# Revisiting moment closure methods with population models

Davin Lunz, J Frédéric Bonnans, Jakob Ruess

# Revisiting moment closure methods with population models

Davin Lunz*[1,2], J. Frédéric Bonnans[3], and Jakob Ruess[1,2]

[1]Inria Paris, 2 rue Simone Iff, 75012 Paris, France
[2]Institut Pasteur, 28 rue du Docteur Roux, 75015 Paris, France
[3]Université Paris-Saclay, CNRS, CentraleSupélec, Inria, Laboratory of signals and systems, 91190, Gif-sur-Yvette, France

December 13, 2021

**Abstract**

Stochastic chemical kinetics at the single-cell level give rise to heterogeneous populations of cells even when all individuals are genetically identical. This heterogeneity can lead to nonuniform behaviour within populations, including different growth characteristics, cell-fate dynamics, and response to stimuli. Ultimately, these diverse behaviours lead to intricate population dynamics that are inherently multiscale: the population composition evolves based on population-level processes that interact with stochastically distributed single-cell states. Therefore, descriptions that account for this heterogeneity are essential to accurately model and control chemical processes. However, for real-world systems such models are computationally expensive to simulate, which can make optimisation problems, such as optimal control or parameter inference, prohibitively challenging. Here, we consider a class of multiscale population models that incorporate population-level mechanisms while remaining faithful to the underlying stochasticity at the single-cell level and the interplay between these two scales. To address the complexity, we study an order-reduction approximations based on the distribution moments. Since previous moment closure work has focused on the single-cell kinetics, extending these techniques to populations models prompts us to revisit old observations as well as tackle new challenges. In this extended multiscale context, we encounter the previously established observation that the simplest closure techniques can lead to non-physical system trajectories. Despite their poor performance in some systems, we provide an example where these simple closures outperform more sophisticated closure methods in accurately, efficiently, and robustly solving the problem of optimal control of bioproduction in a microbial consortium model.

## 1 Introduction

Stochastic chemical kinetics lead to cell-to-cell variability within an isogenic population of cells. This can manifest in heterogeneous cell behaviour, such as response to stimuli or cell-fate decisions [2, 6, 14, 39]. Despite a thoroughly mechanistic understanding of the behaviour of individual cells and communities [40] many population models, particularly for the purposes of control, neglect this heterogeneity due to the complexity it introduces [12, 13]. When this nonuniformity is ignored the resulting dynamics and optimal controls can deviate significantly from the ideal performance [23].

Multiscale models that describe evolving populations of stochastic individual cells pose significant analytical and numerical difficulty, which is exacerbated when trying to optimise such systems. This motivates order-reduction techniques that can mitigate the complexity while retaining sufficient model accuracy. Moment-closure techniques have been a popular order-reduction approach that reduce computational complexity although their use for system optimisation has been limited. For instance, moment-closure techniques are well studied in the chemical kinetics literature [9, 16, 28, 33, 34, 35, 37, 38, 43, and references therein]. Moment closures have also been studied in the context of population balance equations used in aerosol modeling to resolve the evolution of particle distributions under an extensive array of physical phenomena such as evaporation, nucleation, growth, coagulation, and complex mixing [24, 25, 26, 30, 44].

In this work, we focus on heterogeneous population dynamics and their optimisation. Our contribution extends these previous studies in several ways. The models studied here are designed for growing populations of cells that undergo population-level processes while remaining faithful to the underlying stochasticity of single-cell chemical kinetics. This multiscale model is a significant departure from the single-cell perspective [22]. Moreover, much

*davin.lunz@inria.fr

previous moment closure work has focused exclusively on simulating system evolution, whereas we consider the optimisation of such systems. We demonstrate that some techniques that successfully simulate system evolution do not provide a robust basis for system optimisation. This emphasises the need for novel moment closure methods in this multiscale context that may be used for system optimisation.

Here, we study a class of PDE models and seek reduced-order models that preserve the heterogeneity in the original models. We approximate the moment hierarchy of the solution, which reduces a PDE to a system of ODEs by multiplying the governing equation by monomial of the state and integrating over the state space. This ODE system representation preserves the influence of the heterogeneity while significantly reducing the complexity by enveloping the state space. In general, this induces an infinite hierarchy of coupled moment equations, as the dynamics of each moment depend on higher-order moments. For practical purposes, such as numerical simulation, this system must be truncated at finite order and the higher-order terms beyond the truncation order must be approximated by a closure rule. We refer to the resulting finite system of moment equations as the moment-closure approximation. In contrast to previous work, where the same class of PDE models are studied alongside moment-closure approximations up to first order [23], we study moment closures of arbitrarily high order. As we will demonstrate, retaining orders above the first allows us to preserve heterogeneous features not accounted for in the lower-order (homogeneous) models.

In this multiscale setting, we encounter the well-established result that the simple moment-closure technique of approximating the truncated moments by zero can lead to non-physical dynamics while more sophisticated moment-closure approaches can successfully reproduce the exact dynamics. Somewhat counter-intuitively, we study the optimal control of a popular bioproduction system and demonstrate that the sophisticated moment closures fail where the simple moment closure provides robust convergence to give excellent control performance. This example suggests that the naive approach ought not be abandoned too hastily.

In some previous work, reaction propensities are assumed to be of polynomial form so as to provide a self-contained hierarchy of moment equations and avoid non-trivial approximation that introduces both complexity and an additional source of error. However, such simplifying assumptions may limit the scope and applicability of previous methods, leaving important applications beyond the reach of the existing methodology. For example, some non-polynomial functional forms have biological relevance, such as the Hill functional form, which is ubiquitous in biochemistry, with applications ranging from pharmacology to enzyme kinetics and gene expression models.

This problem has been previously addressed by either assuming rational functions [27] or Taylor expanding nonlinear functions [1, 4, 9, 15, 16, 17]. However, we show that, in many practical scenarios, these previous approaches are either infeasible or may produce less robust optimisation convergence. In this work, we derive the moment hierarchies using systematic polynomial approximations of arbitrary nonlinear forms. This approach does not assume that the propensities are rational functions, and provides a polynomial approximation that, as opposed to the Taylor expansion, which is only locally accurate, provides an approximation that covers the state space more broadly. The Weierstrass theorem [8, Ch. 6] (that any continuous function on a closed interval can be uniformly approximated with arbitrary precision by a polynomial function) suggests that polynomial approximations are a promising direction. Nevertheless, choosing an appropriate approximation can be challenging: we provide an example in which the closeness of the polynomial approximation is less important than other features, such as preserving monotonicity of the original function.

The remainder of the paper is organised as follows. In section 2 we introduce a modular class of population models with dynamics grounded in stochastic individual members. In section 3 we derive an associated hierarchy of approximate moment equations. In section 4 we discuss a selection of moment-closure techniques to limit the hierarchy to a finite system. We then study three examples. In sections 5 and 6, we study birth–death–growth and growth–fragmentation models, respectively, where simple moment closures do not perform adequately. In section 7, we study a microbial consortium model where only the simplest moment closure works to solve the bioproduction optimal control problem. Finally, we discuss the implications of our results and future directions in section 8.

## 2 Heterogeneous and modular model class

We begin by considering a collection of *continuum* quantities within each cell, such as levels of a protein of interest or a transcription factor. We denote the $j$th quantity $x_j$ and the vector of the entire collection $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d_+$. We then consider distinct *discrete* classes of cells, with cells in a given class being indistinguishable. Cell classes may represent different genetic constructs or simply distinct states of an identical construct. Our interest is in the dynamics of populations of large numbers of cells, and the primary object we study is the (expected) population density of each class over the continuum quantities. Importantly, heterogeneity at the single-cell level depends on the stochastic chemical kinetics, not the population size, and therefore persists in large populations [22]. We denote the population density of cells of class $k$ over the continuum quantities $\boldsymbol{x}$ at

time $t$ by $p_k(\boldsymbol{x}, t)$. We are interested in models governing the evolution of the collection of population densities in all states, $\{p_k(\boldsymbol{x}, t)\}_k$, of the form [23]

$$\frac{\partial}{\partial t} p_k(\boldsymbol{x}, t) = \text{Population growth} + \text{Population dilution} + \text{Discrete dynamics} + \text{Continuum dynamics}, \tag{1a}$$

$$p_k(\boldsymbol{x}, 0) = p_{0k}(\boldsymbol{x}), \tag{1b}$$

for $k = 1, \ldots, \mathcal{K}$. Each term on the right-hand side of (1) contributes to the rate of change of the density of cells of class $k$ with internal state $\boldsymbol{x}$, and thus each may depend on both $k$ and $\boldsymbol{x}$, as well as time $t$. We now detail a collection of examples for each component on the right-hand side of (1).

## 2.1 Population growth

The population growth refers to the processes by which the cells proliferate. A simple and classical example is exponential growth. If the continuum quantities $\boldsymbol{x}$ represent *concentrations* that remain essentially constant over cell division events [29], then the rate of change of cell density due to growth may be governed by a law of the form

$$\frac{\partial}{\partial t} p_k(\boldsymbol{x}, t) = g_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) + (\text{Non-growth sources}), \tag{2}$$

where the growth rate $g_k(\boldsymbol{x}, t)$ exhibits dependence on the states $k$ and $\boldsymbol{x}$ as well as time $t$.

