Applied Mathematics and Computation 232 (2014) 836–847

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Coarse-grained computational stability analysis and acceleration of the collective dynamics of a Monte Carlo simulation of bacterial locomotion

Constantinos Siettos

School of Applied Mathematics and Physical Sciences, National Technical University of Athens, 9, Heroon Polytechniou str., GR 15780 Athens, Greece

ARTICLE INFO

Keywords:
Complex systems
Bacterial locomotion
Collective dynamics
Moment closures
Monte Carlo simulations
Equation-free

ABSTRACT

Over the past few years it has been demonstrated that the so called equation-free framework establishes a link between traditional computational analysis and microscopic/stochastic simulation of complex systems. The underlying assumption is that macroscopic models can be in principle written in terms of a few statistical moments of the evolving distributions, but they are unavailable in closed form. Here it is shown how this multiscale framework can be used to find the level at which closures should be sought for a Monte Carlo simulation of bacterial locomotion with flagellar motility. For illustration purposes, a simple biologically inspired one-dimensional in space Monte Carlo model of *Escherichia coli* locomotion was used. The coarse-grained stability analysis of the emergent dynamics revealed that the collective behavior for the particular model can be parametrized by just a few moments of the distribution of the cells positions indicating that higher-order moments as well as other "internal" variables of the model become relatively fast in time, functionals of these few coarse-grained observables. Acceleration of the Monte Carlo simulations in time based on these "slow" macroscopic variables is also demonstrated.

1. Introduction

Bacterial locomotion, the process by which bacteria "swim", change their speed and/or orientate in space responding to environmental changes has been thoroughly studied as it constitutes one of the basic survival and growth mechanisms of many micro-organisms. Revealing and understanding these mechanisms that pertain to the function of this biological directional "kinesis" is of outmost importance in several fields of biology, biomechanics and medicine [1,2], ocean ecology [3], biologically-inspired robotics [4] and orthopedics [5,6]. Advances in the field have gained a lot from the extensive study of *Escherichia coli* (*E. coli*) a bacterium that is commonly found in animals intestines. The chemotaxis pathway and biomechanics of *E. coli* have been well investigated through experiments [7–13] providing useful insights also for other species. A recent study provided that more than half of the bacteria have chemotactic ability [14].

The extensive study of *E. coli* has allowed the development of many dynamical models of bacterial locomotion ranging from the macro [15–21] to mechanistic-based micro-scale ones incorporating biological information even at the molecular inner-cell level [22–27] (for a detailed review of these models please refer to [28–30]). No doubt, these mathematical models are invaluable tools that can help shed light on still unknown issues regarding not only chemotaxis but also many other "adaptive" biochemical and genetic networks as well as human health [14,31] and life-cycle [32].

E-mail address: ksiet@mail.ntua.gr

http://dx.doi.org/10.1016/j.amc.2014.01.151
0096-3003/

For example, the collective dynamics for bacterial chemotaxis can be represented – under certain assumptions-by a Partial-Differential Equation (PDE) describing the evolution of the probability density, say, $\rho(\mathbf{x}, t)$, of the bacteria in space [33]:

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} = \mathcal{Q}(\mathbf{x}, \rho, \mathbf{p}), \tag{1}$$

where $\mathcal{Q}$ is the partial differential operator and $\mathbf{p}$ denotes the vector containing other "internal" variables. If cell division of death is ruled out, the above equation reads [34]:

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} = \nabla \cdot (D\nabla\rho - \rho v), \tag{2}$$

where D is the diffusion coefficient and v is the chemotactic velocity.

The above equation is not in a closed form as v is not explicitly related to ρ. Over the years, various approximations have been proposed that express v in terms of other internal variables such as the gradient of the concentration of the attractant (or repellant). For example, a most common approximation, as extracted from kinetic theory is given by the Patlak–Keller/Segel relation reading [15–18]:

$$v = \chi \nabla c, \tag{3}$$

where χ denotes the chemotactic sensitivity and c is the concentration of the chemo-attractant (or repellant). The above closure is derived assuming that the cells emit immediately the chemical chemoattractants which are then diffused.

Many other approaches for the calculation of the chemotactic sensitivity or the locomotion velocity expressed in terms of other internal variables have been proposed (see the discussion in [34] and for a review see also [35]). Yet, such closure relations are based on approximations and assumptions that may introduce certain biases in the analysis of the collective dynamics. This prevents the performance of important computational tasks such as stability and bifurcation analysis, control and optimization, which rely on the availability of models, lumped or distributed, in terms of macroscopic (coarse) variables.

On the other hand, more realistic-non phenomenological models of the underlying physics can be written at the microscopic/cellular level, where modeling is usually performed via stochastic simulators (e.g. Monte Carlo, Brownian dynamics, Lattice-Boltzmann, Markov chains). However, at this fine biological level of representation, due to the emergent complexity, stochasticity of the evolving rules and the "curse of dimensionality" standard system-level analysis cannot be used efficiently to handle the given detailed information.

Towards this aim, it has been shown that "coarse timesteppers" establish a link between traditional numerical analysis and microscopic simulation [36–45]. The underlying assumption of this equation-free procedure is that macroscopic models exist and can, in principle, be derived from microscopic ones (description at a much finer level) but they may lack a closed form description (e.g. in terms of moments of the microscopic distributions) of their governing coarse equations. This system identification-based computational approach sidesteps the necessity of deriving good explicit closures, thus enhancing our efficiency in dealing with the problem in a systematic way. In [46] it is demonstrated how coarse-projective integration can be carried out using a microscopic-stochastic model of chemotaxis. In [47] coarse-projective integration is studied and analyzed in detail for kinetic Monte-Carlo simulations of random walks simulating bacterial chemotaxis.

Here it is demonstrated how the equation-free framework can be applied to (i) perform system (macroscopic) level fixed point and stability analysis and (ii) enable the acceleration of the evolution computations of a Markov chain-Monte Carlo simulation of the collective dynamics of *E. coli* locomotion. Using a simple Gaussian chemoattractant profile, it is shown that near the equilibrium the collective dynamics of this system evolve on a two-dimensional slow manifold that can be parametrized by the mean and the variance of the microscopic distribution. The analysis validates the conjecture that under certain assumptions the macroscopic dynamics can be described by the probability density function (serving as the macroscopic variable) of the positions and that all other "internal" variables get slaved relatively fast to the coarse-grained description.

The paper is organized as follows: in Section 2, the biologically inspired Monte-Carlo model which describes the dynamics of the chemotaxis of *E. coli* at micro level is presented. In Section 3 the proposed framework, the equation-free approach for multiscale computations is briefly presented. In Section 4 it is shown how the concept of coarse timesteppers can be combined with system identification techniques in order to accelerate in time the detailed simulations. Section 5 provides the results of the analysis followed by the conclusions section.

