, and denoting $U_{(m+1)}=1$, there is a constant C_{p_1,p_2} such that:

$$\left|\left(t-tU_{(m)}+\theta\right)^{p_1}+t^{p_1}\sum_{n=1}^{m}\left(U_{(n)}-U_{(n-1)}\right)^{p_1}\right|^{p_2} \le$$
$$C_{p_1,p2}\left(\theta^{p_2p_1}+t^{p_1p_2}(m+1)^{p_2-1}\sum_{n=1}^{m+1}\left(U_{(n)}-U_{(n-1)}\right)^{p_1p_2}\right). \quad \text{(A.6)}$$

Step (iii).
Let us compute $\mathbb{E}_{t,m}\left(\left(U_{(n)}-U_{(n-1)}\right)^{p_1p_2}\right)$ for $n=1,\ldots,m+1$. First, let us recall the probability density of a couple $(U_{(i)},U_{(j)})$ for $1\le i<j\le m$ in an order statistics of size m:

$$\mathbb{P}\left(U_{(i)}\in[u_i,u_i+du_i],U_{(j)}\in[u_j,u_j+du_j]\right)$$
$$= m!\mathbb{P}(U_1<\cdots<U_{i-1}<u_i)\times\mathbb{P}(U_i\in[u_i,u_i+du_i])$$
$$\times\mathbb{P}(u_i<U_{i+1}<\cdots<U_{j-1}<u_j)\times\mathbb{P}(U_j\in[u_j,u_j+du_j])\times\mathbb{P}(u_j<U_{j+1}<\cdots<U_m),$$
$$= m!\frac{u_i^{i-1}}{(i-1)!}\frac{(u_j-u_i)^{j-i-1}}{(j-i-1)!}\frac{(1-u_j)^{m-j}}{(m-j)!}du_i\,du_j. \quad \text{(A.7)}$$

As a consequence,

$$\begin{aligned}\mathbb{E}_{t,m}\left(\left(U_{(n)}-U_{(n-1)}\right)^{p_1p_2}\right) &= \int_{0<u_{n-1}<u_n<1} m!\frac{u_{n-1}^{n-2}}{(n-2)!}(u_n-u_{n-1})^{p_1p_2}\frac{(1-u_n)^{m-n}}{(m-n)!}du_{n-1}\,du_n \\ &= m!(p_1p_2)!\int_{0<u_{n-1}<u_n<1}\frac{u_{n-1}^{n-2}}{(n-2)!}\frac{(u_n-u_{n-1})^{p_1p_2}}{(p_1p_2)!}\frac{(1-u_n)^{m-n}}{(m-n)!}du_{n-1}\,du_n \\ &= \frac{m!(p_1p_2)!}{(m+p_1p_2)!},\end{aligned}$$

where in the last line we have used (A.7) with appropriate (i,j,m) to compute the integral. In the same way, the probability density of $U_{(i)}$ for $1\le i\le m$ in an order statistics of size m is given by:

$$\mathbb{P}\left(U_{(i)}\in[u_i,u_i+du_i]\right)=m!\frac{u_i^{i-1}}{(i-1)!}\frac{(1-u_i)^{m-i}}{(m-i)!}du_i,$$

so that

$$\begin{aligned}\mathbb{E}_{t,m}\left(U_{(1)}^{p_1p_2}\right) &= m!(p_1p_2)!\int_{0<u_1<1}\frac{u_1^{p_1p_2}}{(p_1p_2)!}\frac{(1-u_1)^{m-1}}{(m-1)!}du_1 \\ &= \frac{m!(p_1p_2)!}{(m+p_1p_2)!},\end{aligned}$$

and in the same way,

$$\mathbb{E}_{t,m}\left((1-U_{(m)})^{p_1p_2}\right)=\frac{m!(p_1p_2)!}{(m+p_1p_2)!}.$$

Step (iv).
Remarking that $\mathbb{E}_m(\theta^{p_1p_2})=(p_1p_2)!$, and inserting the computations of Step (ii) in (A.5)-(A.6) yields:

$$\begin{aligned}&\mathbb{E}\left(\left|\sum_{n=1}^{N_t+1}\theta_n^{p_1}\right|^{p_2}\mathbf{1}_{N_t>n_0}\right)\le \\ &\quad C_{p_1p_2}\sum_{m=n_0+1}^{+\infty}\left((p_1p_2)!+t^{p_1p_2}(m+1)^{p_2-1}\frac{m!(p_1p_2)!}{(m+p_1p_2)!}\right)\mathrm{e}^{-t}\frac{t^m}{m!}.\end{aligned}$$

Using the inequalities $\frac{(m+1)^{p_2-1}m!}{(m+p_1p_2)!}\le 1$, $\mathrm{e}^{-t}\le 1$ and

$$\sum_{m=n_0+1}^{+\infty}\frac{t^m}{m!}\le\frac{t^{n_0}}{n_0!},$$

it yields

$$\mathbb{E}\left(\left|\sum_{n=1}^{N_t+1}\theta_n^{p_1}\right|^{p_2}\mathbf{1}_{N_t>n_0}\right)\le C_{p_1p_2}\frac{t^{n_0+p_1p_2}}{n_0!},$$

so that in the end, using the assumption $n_0 \geq 4et$ and the Stirling formula,

$$\mathbb{E}\left(\left|\sum_{n=1}^{N_t+1}\theta_n^{p_1}\right|^{p_2}\mathbf{1}_{N_t>n_0}\right) \leq C_{p_1p_2}(\frac{n_0}{4\mathrm{e}})^{n_0+p_1p_2}\frac{1}{\sqrt{n_0}}(\frac{\mathrm{e}}{n_0})^{n_0}$$
$$\leq C_{p_1p_2}\frac{n_0^{p_1p_2}}{\sqrt{n_0}2^{n_0}}2^{-n_0} \leq C_{p_1p_2}2^{-n_0},$$

which yields the result. □