An alternative is to consider the continuum states $\boldsymbol{x}$ to represent (scaled) copy numbers. In this case, cell division events cause the state $\boldsymbol{x}$ to be partitioned and shared between the two daughter cells, which is described by a fragmentation process. We define the set of states in $\mathbb{R}_+^d$ with each component no greater than each corresponding component of $\boldsymbol{x}$, which we denote $\mathbf{0} \leq \boldsymbol{x} \leq \hat{\boldsymbol{x}} := [0, \hat{x}_1] \times \cdots \times [0, \hat{x}_d]$ where $\mathbf{0} = (0, \ldots, 0) \in \mathbb{N}_0^d$ is the $d$-dimensional vector of zeros. Similarly, $\hat{\boldsymbol{x}} \geq \boldsymbol{x}$ denotes the set where each component is no less than its counterpart in $\boldsymbol{x}$. The rate of change of cell density is then governed by

$$\frac{\partial}{\partial t} p_k(\boldsymbol{x}, t) = -g_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) + 2 \int_{\hat{\boldsymbol{x}} \geq \boldsymbol{x}} g_k(\hat{\boldsymbol{x}}, t) p_k(\hat{\boldsymbol{x}}, t) \theta_k(\boldsymbol{x} \mid \hat{\boldsymbol{x}}) \, \mathrm{d}\hat{\boldsymbol{x}} + (\text{Non-growth sources}), \tag{3}$$

where the growth rate is still $g_k(\boldsymbol{x}, t)$ and $\theta_k(\boldsymbol{x} \mid \hat{\boldsymbol{x}})$ represents the probability density that, for a given daughter cell, a mother cell in state $\hat{\boldsymbol{x}}$ divides to produce a daughter in state $\boldsymbol{x}$. The factor of two reflects the cell division of the mother into two daughter cells. We require that $\theta_k$ be a symmetric probability density:

$$\int_{\mathbf{0} \leq \boldsymbol{x} \leq \hat{\boldsymbol{x}}} \theta_k(\boldsymbol{x} \mid \hat{\boldsymbol{x}}) \, \mathrm{d}\boldsymbol{x} = 1, \qquad \theta_k(\boldsymbol{x} \mid \hat{\boldsymbol{x}}) = \theta_k(\hat{\boldsymbol{x}} - \boldsymbol{x} \mid \hat{\boldsymbol{x}}). \tag{4}$$

Assumptions (4) guarantee that the fragmentation contributes an increase in the total number of cells per unit time while conserving the total protein copy number. To derive the moment hierarchy, we will require the additional assumption of self-similarity

$$\theta_k(\boldsymbol{x} \mid \hat{\boldsymbol{x}}) = \hat{\boldsymbol{x}}^{-\mathbf{1}} \Theta_k(\boldsymbol{x} \oslash \hat{\boldsymbol{x}}), \tag{5}$$

for some probability density $\Theta_k : \mathbf{0} \leq \boldsymbol{y} \leq \mathbf{1} \to \mathbb{R}_+$ where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{N}_0^d$ is the $d$-dimensional vector of ones and we introduce the Hadamard product and division, defined respectively by

$$\boldsymbol{x} \odot \boldsymbol{y} = (x_1 y_1, \ldots, x_d y_d), \qquad \boldsymbol{x} \oslash \boldsymbol{y} = \left( \frac{x_1}{y_1}, \ldots, \frac{x_d}{y_d} \right). \tag{6}$$

To see that $\Theta_k$ is a probability density, note that, by using the change of variables $\boldsymbol{y} := \boldsymbol{x} \oslash \hat{\boldsymbol{x}}$, the density condition in (4) takes the form

$$1 = \int_{\mathbf{0} \leq \boldsymbol{x} \leq \hat{\boldsymbol{x}}} \theta_k(\boldsymbol{x} \mid \hat{\boldsymbol{x}}) \, \mathrm{d}\boldsymbol{x} = \hat{\boldsymbol{x}}^{\mathbf{1}} \int_{\mathbf{0} \leq \boldsymbol{y} \leq \mathbf{1}} \theta_k(\hat{\boldsymbol{x}} \odot \boldsymbol{y} \mid \hat{\boldsymbol{x}}) \, \mathrm{d}\boldsymbol{y} = \int_{\mathbf{0} \leq \boldsymbol{y} \leq \mathbf{1}} \Theta_k(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}. \tag{7}$$

For $\mathbf{0} \leq \boldsymbol{y} \leq \mathbf{1}$, the symmetry condition in (4) takes the form

$$\Theta_k(\boldsymbol{y}) = \Theta_k(\mathbf{1} - \boldsymbol{y}). \tag{8}$$

## 2.2 Population dilution

In many practical scenarios, the population is diluted to preserve healthy growth conditions. For example, bioreactors have modes of continuous operation that allow fresh media to be introduced into the bioreactor, and the existing solution diluted to remove toxins and biproducts as well as maintain an equilibrium population density. This allows longer operation, reducing the burden of frequently resetting the bioreactor, and simplifies the modeling as it allows the colony and nutrients to remain in a steady growth phase [32].

We consider the turbidostat operation mode, where the dilution is tuned in real time so as to maintain a constant optical density. Taking the optical density as a proxy for the population density, we model this contribution via the dilution rate, $\Lambda(t)$,

$$\frac{\partial}{\partial t}p_k(\boldsymbol{x},t) = -\Lambda(t)p_k(\boldsymbol{x},t) + (\text{Non-dilution sources}), \tag{9}$$

where $\Lambda(t)$ does not depend on $k$ or $\boldsymbol{x}$, but does depend on the population density as follows. Integrating (1) over the state space $\boldsymbol{x} \in \mathbb{R}^d_+$ and summing over all discrete states $k = 1, \ldots, \mathcal{K}$, the discrete and continuum dynamics cancel as they are *conservative*; describing how members of the population change their states $k$ and $\boldsymbol{x}$ without changing the total population mass. In this case, where all changes in population mass stem from the "Population growth" terms, it follows that

$$\Lambda(t) = \frac{\displaystyle\sum_{k=1}^{\mathcal{K}} \int_{\mathbb{R}^d_+} [\text{Population growth}]_k(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x}}{\displaystyle\sum_{k=1}^{\mathcal{K}} \int_{\mathbb{R}^d_+} p_k(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x}}, \tag{10}$$

where we have shown the explicit dependence of the population growth contributions on the states $k$ and $\boldsymbol{x}$ and time $t$. The turbidostat dilution rate (10) is simply the mean growth rate, which is precisely the balance required to preserve a constant total population mass. Moreover, since the denominator in (10) is constant, we may assume, without loss of generality, that it is unity. This is equivalent to rescaling the population densities $p_k(\boldsymbol{x},t)$ by this constant value. This gives a simpler form for the dilution rate, namely

$$\Lambda(t) = \sum_{k=1}^{\mathcal{K}} \int_{\mathbb{R}^d_+} [\text{Population growth}]_k(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x}. \tag{11}$$

## 2.3 Discrete dynamics

Discrete dynamics refers to the change of state of a cell in class $k$ to some other class $\ell$, at a rate denoted by $c_{k\ell}(\boldsymbol{x},t)$, whereby

$$\frac{\partial}{\partial t}p_k(\boldsymbol{x},t) = -p_k(\boldsymbol{x},t)\sum_{\ell=1}^{\mathcal{K}} c_{k\ell}(\boldsymbol{x},t) + \sum_{\ell=1}^{\mathcal{K}} c_{\ell k}(\boldsymbol{x},t)p_\ell(\boldsymbol{x},t) + (\text{Non-discrete sources}). \tag{12}$$

The discrete states might represent the copy number of a species produced in numbers below what would justify a continuum approximation [21, 22] or a binary internal state such as governed by binding kinetics. In such cases the discrete dynamics describe stochastic single-cell chemical kinetics. In the current formulation, the copy-number paradigm for continuum species does not carry over to dividing discrete species, however, this is an immediate extension.

## 2.4 Continuum dynamics

Continuum dynamics refers to the change of state of a cell in state $\boldsymbol{x}$ to some other state $\boldsymbol{y}$. For example, the evolution of the continuum state $\boldsymbol{x}$ under a birth–death process with rate $r_k(\boldsymbol{x},t) \geq 0$ is described by the Fokker–Planck operator

$$\frac{\partial}{\partial t}p_k(\boldsymbol{x},t) = -\boldsymbol{e}_k \cdot \nabla\left[r_k(\boldsymbol{x},t)p_k(\boldsymbol{x},t)\right] + \frac{1}{2\Omega}\boldsymbol{e}_k^\top \nabla^2\left[r_k(\boldsymbol{x},t)p_k(\boldsymbol{x},t)\right]\boldsymbol{e}_k + (\text{other sources}), \tag{13}$$

where $\boldsymbol{e}_k \in \mathbb{Z}^d$ is a vector representing the number of reactant molecules minus the number of product molecules of each continuum species [22], $\nabla^2$ denotes the Hessian, and $\Omega$ is the characteristic system size [21]. The operator is associated with the zero normal-flux boundary condition

$$(\boldsymbol{e}_k \cdot \boldsymbol{n})\left\{r_k(\boldsymbol{x},t)p_k(\boldsymbol{x},t) - \frac{1}{2\Omega}\boldsymbol{e}_k \cdot \nabla\left[r_k(\boldsymbol{x},t)p_k(\boldsymbol{x},t)\right]\right\} = 0, \qquad \boldsymbol{x} \in \partial\mathbb{R}^d_+, \tag{14}$$

where $\boldsymbol{n}$ denotes the outward-facing normal to the boundary. There may be many such processes occurring simultaneously on the state $k$, in which case we sum over all the associated Fokker–Planck operators and boundary conditions. This operator may be written in divergence form, from which its conservative nature is apparent, as done in Appendix A.2.