2. *Escherichia coli* locomotion: a bio-mechanical-based biased random walk model

Bacteria, such as *E. coli*, have several flagella that help them direct their movements towards the most likely direction. Each flagellum rotates by the action of a rotary motor. A typical *E. coli* bacterium has up to 6 flagella. These can rotate either in a clockwise (CW) manner, causing the bacterium to tumble, or counter-clockwise (CCW), causing the bacterium to swim in a straight line. The overall movement of a bacterium is the result of altering tumble and swim phases. Tumbling frequency and duration time are most important factors of controlling the locomotion of the bacteria. When a decreasing concentration of an attractant or increasing concentration of a repellant is detected, the bacteria will increase the tumbling frequency and consequently the probability of changing direction.

An *E. coli* bacterium has three types of receptors (transmembrane proteins) for detecting changes in the concentration of attractants and repellents. The signals from these receptors are transmitted through the signal transduction system to the

flagella motors, where Che proteins are activated. Increasing concentrations of the Che-Y protein, which is the output of this system, increases the tumbling frequency, and therefore bias the CW rotation. In general, the evolution of the changes in the transduction system can be represented by an evolution equation of the form:

$$\frac{d\mathbf{u}}{dt} = \mathcal{F}(\mathbf{u}, S), \tag{4}$$

where $\mathbf{u} \in \Re^q$ is the state vector, $S \in \Re$ denotes the stimulus signal to the system, and, $\mathcal{F} : \Re^q \times \Re \to \Re^q$ is the (smooth) evolution operator. In most of the cases, there is some kind of adaptation to the input signal: when S is time-independent, some functional of $\mathbf{u}$ should be also time-independent.

Here a simple two dimensional ($q = 2$) ODEs model for excitation and adaptation is used [11]:

$$\frac{du_1}{dt} = \frac{f(S(t)) - (u_1 + u_2)}{\tau_e}, \tag{5}$$

$$\frac{du_2}{dt} = \frac{f(S(t)) - u_2}{\tau_a}, \tag{6}$$

where f is the function encoding the signal transduction steps, and should satisfy the condition $f(0) = 0$. Here it is taken as:

$$f(S) = amp\frac{S}{1+S}, \tag{7}$$

where the gain amp scales the input signal.

The response of this simple model occurs in two time scales: the scale of excitation, which is characterized by the time constant τ_e and the scale of adaptation, which is characterized by the time constant τ_a. If $\tau_e \ll \tau_a$, then whenever $t \gg \tau_a$, the model relaxes to:

$$u_2 = f(S) - u_1. \tag{8}$$

The probability that the cell is running (i.e., the number of flagella rotating CCW is greater or equal to $N/2$, N being number of flagella) is given by the sum of the binomial probability mass function (voting hypothesis):

$$p_{CCW} = \sum_{j=N/2}^{N} \binom{N}{j} p_{CCW}^j (1 - p_{CCW})^{N-j}. \tag{9}$$

From experimental data we know that the probability bias, p_{CCW} of a single flagella motor in the absence of stimulus, for a 2.95 μM concentration of Che-Y protein is 0.64 [11]. For $N = 6$, Eq. (9) gives that the probability of CCW bias of a single bacteria is $p_{CCW} = 0.88$.

The behavior of a single bacterium is determined by the following evolution rules [11]:

1. Initialize the position of each bacterium, say $x(t)$, its velocity v (here assumed constant for all bacteria) as well as the states of the excitation-adaptation model given in Eq. (6) and the sampling time Δt.
2. Draw a random number ζ from a uniform random distribution on $[0, 1]$.
3. Compare ζ with the probability of switching direction of flagella rotation:

$$p \sim k_{\pm} \pm \Delta t, \tag{10}$$

where $k_{\pm}$ are rational functions of the reaction rate constants. In [11], the authors have measured the equilibrium CW (P_{CW}) and CCW bias (P_{CCW}) as a Hill function of the dissociation constant K_d and the concentration Y of Che-Y as follows:

$$P_{CW} = \frac{Y^n}{K_d^n + Y^n} = \frac{k_+}{k^+ + k^-}. \tag{11}$$

The transition rates between CW and CCW rotations are given by:

$$k_+ = \frac{nY^{n-1}}{K_d^n + Y^n}, \tag{12}$$

$$k_- = \frac{1}{Y}\frac{nK_d^n}{K_d^n + Y^n}, \tag{13}$$

where $n = 10.3$ and $K_d = 3.1$. In order to match the experimental data on the switching behavior, Y was calculated using the following relation:

$$Y = \overline{Y} - gu_{1s}, \tag{14}$$

where $\overline{Y} = 2.95(\mu M)$ is the ligant Che-Yp concentration that gives the correct experimental flagellar bias and u_{1s} is the steady state value of u_1; g is a gaining factor ranging from three to six that here was set equal to five.

```
3.a IF (a flagellum rotates CW AND ζ > p) THEN
        keep rotating CW
    ELSE
        switch to CCW rotation
    ENDIF
3.b IF (a flagellum rotates CCW AND ζ > p) THEN
        keep rotating CCW
    ELSE
        switch to CW rotation
    ENDIF
3.c IF (the number of flagella that rotate CW is less than 3) THEN
        tumble
    ELSE
        run
          IF (previously running) THEN
            direction remains unchanged
          ELSE
          direction = ±1, with equal probability
        ENDIF
    ENDIF
3.d Update the position of the bacteria:
```

$$x(t+\Delta t) = x(t) + direction\, v\, \Delta t \tag{15}$$

3. Multiscale computations of the *E. coli* locomotion using coarse timestepping

Let us start with a brief overview of the concept of the"coarse timestepper" [40] assuming that due to the complexity of the system under study there are no explicit macroscopic equations in a closed form that can approximate the emergent collective dynamics in an efficient manner. Yet, we assume that the physics are known in a more detailed level and thus we can develop a microscopic computational model using statistical mechanics simulation techniques such as Molecular dynamics, Monte-Carlo, Brownian dynamics and cellular automata. In general, the microscopic simulator can be represented by the following map:

$$\mathbf{U}_{k+1} = \mathbf{S}_T(\mathbf{U}_k, \mathbf{p}), \tag{16}$$

where $\mathbf{U}_k \equiv \mathbf{U}(t_k)$ denotes the state vector of the microscopic distribution at time $t_k = kT$ and $\mathbf{S}_T : \Re^N \times \Re^m \to \Re^N$ is the time-evolution operator, $\mathbf{p} \in \Re^m$ is the vector of the system's parameters. Hence, the microscopic simulator reports the values of the states of the microscopic distribution after an arbitrarily chosen macroscopic time interval T.