In many biologically relevant scenarios, production in bursts provides a more realistic description of the underlying production mechanism [10, 20], whereby the continuum dynamics take the form

$$\begin{aligned}\frac{\partial}{\partial t}p_k(\boldsymbol{x},t) = -f_k(\boldsymbol{x},t)p_k(\boldsymbol{x},t) + \|\boldsymbol{e}_k\| \int_{\boldsymbol{x}-z\boldsymbol{e}_k\in\mathbb{R}^d_+} f_k(\boldsymbol{x}-z\boldsymbol{e}_k,t)p_k(\boldsymbol{x}-z\boldsymbol{e}_k,t)Q_k(\boldsymbol{x}-z\boldsymbol{e}_k,z\|\boldsymbol{e}_k\|)\,\mathrm{d}z \\ + (\text{other sources}),\end{aligned} \tag{15}$$

where $\boldsymbol{e}_k \in \mathbb{Z}^d$ again represents the number of reactant molecules minus the number of product molecules, $f_k(\boldsymbol{x},t)$ denotes the rate of production bursts, and $Q_k(\boldsymbol{x},y)$ denotes the probability density, characteristically of exponential form, that the production process in a cell of state $\boldsymbol{x}$ yields a state $\boldsymbol{x}+y\boldsymbol{e}_k/\|\boldsymbol{e}_k\|$. We thus require that, for all $\boldsymbol{x}$,

$$\int_0^\infty Q_k(\boldsymbol{x},y)\,\mathrm{d}y = 1. \tag{16}$$

Here too, if several such process occur simultaneously these are to be summed.

There are no boundary conditions to be imposed with the burst operator, however, the formulation (15) implicitly imposes boundary conditions. The present formulation assumes that all components of the vector $\boldsymbol{e}_k$ are non-negative. Were this not the case, the first term on the right-hand side of (15) would need to be modulated by a coefficient representing the fraction of valid jumps: $v_k(\boldsymbol{x}) := \|\boldsymbol{e}_k\| \int_{\boldsymbol{x}+z\boldsymbol{e}_k\in\mathbb{R}^d_+} Q_k(z\|\boldsymbol{e}_k\|)\,\mathrm{d}z$. This would disallow jumps that use "exhausted" continuum species (that is, jumps to states $\boldsymbol{x}$ with at least one negative component) and instead leave the state unchanged. This treatment can be handled in the present formulation by simply absorbing the coefficient $v_k(\boldsymbol{x})$ into the rate $f_k(\boldsymbol{x},t)$. A more physically relevant model would place the fraction $1-v_k(\boldsymbol{x})$ of invalid jumps at the boundary, however, for the sake of simplicity, we will not account for this.

# 3 Moment hierarchy

The analysis and optimal control of the modular PDE models introduced in section 2 is formidably challenging. The computational cost of their numerical simulation and optimisation quickly becomes prohibitive with increasing resolution and dimension, motivating dimension reduction techniques.

Taking the components introduced in section 2, we derive the associated contributions to the moment hierarchy. We introduce the (scalar) moments

$$X_k^{\boldsymbol{i}}(t) := \int_{\mathbb{R}^d_+} \boldsymbol{x}^{\boldsymbol{i}} p_k(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x}, \tag{17}$$

for the multi-index $\boldsymbol{i} = (i_1,\ldots,i_d) \in \mathbb{N}_0^d$, where $\boldsymbol{x}^{\boldsymbol{i}} = \prod_{j=1}^d x_j^{i_j} \in \mathbb{R}_+$ and we write $|\boldsymbol{i}| := \sum_{j=1}^d i_j$. The governing equation of each moment $X_k^{\boldsymbol{i}}(t)$ is found by multiplying the PDE governing $p_k(\boldsymbol{x},t)$ by $\boldsymbol{x}^{\boldsymbol{i}}$ and integrating over the state space. The details of these calculations appear in Appendix A, and in this section we outline the results.

To produce a closed hierarchy of moments, we approximate the non-polynomial functions of state on the right-hand side of (1) by a polynomial form. So for a function $g_k(\boldsymbol{x},t)$, we take the $N(g_k)$-degree approximation, denoted by

$$g_k(\boldsymbol{x},t) \approx \sum_{|\boldsymbol{j}|\le N(g_k)} g_{k,\boldsymbol{j}}(t)\boldsymbol{x}^{\boldsymbol{j}}, \tag{18}$$

where we use the multi-index notation $\boldsymbol{j} \in \mathbb{N}_0^d$ for which $\boldsymbol{x}^{\boldsymbol{j}} = x_1^{j_1}\cdots x_d^{j_d} \in \mathbb{R}$ is a scalar value. This step raises the question of how to select a polynomial approximation. Later, with application to a concrete example, we discuss important considerations in this choice.

### 3.1 Population growth

The growth law in the concentration paradigm (2) contributes to the moments via

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}X_k^{\boldsymbol{i}}(t) &= \int_{\mathbb{R}_+^d} \boldsymbol{x}^{\boldsymbol{i}} g_k(\boldsymbol{x},t)p_k(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x} + (\text{Non-growth sources}) \\ &\approx \sum_{|\boldsymbol{j}|\le N(g_k)} g_{k,\boldsymbol{j}}(t)\int_{\mathbb{R}_+^d} \boldsymbol{x}^{\boldsymbol{i}+\boldsymbol{j}} p_k(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x} + (\text{Non-growth sources}) \\ &= \sum_{|\boldsymbol{j}|\le N(g_k)} g_{k,\boldsymbol{j}}(t)X_k^{\boldsymbol{i}+\boldsymbol{j}}(t) + (\text{Non-growth sources}).\end{aligned} \tag{19}$$

In the copy-number paradigm, we show in Appendix A that the moment equations take the form

$$\frac{\mathrm{d}}{\mathrm{d}t}X_k^{\boldsymbol{i}}(t) = (2\Theta_k^{\boldsymbol{i}} - 1)\sum_{|\boldsymbol{j}|\le N(g_k)} g_{k,\boldsymbol{j}}(t)X_k^{\boldsymbol{i}+\boldsymbol{j}}(t) + (\text{Non-growth sources}), \tag{20}$$

for

$$\Theta_k^{\boldsymbol{i}} := \int_{\boldsymbol{0}\le\boldsymbol{y}\le\boldsymbol{1}} \boldsymbol{y}^{\boldsymbol{i}}\Theta_k(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{21}$$

Note from (7) that the zeroth-order contribution $\Theta_k^{\boldsymbol{0}} = 1$ for $\boldsymbol{0} = (0,\dots,0)$. This demonstrates how the moment hierarchy preserves the form of the total growth rate: for $\boldsymbol{i} = \boldsymbol{0}$, the right-hand sides of (19) and (20) are identical.

### 3.2 Population dilution

The turbidostat contribution (9), which maintains a constant population mass, is approximated by

$$\frac{\mathrm{d}}{\mathrm{d}t}X_k^{\boldsymbol{i}}(t) = -\Lambda(t)X_k^{\boldsymbol{i}}(t) + (\text{Non-dilution sources}), \tag{22}$$

where the turbidostat dilution rate $\Lambda(t)$ given in (11) may be expressed self-consistently as the sum of the growth contributions of zeroth order $\boldsymbol{i} = \boldsymbol{0}$. In both the concentration (19) and copy-number (20) paradigms the zeroth-order contributions are identical, therefore, in both cases,

$$\Lambda(t) = \sum_{k=1}^{\mathcal{K}}\sum_{|\boldsymbol{j}|\le N(g_k)} g_{k,\boldsymbol{j}}(t)X_k^{\boldsymbol{j}}(t). \tag{23}$$

It follows that, in turbidostat mode, the total population mass $\sum_{k=1}^{\mathcal{K}} X_k^{\boldsymbol{0}}(t)$ remains constant in the moment-closure approximation.

### 3.3 Discrete dynamics

The discrete dynamics (12) lend the contributions

$$\frac{\mathrm{d}}{\mathrm{d}t}X_k^{\boldsymbol{i}}(t) = -\sum_{\ell=1}^{\mathcal{K}}\sum_{|\boldsymbol{j}|\le N(c_{k\ell})} c_{k\ell,\boldsymbol{j}}(t)X_k^{\boldsymbol{i}+\boldsymbol{j}}(t) + \sum_{\ell=1}^{\mathcal{K}}\sum_{|\boldsymbol{j}|\le N(c_{\ell k})} c_{\ell k,\boldsymbol{j}}(t)X_\ell^{\boldsymbol{i}+\boldsymbol{j}}(t) + (\text{Non-discrete sources}). \tag{24}$$

Here too the conservation property of the discrete dynamics is preserved: summing (24) over $k$ we find that $\sum_{k=1}^{\mathcal{K}} X_k^{\boldsymbol{i}}(t)$ is changed only by non-discrete sources.