Let us further assume that the macroscopic (system-level) dynamics of the system under study can be described by a map of the form:

$$\mathbf{u}_{k+1} = \mathbf{F}_T(\mathbf{u}_k, \mathbf{p}), \tag{17}$$

where $\mathbf{u} \in \Re^n$, $n \ll N$, denotes the macroscopic state vector, and $\mathbf{F}_T : \Re^n \times \Re^m \to \Re^n$ is a smooth macroscopic time-evolution operator. We furthermore assume that due to the complexity of the problem under study such a macroscopic model is not available, or it is overwhelming difficult to derive in a closed form.

When this is the case, a series of important system level tasks such as the tracing of coarse-grained solution branches in the parameter space, stability analysis, optimization and design of control systems cannot be performed by exploiting the arsenal of well-established techniques in the continuum.

The question that naturally arises is how one can systematically study and analyze the collective (macroscopic) dynamics, when the macroscopic evolution operator is not explicitly available in a closed form. The answer comes from the concept of timestepping which constitutes the "heart" of the equation-free computational approach [36–45]. The main idea is to sidestep the derivation of system-level equations and to use appropriately-initialized short runs in time of the microscopic/stochastic models to estimate necessary quantities "on demand". What the equation-free approach does, in fact, is to provide closures "on demand"; relatively short bursts of the fine scale simulator naturally establish in a strictly numerical manner the slaving relation between the fast and slow dynamics of the system under study (refer to [40] for more detailed discussions).

In a nutshell, the coarse timestepper consists of the following steps (see also Fig. 1):

(a) Prescribe a macroscopic initial condition (e.g., concentration profile) $\mathbf{u}(t_0)$;

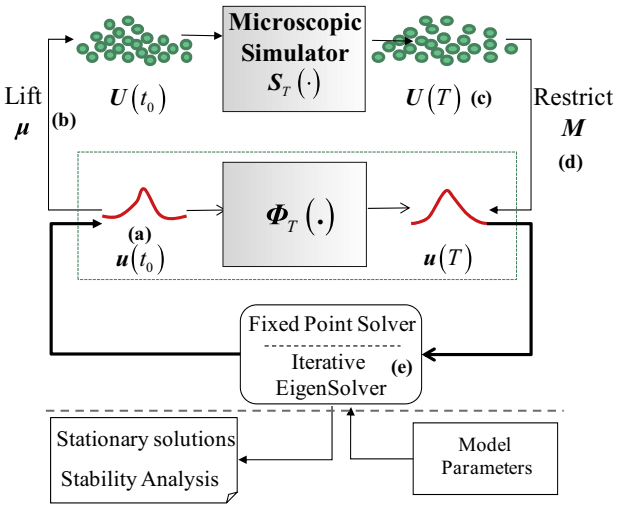

Fig. 1. The coarse timestepper: (a) Prescribe a macroscopic initial condition, (b) Transform it through lifting to consistent microscopic realizations, (c) Evolve these realizations using the microscopic simulator for the desired short macroscopic time T, (d) Obtain the restrictions, (e) Wrap around the map established numerical analysis algorithms to perform fixed point and stability analysis computations.

(b) Transform it through lifting to *consistent* microscopic realizations $\mathbf{U}(t_0) = \mu\mathbf{u}(t_0)$, where μ is the lifting operator.
(c) Evolve these realizations using the microscopic simulator for the desired short macroscopic time (time horizon) T, generating $\mathbf{U}(T)$. An appropriate choice of T can be estimated from the spectrum derived from the mean field equations linearization around their steady states [40].
(d) Obtain the restrictions $\mathbf{u}(T) = M\mathbf{U}(T)$; M denotes the restrict operator. The lift and restrict operators should satisfy $\mu M = \mathbf{I}$.
The above procedure can be considered as a "black box" coarse timestepper:

$$\mathbf{u}_{k+1} = \Phi_T(\mathbf{u}_k, \mathbf{p}). \tag{18}$$

Under certain assumptions regarding the separation between the time-scales of the system, $\Phi_T : \Re^n \times \Re^m \to \Re^n$ provides a numerical approximation of the unavailable macroscopic operator $\mathbf{F}_T$.
Thus
(e) At an outer level, and depending on the task we want to carry out, (such as the computation of fixed points, the stability analysis and optimization), well established numerical analysis algorithms can be utilized to estimate "on demand" the required quantities such as residuals, Jacobians, control matrices and Hessians. These algorithms call the timestepper as a black-box subroutine from nearby appropriately perturbed initial conditions and for relatively short time intervals.

3.1. Computational framework for steady-state and stability analysis of the collective dynamics

Coarse-grained (macroscopic) steady states can be obtained as fixed points, using T as reporting horizon, of the mapping:

$$\Phi_T : \mathbf{u} - \Phi_T(\mathbf{u}, \mathbf{p}) = \mathbf{0}, \tag{19}$$

using Newton–Raphson-like methods [48]. The procedure involves the iterative solution of the following linearized system:

$$\left[\mathbf{I} - \frac{\partial \Phi_T}{\partial \mathbf{u}}\right]\delta\mathbf{u} = -[\mathbf{u} - \Phi_T(\mathbf{u}, \mathbf{p})]. \tag{20}$$

The computation of the Jacobian $D\Phi_T \equiv \partial\Phi_T/\partial\mathbf{u}$ can be achieved in a fully numerical manner by calling the coarse timestepper from appropriately perturbed initial conditions. For example, using center finite differences the (i,j) element of the Jacobian is estimated by:

$$\frac{\partial \Phi_{Tj}}{\partial u_i} = \frac{\Phi_{Tj}(u_i + \varepsilon u_i) - \Phi_{Tj}(u_i - \varepsilon u_i)}{2\varepsilon u_i} + \mathcal{O}(\varepsilon u_i^2), \tag{21}$$

where Φ_{Tj}, u_i denote the jth and ith elements of Φ_T and $\mathbf{u}$ respectively; ε is a small and appropriately chosen scalar.

The above framework enables the temporal simulator to converge to both stable and unstable solutions and trace their locations through bifurcation points utilizing techniques such as the pseudo-arc length continuation approach.