### 3.4 Continuum dynamics

In Appendix A we show that the contributions stemming from the Fokker–Planck terms (13) take the form

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}X_k^{\boldsymbol{i}}(t) \approx{}& \sum_{\ell=1}^{d} i_\ell e_{k,\ell}\sum_{|\boldsymbol{j}|\le N(r_k)} r_{k,\boldsymbol{j}}(t)X_k^{\boldsymbol{i}-\boldsymbol{v}_\ell+\boldsymbol{j}}(t) + \frac{1}{2\Omega}\sum_{\ell=1}^{d}\sum_{m=1}^{d} i_\ell(i_m-\delta_{\ell m})e_{k,\ell}e_{k,m}\sum_{|\boldsymbol{j}|\le N(r_k)} r_{k,\boldsymbol{j}}(t)X_k^{\boldsymbol{i}-\boldsymbol{v}_\ell-\boldsymbol{v}_m+\boldsymbol{j}}(t) \\ &+ (\text{other sources}),\end{aligned} \tag{25}$$

where $\delta_{\ell m}$ is the Kronecker Delta function, which takes the value one when $\ell = m$ and zero otherwise, and $\boldsymbol{v}_\ell$ denotes the vector of zeros except for the $\ell$th component whose value is one. The moment approximation (25) of the Fokker–Planck terms has no zeroth-order ($\boldsymbol{i} = \boldsymbol{0}$) contribution, thus preserving the conservation property that $X_k^{\boldsymbol{0}}(t)$ is changed only by non-continuum sources.

The moments for the bursty production (15) are calculated in Appendix A. There is similarly no zeroth-order contribution. For higher-order moments,

$$\frac{\mathrm{d}}{\mathrm{d}t} X_k^{\boldsymbol{i}}(t) = \sum_{\boldsymbol{j}<\boldsymbol{i}} b_{\boldsymbol{i},\boldsymbol{j}} \boldsymbol{e}_k^{\boldsymbol{i}-\boldsymbol{j}} \|\boldsymbol{e}_k\|^{-|\boldsymbol{i}-\boldsymbol{j}|} \int_{\mathbb{R}_+^d} \boldsymbol{x}^{\boldsymbol{j}} Q_{k,|\boldsymbol{i}-\boldsymbol{j}|}(\boldsymbol{x}) f_k(\boldsymbol{x},t) p_k(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x} + (\text{other sources}), \tag{26}$$

where the coefficients $b_{\boldsymbol{i},\boldsymbol{j}}$ are defined by (82), and we have defined, for $j \in \mathbb{N}_0$,

$$Q_{k,j}(\boldsymbol{x}) := \int_0^\infty y^j Q_k(\boldsymbol{x}, y)\,\mathrm{d}y. \tag{27}$$

Denoting the polynomial approximation

$$Q_{k,j}(\boldsymbol{x}) \approx \sum_{|\boldsymbol{\ell}| \le jN(Q_{k,j})} w_{k,j,\boldsymbol{\ell}} \boldsymbol{x}^{\boldsymbol{\ell}}, \tag{28}$$

we see that

$$\frac{\mathrm{d}}{\mathrm{d}t} X_k^{\boldsymbol{i}}(t) \approx \sum_{\boldsymbol{j}<\boldsymbol{i}} b_{\boldsymbol{i},\boldsymbol{j}} \boldsymbol{e}_k^{\boldsymbol{i}-\boldsymbol{j}} \|\boldsymbol{e}_k\|^{-|\boldsymbol{i}-\boldsymbol{j}|} \sum_{|\boldsymbol{\ell}| \le |\boldsymbol{i}-\boldsymbol{j}| N(q_k)} \sum_{|\boldsymbol{m}| \le N(f_k)} w_{k,|\boldsymbol{i}-\boldsymbol{j}|,\boldsymbol{\ell}} f_{k,\boldsymbol{m}}(t) X_k^{\boldsymbol{j}+\boldsymbol{\ell}+\boldsymbol{m}}(t) + (\text{other sources}). \tag{29}$$

The moment approximation (29) of the bursty production terms preserves the conservation property that $X_k^{\boldsymbol{0}}(t)$ is changed only by non-continuum sources (when $\boldsymbol{i} = \boldsymbol{0}$ the right-hand side of (25) vanishes because there are no multi-indices $\boldsymbol{j} < \boldsymbol{i}$).

It is instructive to consider the concrete example of the typical exponential jump kernel, namely,

$$Q_k(\boldsymbol{x}, y) = \frac{\mathrm{e}^{-y/q_k(\boldsymbol{x})}}{q_k(\boldsymbol{x})}, \tag{30}$$

for a (possibly state-dependent) mean burst size $q$. This satisfies (16), $Q_{k,0}(\boldsymbol{x}) = 1$, and gives, by integrating by parts, the recursive relation $Q_{k,j}(\boldsymbol{x}) = jq(\boldsymbol{x})Q_{k,j-1}(\boldsymbol{x})$ for $j \ge 1$. It follows that

$$Q_{k,j}(\boldsymbol{x}) = j! q_k(\boldsymbol{x})^j. \tag{31}$$

Taking the polynomial approximation of $q_k$, we then see that

$$Q_{k,j}(\boldsymbol{x}) \approx j! \left( \sum_{|\boldsymbol{\ell}| \le N(q_k)} q_{k,\boldsymbol{\ell}} \boldsymbol{x}^{\boldsymbol{\ell}} \right)^j = \sum_{|\boldsymbol{\ell}| \le jN(q_k)} w_{k,j,\boldsymbol{\ell}} \boldsymbol{x}^{\boldsymbol{\ell}}, \tag{32}$$

for some coefficients $w_{k,j,\boldsymbol{\ell}}$.

### 3.5 Initial conditions

The moments inherit their initial conditions from the moments of the initial density (1b), that is,

$$X_k^{\boldsymbol{i}}(0) = \int_{\mathbb{R}_+^d} \boldsymbol{x}^{\boldsymbol{i}} p_{0k}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \tag{33}$$

This concludes the derivation of the ODE components of the moment hierarchies from the components of the PDE model class.

## 4 Moment closures

Equipped with the infinite moment hierarchies, we must decide how to close the moment equations so as to be able to solve a finite collection numerically. We will discuss three closure schemes.

The simplest thing to do is to consider all higher-order moments to be zero: for some $N$, set

$$X_k^{\boldsymbol{i}}(t) = 0, \qquad \text{for all } k \text{ and all } |\boldsymbol{i}| \ge N. \tag{34}$$

While this is straightforward, it implicitly relies on the assumption that the higher-order moments do not have an important influence on the lower-order moments. This is a very strong assumption, especially in the context of nonlinear functions subject to polynomial approximation. Nevertheless, we will discover that, despite its shortcomings, the zero closure, as we will call it, can be surprisingly useful.

A more appealing idea, for distributions concentrated near their mean, is to assume that the *central* higher-order moments vanish, that is,

$$\int_0^\infty (\boldsymbol{x} - \boldsymbol{X}_k(t))^{\boldsymbol{i}}\, p_k(x,t)\,\mathrm{d}x, \qquad \text{for all } k \text{ and all } |\boldsymbol{i}| \geq N = 0, \tag{35}$$

where $\boldsymbol{X}_k(t)$ is the vector of means

$$\boldsymbol{X}_k(t) = \left( \frac{X_k^{\boldsymbol{v}_1}(t)}{X_k^{\boldsymbol{0}}(t)}, \ldots, \frac{X_k^{\boldsymbol{v}_d}(t)}{X_k^{\boldsymbol{0}}(t)} \right), \tag{36}$$

with $\boldsymbol{v}_i$ denoting the $d$-dimensional vector of zeros except the $i$th entry which is one. Performing the binomial expansion, the condition (35) gives the central closure, as we will call it. For example, in dimension $d = 1$, we obtain the formula

$$X_k^i(t) = -\sum_{j=0}^{i-1} \binom{i}{j} \left( -\frac{X_k^1(t)}{X_k^0(t)} \right)^{j-k} X_k^j(t), \qquad \text{for all } k \text{ and all } i \geq N. \tag{37}$$

Finally, we consider the zero-information closure [38]. At any point in time $t$, the distribution is assumed to be the distribution of maximum-entropy whose first $N$ moments coincide with the given lower-order moments $\{X^i(t)\}_{i=0}^{N-1}$. By solving the constrained optimisation problem of maximising the entropy subject to the constraint of given moments, we may reconstruct the distribution, and, in turn, we may calculate the higher-order moments. We will demonstrate that this technique can succeed where the previous closure methods fail, nevertheless, the reconstruction and subsequent moment calculation is nontrivial and costly, which makes them difficult to use in optimisation loops.

Armed with techniques to close the moment systems, we proceed to consider three examples: a birth–death single-cell model coupled to selective growth, a family of growth–fragmentation models, and a microbial consortium model. In the first two examples, we discuss simulating the moment equations using the three different closure schemes, while in the final example we seek to solve an optimal control problem of maximising bioproduction yield.