For large-scale systems, fixed point calculations can be performed in a more efficient manner using matrix-free iterative solvers such as the Newton-Generalized Minimum Residual (Newton-GMRES) method [48]. The advantage of using matrix-free methods over more "traditional" techniques is that the explicit calculation and storage of the Jacobian $\partial\Phi_T/\partial\mathbf{u}$ is not required. Instead, what is really needed is matrix–vector multiplications which can be obtained at low computational cost by calling the timestepper from *nearby* initial conditions allowing the estimation of the action of the linearization of the map Φ_T on known vectors, as:

$$D\Phi_T(\mathbf{u})\cdot\mathbf{v}\approx\frac{\Phi_T(\mathbf{u}+\varepsilon\mathbf{v})-\Phi_T(\mathbf{u})}{\varepsilon}. \tag{22}$$

Alternative algorithms such as the Recursive Projection method [49] and other Newton–Picard methods such as the ones presented in [50,51] can be also used to compute both steady states and periodic solutions and construct their bifurcation diagrams.

The leading (algebraically largest) eigenvalues of the matrix $D\Phi_T$ determine the local stability of the system. For large-scale systems these can be estimated using again a matrix-free iterative eigensolver such as the Arnoldi procedure [52,53] exploiting the same timestepper approach. These algorithms share common procedures that in short can be described as follows:

Choose $\mathbf{v}_1\in\Re^n$ with $||\mathbf{v}_1||=1$.

For j = 1 Until Convergence.

1. Compute and store $D\Phi_T\cdot\mathbf{v}_j$ using Eq. (6)
2. Compute and store $h_{i,j}=<D\Phi_T\cdot\mathbf{v}_j,\mathbf{v}_i>,\ i=1,2,\ldots,j$
3. $r_j=D\Phi_T\cdot\mathbf{v}_j-\sum_{t=1}^{j}h_{ij}\mathbf{v}_i$
4. $h_{j+1,j}=<r_j,r_j>^{1/2}$
5. $\mathbf{v}_{j+1}=r_j/h_{j+1,j}$

End For

At step j, the algorithm produces an orthonormal basis $\{\mathbf{v}_1,\mathbf{v}_2,\ldots\mathbf{v}_m\}$ of the Krylov subspace K_j spanned by:

$$\{\mathbf{v}_1,D\Phi_T\cdot\mathbf{v}_1,\ldots,D\Phi_T^{j-1}\cdot\mathbf{v}_{j-1}\}. \tag{23}$$

The projection of $D\Phi_T$ on K_j is represented in the basis $\{\mathbf{v}_j\}$ by the upper Hessenberg matrix:

$$H_j=\mathbf{v}_j^T D\Phi_T\cdot\mathbf{v}_j, \tag{24}$$

whose elements are the coefficients $h_{i,j}$. At each j step, GMRES minimizes the residual $R=||\mathbf{u}-\Phi_T(\mathbf{u},\mathbf{p})||$.

Regarding the stability analysis, the eigenvalues of H_j provide approximations of the eigenvalues of $D\Phi_T$ for the outermost spectrum of $D\Phi_T$; the eigenvectors of $D\Phi_T$ are approximated by:

$$\mathbf{x}_j=\mathbf{v}_j\mathbf{z}_j, \tag{25}$$

where $\mathbf{z}_j$ are the eigenvectors of the Hessenberg (it may be computed using standard packages like EISPACK or LAPACK).

The approximation character of the algorithm, flashes a note of caution in the choice of the perturbation parameter ε since at the end of the algorithm what is usual done is the formation of the converged Hessenberg through the multiplication:

$$H_j=\mathbf{v}_j^T(A\mathbf{v}_j)\approx\mathbf{v}_j^T\frac{\Phi_T(\mathbf{u}+\varepsilon\mathbf{v}_j)-\Phi_T(\mathbf{u})}{\varepsilon}, \tag{26}$$

which is only an approximation to the actual Hessenberg. This may reflect to some "distortion" to the computation of the leading eigenvectors. One should choose according to the accuracy of the already converged solution.

3.2. The lift and restrict operators for the chemotaxis problem

As explained in Section 3.1, an explicit derivation of closures is not required. Instead the concept of coarse timestepping can be exploited to sidestep the derivation of such approximations. Here, the spatial distribution in $\mathbf{x}$, $\rho(\mathbf{x},t)$ is computed by approximating the corresponding inverse cumulative distribution function $ICDF(\rho)$ using n_S orthogonal polynomials of order q. Let us denote by $\mathbf{W}$ the matrix of dimension $(m\times n_s)$ containing the values of such n_s orthogonal polynomials over the m points in space on which the $ICDF(\rho)$ is computed.

The restrict operator, M, is defined as the product of $ICDF(\rho)$ with $\mathbf{W}$ resulting to the coefficients of the orthogonal polynomials $\mathbf{a}$, i.e.:

$$M\equiv\mathbf{a}=ICDF(\rho)\cdot\mathbf{W}. \tag{27}$$

Since $\mathbf{W}\mathbf{W}^T=\mathbf{I}$, the lift operator μ creates $ICDF(\rho)$ (i.e., a distribution in space) as:

$$\mu\equiv ICDF(\rho)=\mathbf{a}\cdot\mathbf{W}^T. \tag{28}$$

Based on the above coarse-grained steady states can be computed as fixed points of:

$$\mathbf{a} - \Phi_T(\mathbf{a}, \mathbf{p}) = \mathbf{0}, \tag{29}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_{n_S}]$

4. Coarse projective integration of the Markov chain Monte Carlo model

The coarse timestepper can be also used to perform coarse projective integration. The basic idea is that the coarse timestepper can be used to approximate the time derivatives of the corresponding continuum formulation, even if the continuum equations are not known in closed form. Specifically, the following steps are executed [40]:

(f) Repeat step (d) over several time steps, giving several $\mathbf{U}(t_i)$, as well as their restrictions $\mathbf{u}(t_i) = M\mathbf{U}(t_i)$, $i = 1, 2, \ldots, k+1$.
(g) Use the chord connecting these successive time-stepper output points to estimate the derivative of the continuum variables. Note that this does not require that we know the explicit continuum equations.
(h) Use this derivative in an outer integrator (such as forward Euler) to estimate the continuum state $\mathbf{u}(t_{k+1+m})$ much later in time.
(i) Go back to step (b).

For the chemotaxis problem, the proposed procedure can be summarized in a two-tier level pseudo-code as follows:

Do while {desired time T_{final}}
I. For $t = n, n+1, n+2, \ldots, n+l$

I1. Compute the Cumulative distribution function of x-positions $CDF(\rho)_t$ from the probability distribution function $PDF(\rho)_t$
I2. Approximate the Inverse Cumulative Distribution function of x-positions $ICDF(\rho)_t$ using n_S orthogonal polynomials.
I3. Restore the coefficients $\mathbf{a} = [a_1, a_2, \ldots, a_{n_S}]$ of the approximating polynomials.

End for
II. Project the polynomial coefficients k-times instances ahead to estimate the $ICDF(\rho)_{t+l+k}$
Several techniques can be used for that purpose, including ARMAX models [54] and polynomial curve fitting using Least Squares. At this point one can also incorporate discrete time filters to smooth the time-series signal. This step involves the selection of an appropriate model structure (among a set of candidate models). Here, an Autoregressive (AR)-based system identification technique is adopted:

II.1 Perform system identification of the approximating polynomial coefficients of the *ICDF* by a discrete model of the form:

$$a_i(t) = c_1 a_i(t-1) + c_2 a_i(t-2) + \cdots + c_{n-k} a_i(t-n_k)$$

The above expression relates a_i at time t to a finite number of past values $a_i(t-n_k)$. At this point it should be noted that, if necessary, a smoothing filter can be applied to the a_i time series. In this work, a second order (in a window of 5 sample points), Savitzky Golay filter [55] is used, which is a time varying FIR filter. Savitzky–Golay smoothing filters (also called digital smoothing polynomial filters or least squares smoothing filters) are typically used to "smooth out" a noisy signal whose frequency span (without noise) is large. Such filters are optimal in the sense that they minimize the least-squares error in fitting a polynomial to each frame of noisy data. The estimation of the vector of parameters $\Theta^T = [c_1, c_2, \ldots, c_{n-k}]$ at time t is carried out using least squares (LS) over the "raw" or smoothed data at time instances $t = n, n+1, \ldots n+l$.

II.2 Use the derived model to project a_i to time $t+l+k$.
II.3 Use the projected values to lift to one (or more) consistent microscopic distributions of positions $\mathbf{x}$.

End While

It is important to note that one must integrate the microscopic rules for some time before estimating the time derivative of the continuum variables. This allows higher moments of the continuum description to become slaved to the statistics of interest.

5. Simulation results

Simulation results were obtained using 2000 cells (N_{cells}), $amp = 15$, $g = 5$, $t_e = 0.1$, $t_a = 20$, $v = 0.002\ m/s$ and $dt = 0.01$ as the Monte Carlo time step. The chemoattractant profile (see Fig. 2) is a Gaussian-like function given by $S = \frac{1}{\sqrt{\pi}} \exp^{-\frac{(x-6)^2}{2}}$.

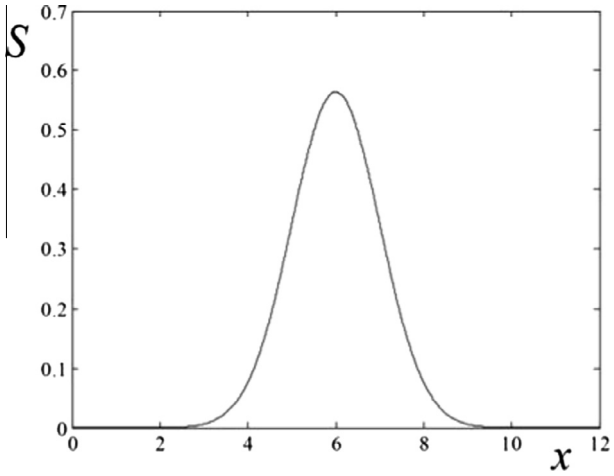

Fig. 2. The chemo-attractant distribution.

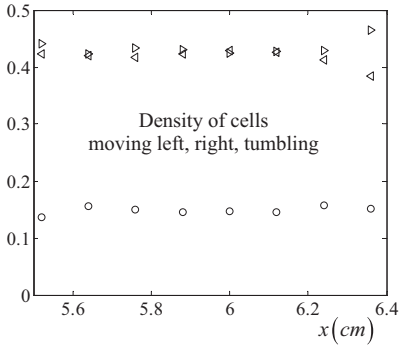

Fig. 3. Distributions of bacteria: running left ◁, right ▷ and tumble o.

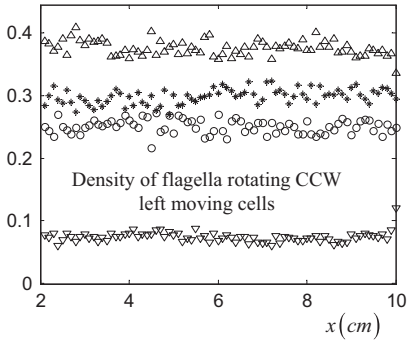

Fig. 4. Flagella rotating CCW Distributions for the left moving cells; △ correspond to 4, ∗ to 3, o to 5 and ▽ to 6 CCW rotating flagella.

Fig. 3 illustrates the distributions of cells moving left, right and tumbling at equilibrium; they are derived by averaging over time from $t = 10000$ to $t = 15000$ time steps using a time horizon of $T = 100$ time steps (i.e. averaging over a total of 50 samples) and over the total distribution of positions. The initial condition was a uniform (flat) profile of positions. As it is shown, simulation results are consistent with those dictated by the voting hypothesis giving a probability of 0.88 that the cell is running either left or right. Fig. 4 depicts the distributions of the number of left moving flagella that rotate CCW and Fig. 5 the distributions of the flagella that rotate CCW for the tumbling cells. Simulation results are consistent with those dictated by the binomial probability mass functions. All these variables are internal detailed variables with their own dynamics.

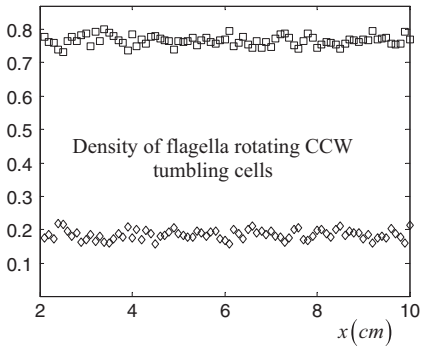

Fig. 5. Flagella rotating CCW Distributions for the tumbling cells.

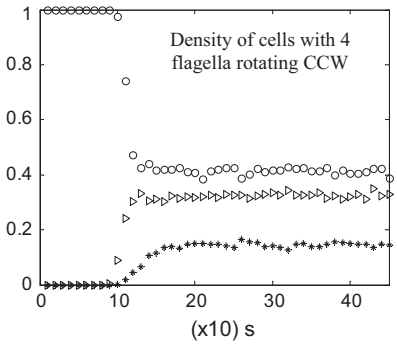

Fig. 6. Quick dynamic slaving of "internal" variables. When lifting, we set $u_1 = u_2 = 0$ for all cells and that cells are running right with 4 flagella CCW: time evolution of number of cells running right (o), running cells (either left or right) having 4 flagella rotating CCW (▷), tumbling cells (∗).