# 5 Birth–death–growth models

We begin with a birth–death process of a single species $X$, which models the single-cell reactions

$$\varnothing \quad \xrightarrow{\lambda} \quad \mathrm{X} \quad \xrightarrow{\mu} \quad \varnothing, \tag{38}$$

for the non-negative birth and death rates $\lambda$ and $\mu$, respectively, which may depend on the state.

In addition to the single-cell dynamics, there is a population-level selection due to a selective growth rate $G$ that may also depend on the state. Under the continuum scaling [22], the population density is governed by

$$\frac{\partial}{\partial t} p(x,t) = G(x)p(x,t) - \frac{\partial}{\partial x}\left[(\lambda(x) - \mu(x))p(x,t)\right] + \frac{1}{2\Omega}\frac{\partial^2}{\partial x^2}\left[(\lambda(x) + \mu(x))p(x,t)\right], \quad p(x,0) = p_0(x), \tag{39a}$$

for some initial distribution $p_0(x)$, and the zero normal flux boundary conditions

$$(\lambda(x) - \mu(x))p(x,t) - \frac{1}{2\Omega}\frac{\partial}{\partial x}[(\lambda(x) + \mu(x))p(x,t)] = 0, \qquad \text{at } x = 0 \text{ and as } x \to \infty. \tag{39b}$$

It will be convenient to choose the polynomial birth, death, and growth rates studied in Ref. [22] since analytical approximations are available in different limiting cases. These are given by

$$\lambda(x) = \Lambda x(1-x), \qquad \mu(x) = x, \qquad G(x) = gx, \tag{40}$$

where we assume $\Lambda > 1$ such that we are in the supercritical case (for $\Lambda \leq 1$, there is no state $x > 0$ that constitutes a metastable equilibrium for the single-cell stochastic dynamics).

It can be shown [22] that the large-time behaviour of the solution is to have a population distribution around a critical point $x = x_c$ and growing exponentially at a rate $G(x_c)$. Up to exponentially small corrections in the large system-size limit $\Omega \gg 1$, the distribution is

$$p(x,t) \sim \sqrt{\frac{a\Omega}{\pi}}\, \mathrm{e}^{G(x_c)t - a\Omega(x-x_c)^2}, \tag{41a}$$

where $x_c$ is the unique (positive) point where the birth and death rates match, namely $\lambda(x_c) = \mu(x_c)$ at $x_c = 1 - 1/\Lambda$, and

$$a := -\frac{\lambda'(x_c) - \mu'(x_c)}{\lambda(x_c) + \mu(x_c)}, \tag{41b}$$

with the primes denoting derivatives. Since $x_c \in (0,1)$, the distribution is centred around a point for which the birth rate is positive. Multiplying (41) by $x^i$, integrating, we deduce that, up to exponentially small corrections, the long-term moments are of the form

$$\begin{aligned} X^i(t) &\sim \int_0^\infty x^i p(x,t)\,\mathrm{d}x \\ &\sim \frac{1}{\sqrt{\pi}(a\Omega)^{i/2}}\, \mathrm{e}^{G(x_c)t} \int_{-\infty}^{\infty} y^i \mathrm{e}^{-(y-y_c)^2}\,\mathrm{d}y \\ &\sim \frac{1}{\sqrt{\pi}(a\Omega)^{i/2}}\, \mathrm{e}^{G(x_c)t} \int_{-\infty}^{\infty} (z+y_c)^i \mathrm{e}^{-z^2}\,\mathrm{d}z \\ &= \frac{1}{\sqrt{\pi}}\, \mathrm{e}^{G(x_c)t} \sum_{j=0}^{i} \binom{i}{j} x_c^{i-j} (a\Omega)^{-j/2} \begin{cases} \Gamma\left(\frac{j+1}{2}\right), & j \text{ even}, \\ 0, & j \text{ odd}, \end{cases} \end{aligned} \tag{42}$$

where $y_c = x_c\sqrt{a\Omega}$ and $\Gamma$ is the Gamma function. Notably, we see that $X^i(t) \propto \mathrm{e}^{G(x_c)t}$ for some constant of proportionality (that depends on $i$).

In addition to the large-time benchmark, we may also compute an analytical solution in the limiting case of $\Omega \to \infty$ where we neglect the second-order term in (39) to give the hyperbolic PDE

$$\frac{\partial}{\partial t}p(x,t) = G(x)p(x,t) - \frac{\partial}{\partial x}\left[(\lambda(x) - \mu(x))p(x,t)\right], \qquad p(x,0) = p_0(x), \tag{43a}$$

with zero normal flux boundary conditions

$$(\lambda(x) - \mu(x))p(x,t) = 0, \qquad \text{at } x = 0 \text{ and as } x \to \infty. \tag{43b}$$

This is a singular limit: far from $x_c$ the solution of (39) will be well approximated by the solution of (43), however, in a vicinity of $x = x_c$, there will be a boundary layer where this approximation breaks down (and similarly if the initial distribution has steep gradients). Nevertheless, it is enlightening to study this reduced problem.

The hyperbolic PDE (43) may be solved by the method of characteristics. Substituting the particular forms (40) into (43), we find that characteristic curves satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \Lambda x(t)(1 - x(t)) - x(t), \qquad x(0) = x_0, \tag{44}$$

which admits the general solution

$$x(t) = \left[\frac{1}{x_c} + \left(\frac{1}{x_0} - \frac{1}{x_c}\right)\mathrm{e}^{-(\Lambda-1)t}\right]^{-1}. \tag{45}$$

Along characteristics, the density satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}p(t) = \left[(2\Lambda + g)x(t) - (\Lambda - 1)\right]p(t), \qquad p(0) = p_0(x_0). \tag{46}$$

Substituting the characteristic form of $x(t)$ from (45) and integrating, we find that the density satisfies

$$p(t) = p_0(x_0)\mathrm{e}^{-(\Lambda-1)t}\left[\frac{x_0}{x_c}\,\mathrm{e}^{(\Lambda-1)t} + 1 - \frac{x_0}{x_c}\right]^{2+g/\Lambda}. \tag{47}$$

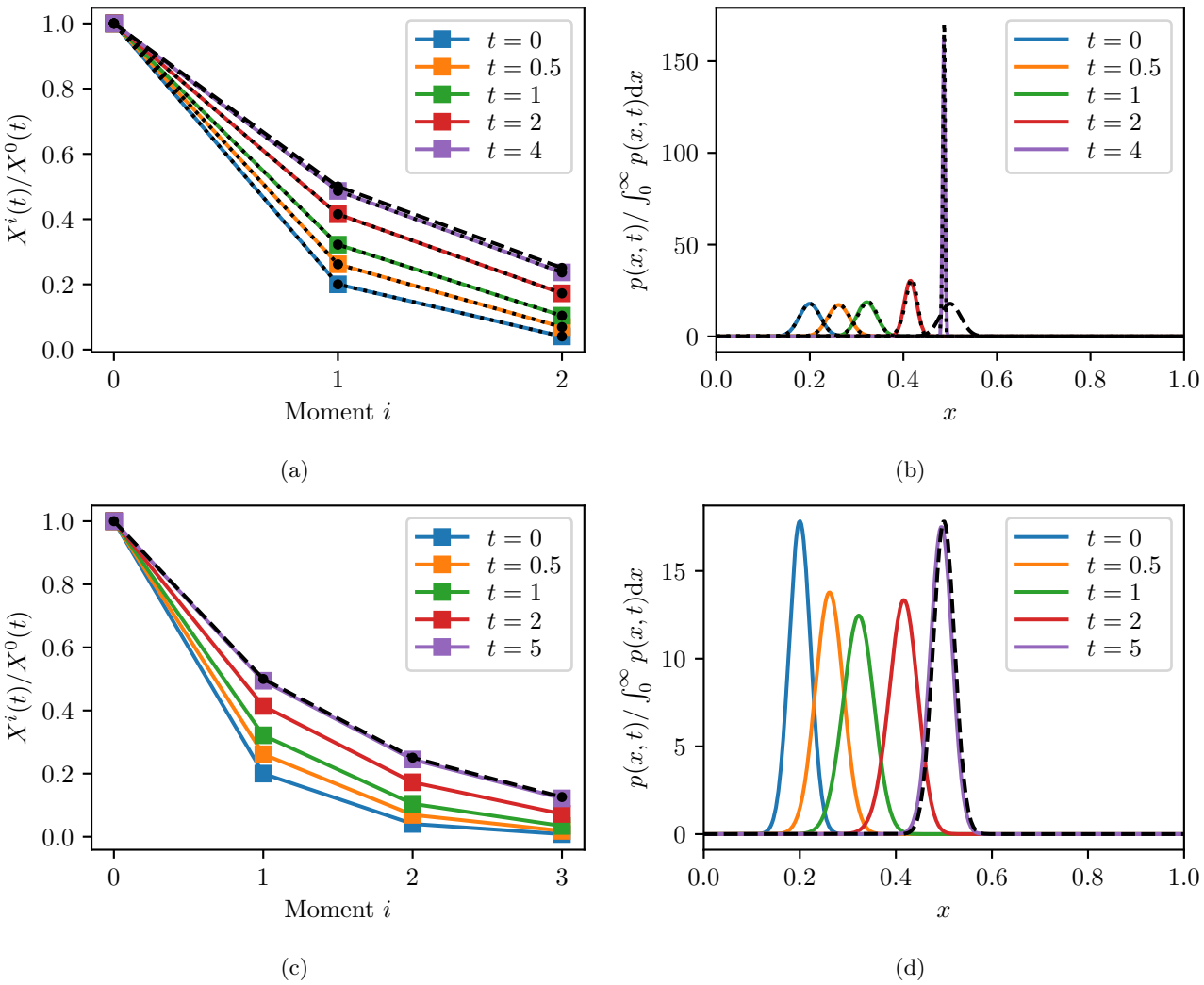



Figure 1: Numerical solutions $X^i(t)$ of the moment equations (48) using the zero-information closure (coloured curves), compared to the moments of the long-time distribution (41) given in (42) (black dashed curve), alongside the corresponding (reconstructed) distributions $p(x,t)$ at different times $t$. A truncation of $N = 3$ moments is employed along with $\Lambda = 2$, $g = 1$, $p_0(x) = \mathrm{e}^{-\Omega(x-0.2)^2}\sqrt{\Omega/\pi}$ and different values of $\Omega$. In (a,b) we use $\Omega = \infty$ and the black dotted curves depict the solution (45) and (47) of the hyperbolic problem, while in (c,d) we use $\Omega = 1000$.