For the problem under study the analysis is based on the hypothesis that a macroscopic model exists and closes for the microscopic distribution of bacteria positions $\rho(\mathbf{x}, t)$. This implies that all the other "internal" variables, such as number of flagella rotating CCW or CW, the excitation u_1 and adaptation u_2 signals, become quickly slaved to the spatial distribution (they evolve towards a "slow manifold" parameterized by the underlying microscopic distribution). Computational results corroborating this closure assumption can be seen in Fig. 6, which illustrates the relatively fast evolution of the "internal" variables. Here the state of the "bacteria" are initialized by setting $u_1 = u_2 = 0$ for all cells, running right with 4 flagella rotating CCW. Fig. 7 shows the evolution of the internal variable u_2.

For the coarse-projective integration, the approximation of the *ICDF* is performed with linear extrapolation in time, $l = 10$ and $k = 10$ using eight orthogonal basis functions until $t = 6000$ time steps and then $l = 10,\ k = 20$ until $t = 25000$ time steps, while after each lifting, a period of $5T$ is used as "healing" of the errors made in the lifting step before the acquisition of the training data. In Fig. 8, are given the evolution of both normal integration (blue lines) and coarse-projective integration (red lines) for the density and cumulative distribution functions.

The coarse-grained steady-state and stability analysis was performed using two different methods, namely (i) through Newton–Raphson's method and (ii) through matrix-free methods.

Newton–Raphson's method was applied in terms of the mean value and the variance of a normal distribution function. Under this assumption, lifting was done with the inverse normal distribution. The fixed point was computed through the coarse timestepper with $T = 400$ time steps and 10 realizations of cell distributions. Upon convergence of the Newton–Raphson (to a residual of $O(10^{-3})$, for perturbations $\varepsilon \sim 5 \times 10^{-2}$) the mean value was found to be ~ 6.005 and the final variance ~ 0.13. Upon convergence, the Jacobian was found to be $A = \begin{pmatrix} 0.90521 & -0.0010 \\ -0.012 & 0.8423 \end{pmatrix}$. The eigenvalues of the 2×2 matrix are: $\lambda_1 \sim 0.910$ and $\lambda_1 \sim 0.845$. Looking at the Jacobian, it is observed that the interaction of the non-diagonal terms is very small when compared with the diagonal ones. This is an indication that near the fixed point the particular system can be considered for practical purposes uncoupled.

Newton-GMRES was also used to compute the steady states with respect to the coefficients of the basis functions and Arnoldi's method has been employed to perform stability analysis by performing on the actual microscopic distribution. These matrix-free algorithms were wrapped around the microscopic simulator averaging over 20 realizations of cells distributions. For the approximation of the inverse cumulative distribution function eight orthogonal basis functions were used.

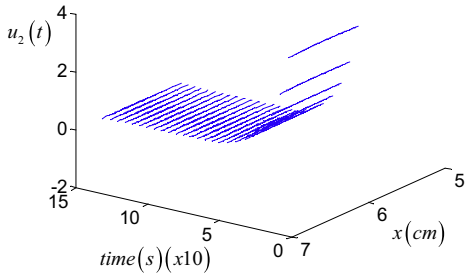

Fig. 7. Quick dynamic slaving of "internal" variables: time evolution of u_2.

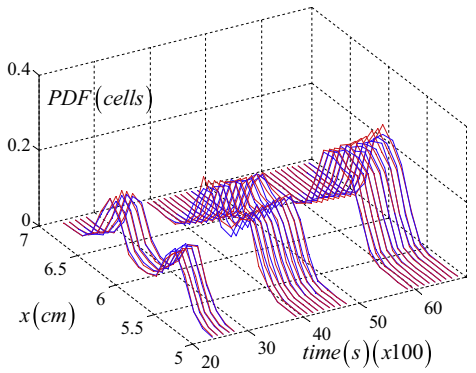

Fig. 8. Coarse projection (5 healing, $m = 5$ acquisition, $k = 10$ projection and 8 basis functions till $t = 6000$ time steps and then 5 healing, 5 acquisition, $k = 30$ projection and 8 basis functions till $t = 25000$ time steps). Blue lines correspond to temporal simulations while red ones to coarse projected simulation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The three leading eigenpairs with a residual of $O(10^{-4})$ computed with Arnoldi's method were $\lambda_1 \sim 0.904$, $\lambda_2 \sim 0.848$, $\lambda_3 \sim 0.571$. The first two leading eigenvalues and their corresponding eigenvectors as computed with Newton–Raphson and Arnoldi coincide for any practical means. Different choice of T (e.g., setting $T = 1000$ time steps) did not affect the results of the stability analysis.

The eigenvalues computed with the Arnoldi method show that there is a clear separation of time scales between the first two dominant macroscopic modes and the third one. This reveals that near the equilibrium and for the given chemoattractant profile, the coarse-grained dynamics evolve on a two dimensional manifold that can be in principle parameterized by the first two moments of the underlying microscopic distribution, namely the mean and the variance of the spatial distribution.

6. Discussion and conclusions

The equation-free approach [40] is a computational framework that provides a systematic approach for analyzing the parametric behavior of complex/ multiscale simulators much more efficiently than simply simulating forward in time. Acceleration of simulations in time, regime and stability computations, as well as continuation and numerical bifurcation analysis and other important tasks such as rare-events analysis of the complex-emergent dynamics can be performed in a straightforward manner bypassing the explicit extraction of closures. In this work, it is demonstrated how this multiscale approach can be used to extract system-level information from a Monte-Carlo model simulating the locomotion of *E. coli*. Fixed point and stability analysis were performed using both "conventional" contraction maps such as Newton–Raphson and matrix-free iterative algorithms (Newton-GMRES for fixed point computations and Arnoldi eigensolver). Through the numerical analysis it was shown that near the equilibrium the coarse-grained dynamics evolve on a two dimensional manifold that can be defined as a function of the first two moments of the evolving spatial distribution. This result indicates that all other internal model variables get relatively fast functionals of the "slow" macroscopic variables. The separation between scales, i.e. the gap between the internal fast and the macroscopic variable is given by the eigenspectrum; this provides a measure of how fast the other "internal" model variables get slaved to the coarse-grained one. Such analysis can provide (computational) justification on the selection of good coarse-grained variables that can be used to describe the emergent dynamics.