Using the approach described in section 3, we determine that the associated moment hierarchy satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}X^i(t) = (g - i\Lambda)X^{i+1}(t) + i(\Lambda - 1)X^i(t) + \frac{i(i-1)}{2\Omega}\left[(\Lambda+1)X^{i-1}(t) - \Lambda X^i(t)\right], \tag{48a}$$

$$X^i(0) = \int_0^\infty x^i p_0(x)\,\mathrm{d}x =: X_0^i. \tag{48b}$$

We emphasise that the only approximation made is that the distribution is not large at the boundary. Since the forms (40) are all polynomial, there is no other approximation, and thus any appreciable error in the moment equations (48) is the result of the moment closure.

We solve the system (48) numerically using an implicit BDF solver [7, 11, 36]. As described analytically, numerical solution using the zero closure quickly leads to negative, non-physical moments. For the central closure, moments of odd order may be negative. In our simulation results, both odd and even moments quickly become negative, demonstrating non-physical behaviour. The zero-information closure on the other hand provides solutions that agree well with the closed-form solutions. For this reason, we illustrate solutions only for the zero-information closure. We reconstruct the distributions corresponding to the solutions of the moment equations by finding the maximum-entropy distribution that shares the same first $N$ moments (a component of the zero-information closure as described in section 4).

In figs. 1a and 1b we solve the moment equations (48) in the limit $\Omega \to \infty$ by neglecting the term in square

brackets. We find that the solution closely matches the solution of the hyperbolic PDE (43) (black dotted curves). This solution tends towards a Dirac mass at $x = x_c$, which shares similar moments as the large-time solution (fig. 1a) despite having noticeably different distributions (fig. 1b, compare black dotted curve at $t = 4$ to black dashed curve). This discrepancy is precisely the result of neglecting the second-order noise term. When reinstated, the solution converges towards a distribution that demonstrates good agreement with the large-time solution (figs. 1c and 1d). We mention, a posteriori, that the reconstructed distributions in fig. 1 are consistent with the assumption of negligible mass at the boundary.

In this first example, we demonstrate that the $\mathcal{O}(1/\Omega)$ contributions make a substantial difference in the distribution when solving the moment equations corresponding to the Fokker–Planck terms. The zero and central closures do not exhibit adequate performance with moments quickly becoming negative. The zero-information closure proves more stable. We proceed to study the moment equations and their closure for non-local terms.

# 6 Growth–fragmentation benchmarks

Having explored the behaviour of the Fokker–Planck dynamics using the moment equations, we now turn our attention to the non-local terms. We study two examples where again the reaction rates are of polynomial form, such that any error stems exclusively from the closure rather than the polynomial approximation. The growth–fragmentation benchmarks studied in Ref. [22] are of this form, with the additional benefit of admitting explicit solutions.

In the first scenario we model a cell population in the copy-number paradigm. Each cell contains some initial copy number $x$ of a static substance, that is, a substance that is no longer being produced but is also not decaying. Mother cells divide into two daughters with uniformly random copy numbers. The colony is housed in a bioreactor running in batch mode, and there are no other processes of relevance. The PDE takes the form

$$\frac{\partial}{\partial t}p(x,t) = -ax^k p(x,t) + \int_x^\infty 2ay^{k-1}p(y,t)\,\mathrm{d}y, \qquad p(x,0) = p_0(x). \tag{49}$$

The division rate is $ax^k$ and the division kernel is self-similar and uniform: in the notation of section 2.1, $\theta_1(x \mid \hat{x}) = 1/\hat{x}$ for $x \in [0, \hat{x}]$ and $\Theta_1(y) = 1$ for $y \in [0, 1]$.

Equation (49) admits the explicit self-similar solution [22]

$$\tilde{p}(x,t) = c\,(at)^{2/k}\mathrm{e}^{-atx^k}. \tag{50}$$

It will be helpful to also calculate the moments of solution (50),

$$\tilde{X}^i(t) := \int_0^\infty x^i\tilde{p}(x,t)\,\mathrm{d}x = c\,(at)^{(1-i)/k}\int_0^\infty x^i\mathrm{e}^{-x^k}\,\mathrm{d}x = \frac{c\,(at)^{(1-i)/k}}{k}\Gamma\left(\frac{i+1}{k}\right), \tag{51}$$

where $\Gamma$ is the Gamma function. From the properties of the Gamma function, explicit expressions are available for $k = 1$ and $k = 2$.

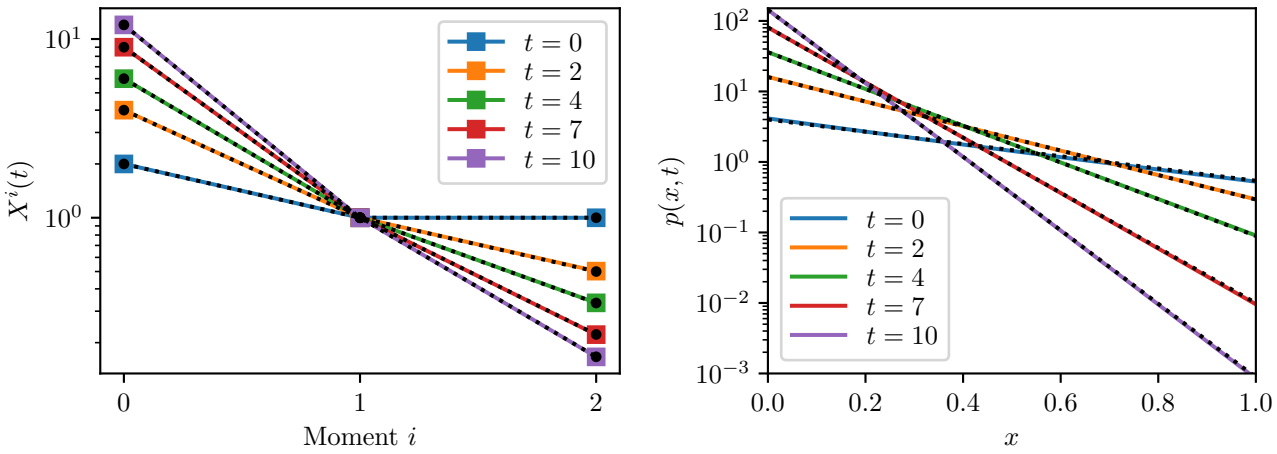
For integer $k > 0$, the moment hierarchy associated with (49) is of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}X^i(t) = a\left(\frac{1-i}{1+i}\right)X^{i+k}(t), \qquad X^i(0) = \int_0^\infty x^i p_0(x)\,\mathrm{d}x =: X_0^i. \tag{52}$$

Here again, since no approximations were made, these two moment hierarchies describe the exact evolution of the distribution's moments.

The self-similar solution (50) does not allow for an arbitrary initial condition, however, we assume that solutions for arbitrary initial conditions tend to the self-similar form in large time. We make some preliminary remarks without assuming the self-similar form. Since the distribution $p(x,t)$ must be non-negative, and assuming its continuity, we deduce that all moments are positive at time $t$ if $p(x,t)$ is positive at any point $x \geq 0$. In this case, $X^i(t)$ is strictly decreasing for all $i > 1$ while $X^0(t)$ is increasing. The first moment $X^1(t)$ is time-invariant regardless of the other moments, as the fragmentation acts to increase the number of cells, but does not change the total copy-number, leaving this a conserved quantity. It follows that the steady solution is a distribution whose $i$th moment is zero for all $i > 1$. The steady state cannot be reached in finite time: for any fixed $x > 0$, $p(x,t)$ decreases at most exponentially, therefore, the distribution remains positive for all time. From the self-similar solution (50) we may further observe that the $i$th moment is $\mathcal{O}(t^{(1-i)/k})$. In particular the 0th moment diverges at a rate $\mathcal{O}(t^{1/k})$ while the moments $i > 1$ decay algebraically. Physically, all cells have divided infinitely to accumulate their mass at the $x = 0$ state.

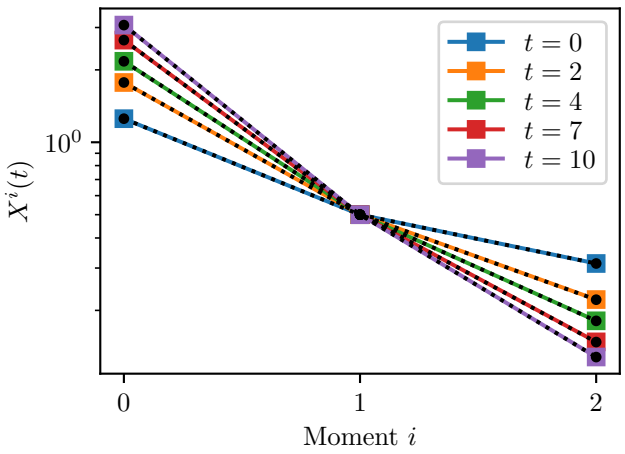
What are the moment dynamics with a zero closure? The $k$ highest-order moments, from order $N - k$ to order $N - 1$, are time-invariant because the right-hand side of (52) vanishes in the closure. These moments thus



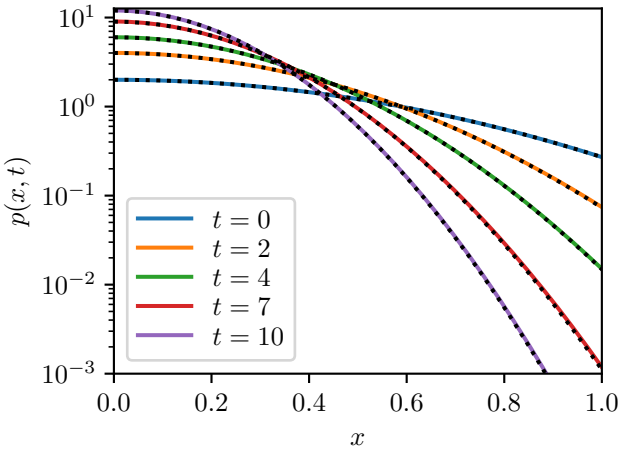

(a) $k = 1$.




(b) $k = 1$.




(c) $k = 2$.




(d) $k = 2$.

Figure 2: Numerical solutions $X^i(t)$ of the moment equations (52) using the zero-information closure (coloured curves), compared to the moments of the self-similar solutions (51) (black curves), alongside the corresponding reconstructed distributions $p(x,t)$ at different times $t$ compared to the self-similar distribution given in (50). The initial condition is given by the self-similar solution at time $t = 2$: $p_0(x) = \tilde{p}(x, 2)$. The time parameter $a = 1$ is used with a truncation of $N = 3$ moments and different model parameters $k$.

retain their initial values. If these initial value are nonzero, the system cannot converge to the exact steady state. Fatally, the $i$th moment where $i \geq \max(2, N - k)$, must decrease monotonically at constant rates (assuming that the initial moments are positive, as must be the case physically) thus eventually becoming negative and tending to $-\infty$, whereupon they are no longer physically meaningful. Similarly, the $i$th moment, where $i \geq \max(2, N - 2k)$, will be monotonically increasing and diverge to $\infty$. This reasoning may be extended to all moments $i > 1$. We can only avoid this by setting the $k$ highest moments to zero initially, which makes them vanish identically. However, the same reasoning may then be applied to the next $k$ moments, and we find that the only way to avoid the diverging moments is to set all moments $i > 1$ to zero initially. The zero closure fails irreparably, and a more informed moment closure is required.

We now study numerical solutions of the moment equations (52), which we again compute using an implicit BDF solver. Both the zero and central closures exhibit non-physical behaviour (negative moments and negative even moments, respectively). On the other hand, the zero-information closure is capable of accurately reproducing the dynamics for which we plot solutions in fig. 2. We again reconstruct the underlying distributions by finding the distribution of maximum-entropy which shares its first $N$ moments with the computed solution, as described in section 4. We see in figs. 2a and 2c how the zeroth-order moment is increasing, while the first-order order moment remains constant, and the second order decreases.

We now move to a second scenario, in which we no longer consider the substance $x$ to be static, and include a self-reinforcing production whose rate is equal to the copy-number (per unit time). Moreover, we consider a random (discrete) transition that all cells can make irrespective of their internal state, say cell death, after

which we no longer want to account for the substance. The growth rate in (3), the rate of cell division events, is $g_1(x) = akx^k$ (since we have just a single discrete state $\mathcal{K} = 1$). The division kernel in (3) is uniform, $\theta_1(x \mid \hat{x}) = 1/\hat{x}$, which, by (5), admits the self-similar representation $\Theta_1(y) = 1$ for $y \in [0, 1]$. Combining these ingredients, the PDE takes the form

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}\left[xp(x,t)\right] - (akx^k + 1)p(x,t) + \int_x^\infty 2aky^{k-1}p(y,t)\,\mathrm{d}y, \qquad p(x,0) = p_0(x). \tag{53}$$

There is no second-order term associated with the production (as obtained in the deterministic limit $\Omega \to \infty$). The equation (53) admits the self-similar solution

$$\hat{p}(x,t) = c\left[a(1-\mathrm{e}^{-kt})\right]^{2/k}\mathrm{e}^{-a(1-\mathrm{e}^{-kt})x^k}. \tag{54}$$

Here too the moments may be calculated as follows:

$$\hat{X}^i(t) := \int_0^\infty x^i\hat{p}(x,t)\,\mathrm{d}x = c\left[a(1-\mathrm{e}^{-kt})\right]^{(1-i)/k}\int_0^\infty x^i\mathrm{e}^{-x^k}\,\mathrm{d}x = \frac{c\left[a(1-\mathrm{e}^{-kt})\right]^{(1-i)/k}}{k}\Gamma\left(\frac{i+1}{k}\right). \tag{55}$$

As opposed to the first model, where the dynamics are purely division whereby the population tends to accumulate zero copy numbers in the diverging population, in this case, the production balances the copy-number loss due to cell division and the cell death balances the increasing cell count due to division, and the population tends to a finite non-singular distribution. For large time, $t \gg 1$, the distribution (53) has, up to exponentially small corrections, the limiting behaviour

$$\hat{p}(x,t) \sim c\,a^{2/k}\mathrm{e}^{-ax^k}, \tag{56}$$

and similarly the moments have the large-time behaviour

$$\hat{X}^i(t) \sim \frac{c\,a^{(1-i)/k}}{k}\Gamma\left(\frac{i+1}{k}\right). \tag{57}$$

For integer $k > 0$, the moment hierarchy corresponding to (53) is

$$\frac{\mathrm{d}}{\mathrm{d}t}X^i(t) = (i-1)X^i(t) + ak\left(\frac{1-i}{1+i}\right)X^{i+k}(t), \qquad X^i(0) = \int_0^\infty p_0(x)\,\mathrm{d}x =: X_0^i. \tag{58}$$

The hierarchy (58) admits steady solutions that satisfy the recurrence relation

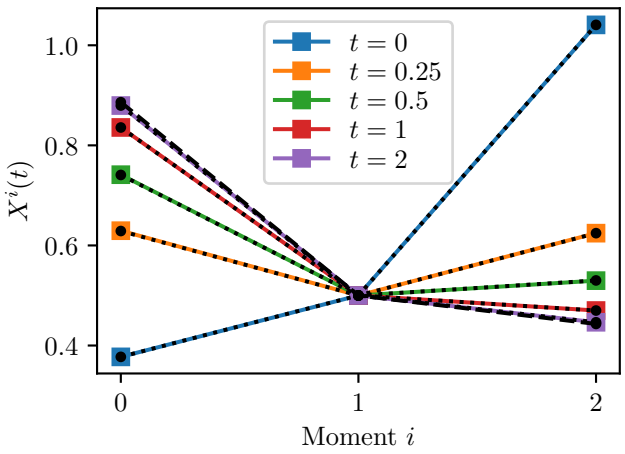
$$X^i = \frac{ak}{1+i}X^{i+k}, \tag{59}$$

for all $i \neq 1$, where $X^1$ is constant for all $t$ and thus retains its initial value. It may be verified directly that the large-time solution (57) satisfies the recurrence relation (59).

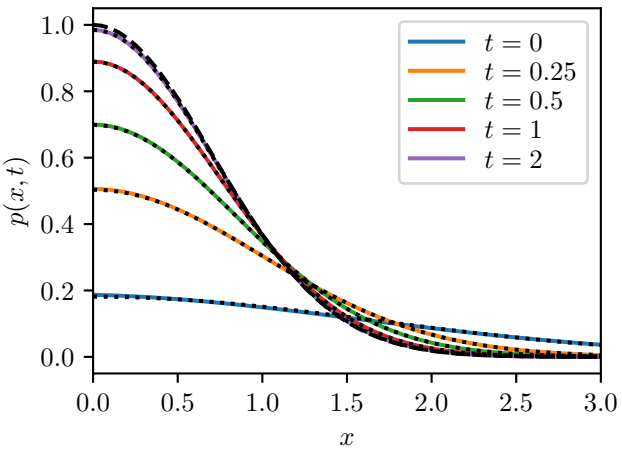
For this growth–fragmentation example, we see that the zero closure leaves the moments $X^i(t)$ growing exponentially for $i \geq \max(2, N-k)$ and for all time. This disagrees with the finite steady-state moments (57). More fatally, this closure leads to (non-physical) negative moments. To see this, note that for all $i$ such that $\max(2, N-k) \leq i < N$, the $i$th moment is an exponential with growth rate $i-1$. Then, for all $i$ such that $\max(2+k, N-k) \leq i+k < N$, the $i$th moment satisfies an ODE of the form $\dot{x}(t) = (i-1)x(t) - c\,\mathrm{e}^{(i+k-1)t}$ for some $c > 0$, whose solution is of the form $x(t) = d\mathrm{e}^{(i-1)t} - c\,\mathrm{e}^{(i+k-1)t}/k$ for some constant $d$. Therefore, for large enough time, the solution behaves like $-\mathrm{e}^{(i+k-1)t}$, diverging to $-\infty$. This reasoning may be repeated for moments $i$, where $\max(2+2k, N-k) \leq i+2k < N$, who will eventually behave as $\mathrm{e}^{(i+k-1)t}$. We may continue in this manner for all moments beyond first order to demonstrate that they will diverge. We deduce that the structure neglected in the zero closure leaves a non-physical moment system.