This task, i.e. trying to find good macroscopic variables in order to extract good closures is of outmost importance and one of the open problems in the area. Towards this aim the integration of data mining techniques such as Diffusion maps [56] can further enhance the efficiency of the equation-free framework. Coarse projective integration was also demonstrated combining the concept of coarse-timestepping with system identification techniques. The illustrative one-dimensional in space stochastic model does not exhibit any striking nonlinear behavior such as appearance of critical points, phase transitions or blow-up solutions in finite time. Such phenomena including complex pattern formation have been observed in experimental-observed bacterial responses [57–65] and their dynamics have been approximated and analyzed using macroscopic level models [16–18,57–65]. The proposed framework appears promising as a tool to better understand such complex collective behavior through the systematic coarse-grained bifurcation and stability analysis for both steady state and periodically oscillatory solutions [45,51], the adaptive detection of the critical points [44] exploiting the detailed biological knowledge that is incorporated into state-of-the art-microscopic models.

Acknowledgments

The author would like to thank Prof. Yannis Kevrekidis (Dept. of Chemical and Biological Engineering, Dept. of Mathematics and PACM, Princeton University), Dr. C.W. Gear (Dept. of Chemical and Biological Engineering, Princeton University) and Prof. Hans Othmer (Dept. of Mathematics, University of Minnesota) for their collaboration, insightful comments and many fruitful and helpful discussions.

References

[1] T-M. Yi, Y. Huang, M.I. Simon, J. Doyle, Robust perfect adaptation in bacterial chemotaxis through integral feedback control, Proc. Nat. Acad. Sci. USA 97 (2000) 4649–4653.
[2] M. Ito, H. Xu, A.A. Guffanti, Y. Wei, L. Zvi, D.E. Clapham, T.A. Krulwich, The voltage-gated Na+ channel NaVBP has a role in motility, chemotaxis and pH homeostasis of an alkaliphilic Bacillus, Proc. Nat. Acad. Sci. USA 101 (2004) 10566–10571.
[3] R. Stocker, J.R. Seymour, A. Samadani, D.E. Hunt, M.F. Polz, Rapid chemotactic response enables marine bacteria to exploit ephemeral microscale nutrient patches, Proc. Nat. Acad. Sci. USA 105 (2008) 4209–4214.
[4] J.H. Long, A.C. Lammert, C.A. Pell, M. Kemp, J.A. Strother, H.C. Crenshaw, M.J. McHenry, A navigational primitive: biorobotic implementation of cycloptic helical klinotaxis in planar motion, IEEE J. Oceanic Eng. 29 (2004) 795–806.
[5] M. Lind, B. Deleuran, K. Thestrup-Pedersen, K. Soballe, Chemotaxis of human osteoblasts: effects of osteotropic growth factors, APMIS 103 (1995) 140–146.
[6] D.G. Poitout, Biomechanics and Biomaterials in Orthopedics, Springer, New York, 2004.
[7] R.M. Macnab, D.E. Koshland, The gradient-sensing mechanism in bacterial chemotaxis, Proc. Nat. Acad. Sci. USA 69 (1972) 2509–2512.
[8] D. Brown, H. Berg, Temporal stimulation of chemotaxis in *Escherichia coli*, Proc. Nat. Acad. Sci. USA 71 (1974) 1388–1392.
[9] H.C. Berg, Chemotaxis in bacteria, Annu. Rev. Biophys. Bioeng. 4 (1975) 119–136.
[10] J.S. Parkinson, Signal transduction schemes of bacteria, Cell 73 (1993) 857–871.
[11] P.A. Spiro, J.S. Parkinson, H.G. Othmer, A model of excitation and adaptation in bacterial chemotaxis, Proc. Nat. Acad. Sci. USA 94 (1997) 7263–7268.
[12] G.S. Waldo, B.M. Standish, J. Berendzen, T.C. Terwilliger, Rapid protein-folding assay using green fluorescent protein, Nat. Biotechnol. 17 (1999) 691–695.
[13] P. Cluzel, M. Surette, S. Leibler, An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells, Science 287 (2000) 1652–1655.
[14] K. Wuichet, I.B. Zhulin, Origins and diversification of a complex signal transduction system in prokaryotes, Sci. Signal. 3 (2010) ra50.
[15] C.S. Patlak, Random walk with persistence and external bias, Bull. Math. Biophys. 15 (1953) 311–338.
[16] E.F. Keller, L.A. Segel, Initiation of slime mold aggregation viewed as an instability, J. Theor. Biol. 26 (1970) 399–415.
[17] E.F. Keller, L.A. Segel, Model for chemotaxis, J. Theor. Biol. 30 (1971) 225–234.
[18] E.F. Keller, L.A. Segel, Traveling bands of chemotactic bacteria: a theoretical analysis, J. Theor. Biol. 30 (1971) 377–380.
[19] R. Lapidus, R. Schiller, A mathematical model for bacterial chemotaxis, Biophys. J. 14 (1974) 825–834.
[20] P. Biler, Local and global solvability of some parabolic systems modeling chemotaxis, Adv. Math. Sci. Appl. 8 (1998) 715–743.
[21] Y. Dolak, C. Schmeiser, The Keller–Segel model with logistic sensitivity function and small diffusivity, SIAM J. Appl. Math. 66 (2005) 286–308.
[22] W. Alt, Biased random walk model for chemotaxis and related diffusion approximation, J. Math. Biol. 9 (1980) 147–177.
[23] M. Rivero, R. Tranquillo, H. Buettner, D. Lauffenburger, Transport models for chemotactic cell populations based on individual cell behavior, Chem. Eng. Sci. 44 (1989) 2881–2897.
[24] H.G. Othmer, A. Stevens, Aggregation blowup and collapse: the ABC's of taxis in reinforced random walks, SIAM J. Appl. Math. 57 (1997) 1044–1081.
[25] A. Stevens, The derivation of chemotaxis-equations as limit dynamics of moderately interacting stochastic many particle systems, SIAM J. Appl. Math. 61 (2000) 183–212.
[26] T. Hillen, H.G. Othmer, The diffusion limit of transport equations derived from velocity jump processes, SIAM J. Appl. Math. 61 (2000) 751–775.
[27] K.J. Painter, Complex spatial patterns in a hybrid chemotaxis reaction-diffusion model, J. Math. Biol. 41 (2000) 285–314.
[28] D. Horstmann, From 1970 until present: the Keller–Segel model in chemotaxis and its consequences, Jahresberichte DMV 105 (2003) 103–165.
[29] M.J. Tindall, P.K. Maini, S.L. Porter, J.P. Armitage, Overview of mathematical approaches used to model bacterial chemotaxis II: bacterial populations, Bull. Math. Biol. 70 (2008) 1570–1607.
[30] T. Hillen, K.J. Painter, A user's guide to PDE models for chemotaxis, J. Math. Biol. 58 (2009) 183–217.
[31] D. Wu, Signaling mechanisms for regulation of chemotaxis, Cell Res. 15 (2005) 52–56.
[32] B.M. Friedrich, F. Juelicher, Chemotaxis of sperm cells, Proc. Nat. Acad. Sci. USA 104 (2007) 13256–13261.
[33] M.J. Schnitzer, Theory of continuum random walks and application to chemotaxis, Phys. Rev. E 48 (1993) 2553–2568.
[34] H.G. Othmer, P. Schaap, Oscillatory cAMP signaling in the development of Dictyostelium discoideum, Comments Theor. Biol. 5 (1998) 175–282.
[35] P.H. Chanavis, Nonlinear mean field Fokker–Planck equations. Application to the chemotaxis of biological populations, Eur. Phys. J. B 62 (2008) 179–208.
[36] C.W. Gear, I.G. Kevrekidis, C. Theodoropoulos, Coarse integration/bifurcation analysis via microscopic simulators: micro-Galerkin methods, Comput. Chem. Eng. 26 (2001) 941–963.
[37] A. Makeev, D. Maroudas, I.G. Kevrekidis, Coarse stability and bifurcation analysis using stochastic simulators: kinetic Monte Carlo examples, J. Chem. Phys. 116 (2002) 10083–10091.
[38] O. Runborg, C. Theodoropoulos, I.G. Kevrekidis, Effective bifurcation analysis: a time-stepper-based approach, Nonlinearity 15 (2002) 491–511.
[39] I.G. Kevrekidis, C.W. Gear, G. Hummer, Equation-free: the computer-assisted analysis of complex multiscale systems, AI.Ch.E.J. 50 (2004) 1346–1354.

[40] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, C. Theodoropoulos, Equation-free coarse-grained multiscale computation: enabling microscopic simulators to perform system-level tasks, Commun. Math. Sci. 1 (2003) 715–762.
[41] C.I. Siettos, M. Graham, I.G. Kevrekidis, Coarse Brownian dynamics for nematic liquid crystals: bifurcation diagrams via stochastic simulation, J. Chem. Phys. 118 (2003) 10149–10156.
[42] J. Mooller, O. Runborg, P.G. Kevrekidis, K. Lust, I.G. Kevrekidis, Equation-free, effective computation for discrete systems: a time stepper based approach, Int. J. Bifur. Chaos 15 (2005) 975–996.
[43] S.J. Moon, R. Ghanem, I.G. Kevrekidis, Coarse graining the dynamics of coupled oscillators, Phys. Rev. Lett. 96 (2006) 144101–144104.
[44] C.I. Siettos, R. Rico-Martinez, I.G. Kevrekidis, A systems-based approach to multiscale computation: equation-free detection of coarse-grained bifurcations, Comput. Chem. Eng. 30 (2006) 1632–1642.
[45] L. Russo, C.I. Siettos, I.G. Kevrekidis, Reduced computations for nematic-liquid crystals: a timestepper approach for systems with continuous symmetries, J. Non-Newtonian Fluid Mech. 146 (2007) 51–58.
[46] S. Setayeshgar, C.W. Gear, H.G. Othmer, I.G. Kevrekidis, Application of coarse integration to bacterial chemotaxis, Multiscale Model. Simul. 4 (2005) 307–327.
[47] R. Erban, I.G. Kevrekidis, H.G. Othmer, An equation-free computational approach for extracting population-level behavior from individual-based models of biological dispersal, Physica D 215 (2006) 1–24.
[48] C.T. Kelley, Iterative Methods for Linear and Nonlinear Equations, SIAM, Philadelphia, 1995.
[49] G.M. Shroff, H.B. Keller, Stabilization of unstable procedures – the recursive projection method, SIAM J. Numer. Anal. 30 (1993) 1099–1120.
[50] K. Lust, D. Roose, A. Spence, A.R. Champneys, An adaptive Newton–Picard algorithm with subspace iteration for computing periodic solutions, SIAM J. Sci. Comput. 19 (1998) 1188–1209.
[51] M. Kavousanakis, L. Russo, C.I. Siettos, A.G. Boudouvis, G.C. Georgiou, A timestepper approach for the systematic bifurcation and stability analysis of polymer extrusion dynamics, J. Non-Newtonian Fluid Mech. 151 (2008) 59–68.
[52] Y. Saad, Numerical Methods for Large Eigenvalue Problems, Manchester University Press, Oxford-Manchester, 1992.
[53] K.N. Christodoulou, L.E. Scriven, Finding leading modes of a viscous free surface low: an asymmetric generalized problem, J. Non-Newtonian Fluid Mech. 3 (1998) 355–406.
[54] L. Ljung, System Identification: Theory for the User, PrenticeHall, Englewood Cliffs, 1987.
[55] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least-squares procedures, Anal. Chem. 36 (1964) 1627–1639.
[56] A. Kolpas, J. Moehlis, I.G. Kevrekidis, Coarse-grained analysis of stochasticity-induced switching between collective motion states, Proc. Nat. Acad. Sci. USA 104 (2007) 5931–5935.
[57] J. Adler, B. Templeton, The effect of environmental conditions on the motility of *Escherichia coli*, J. Gen. Microbiol. 46 (1967) 175–184.
[58] E. Budrene, H. Berg, Complex patterns formed by motile cells of *Escherichia coli*, Nature 349 (1991) 630–633.
[59] E. Budrene, H. Berg, Dynamics of formation of symmetrical patterns by chemotactic bacteria, Nature 376 (1995) 49–53.
[60] M.P. Brenner, L.S. Levitov, E.O. Budrene, Physical mechanisms for chemotactic pattern formation by bacteria, Biophys. J. 74 (1998) 1677–1693.
[61] M.R. Myerscough, P.K. Maini, K.J. Painter, Pattern formation in a generalized chemotactic model, Bull. Math. Biol. 60 (1998) 1–26.
[62] A.A. Polezhaev, R.A. Pashkov, A.I. Lobanov, I.B. Petrov, Spatial patterns formed by chemotactic bacteria *Escherichia coli*, Int. J. Dev. Biol. 50 (1998) 309–314.
[63] F. Michon, L. Forest, E. Collomb, J. Demongeot, D. Dhouailly, BMP2 and BMP7 play antagonistic roles in feather induction, Development 135 (2008) 2797–2805.
[64] H.G. Othmer, K. Painter, D. Umulis, C. Xue, The intersection of theory and application in elucidating pattern formation in developmental biology, Math. Model. Nat. Phenom. 4 (2009) 3–82.
[65] S. Payne, B. Li, Y. Cao, D. Schaeffer, M.D. Ryser, L. Youa, Temporal control of self-organized pattern formation without morphogen gradients in bacteria, Mol. Syst. Biol. 9 (2013) 697.