Numerical solution using the central closure exhibits similar non-physical behaviour of negative even moments and exponentially diverging moment magnitudes. In this case too the zero-information closure is capable of accurately simulating the system (fig. 3). For an initial condition of self-similar form, the moment-closure solution (coloured curves) aligns with the self-similar solutions (black dotted curves, figs. 3a to 3d). For an initial condition not of self-similar form, we observe rapid convergence to the large-time self-similar solution (56) (black dashed curves, figs. 3e to 3f). Note how the zeroth-order self-similar moment $\hat{X}^0(t)$ in (55) is increasing while moments of order two and above are decreasing (see also figs. 3a and 3c) whereas for the non-self-similar example we chose the zeroth-order moment above its steady-state value and the higher-order moments below the steady-state value, such that the moment dynamics are qualitatively different from the self-similar case.

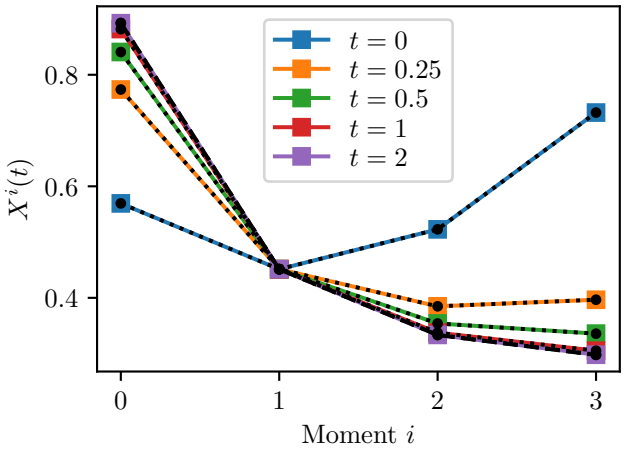
We conclude that naive zero closure is not always a feasible strategy. This observation is already well established in the biochemical literature for discrete master equations, for example, spurious dynamics including



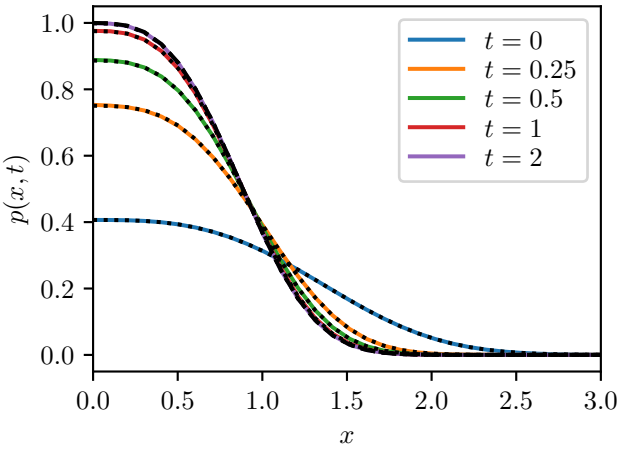

(a) $N = 3$, $k = 1$.



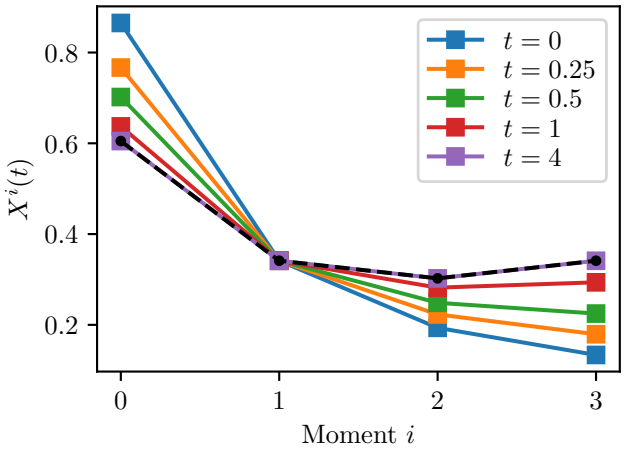

(b) $N = 3$, $k = 1$.



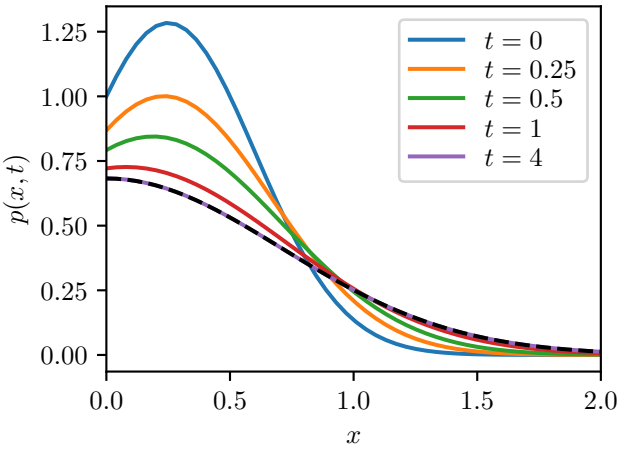

(c) $N = 4$, $k = 3$.




(d) $N = 4$, $k = 3$.




(e) $N = 4$, $k = 2$.




(f) $N = 4$, $k = 2$.

Figure 3: Numerical solutions $X^i(t)$ of the moment equations (58) using the zero-information closure (coloured curves), compared to the self-similar solutions (55) (black dotted curves) and the large-time solution (57) (black dashed curves), alongside the corresponding reconstructed distributions $p(x, t)$ at different times $t$ compared with the self-similar distribution given in (54) with its large-time form in (56). The initial condition in (a–d) is given by the self-similar solution at time $t = 0.1$: $p_0(x) = \hat{p}(x, 0.1)$, while the initial condition in (e,f) is given by $p_0(x) = \mathrm{e}^{2x-4x^2}$. The time parameter $a = 1$ is used with different truncation orders $N$ and model parameters $k$.

non-physical states where moments become negative have been studied in Refs. [28, 33, 34, 37]. However, we reiterate its relevance in the case of continuum models. Moreover, we stress that the cell-division explored here is

a population-level process that extends beyond the realm of the master equation [22], and thus the observation is extended to an altogether new class of processes. We further note that implementing the zero-information closure in the continuum (rather than a discrete state space, as was done in the original paper Ref. [38]) introduces new sources of error, as sums are replaced by integrals, which must be approximated via numerical quadrature.

With this in mind, it might be tempting to think that optimising high-order moment systems is a task doomed to failure: naive closures are typically expected to fail, and sophisticated closures, such as the zero-information closure, are difficult to implement in an optimisation framework and prohibitively costly. The primary motivation for introducing a moment approximation and closure is to avoid optimising the original heterogeneous model. However, if the reduced-order system is similarly intractable, then we have not achieved any improvement. It is for this reason that the result in the following section is both useful and unexpected.

# 7 Controlling a microbial consortium

Microbial consortia are colonies of two or more interacting microorganisms living together. Such colonies offer increased application potential over cultures of single strains since each organism can be specialised for different tasks, allowing a diverse set of mechanisms that need not coexist in a single organism. This promises improvements in bioproduction, bioremediation, and other areas [5, 18, 31, 45].

In this section we study an example of a microbial consortium (first studied in §6 of Ref. [23]), where two strains are maintained within a single bioreactor with the aim of producing a protein of interest (denoted by the state $y$). We consider a construct in which a photoreceptive TF is constitutively produced. Upon light induction, the TF (denoted by the state $x$) is recruited in the production of recombinase which leads to recombination. Only in the differentiated cell construct is the protein of interest produced. The light induction leads to a community of two genetically distinct subpopulations. We assume that the undifferentiated cells grow at some rate $A$, while the differentiated cells have an inhibited growth rate $0 \leq a(y) < A$, due to the presence of the protein, which is toxic to the cell. In fact, in such cases, microbial consortia are considered particularly useful, since a healthy and growing subpopulation can be maintained alongside the subpopulation that is effectively poisoning itself to produce the protein of interest, thereby retaining both a growing population and protein production [31]. We thus call the undifferentiated cells the "growers" and the differentiated cells the "producers". We further assume that the TF recruitment and cell recombination is faster than the other transient timescales and may be neglected.

The inhibited growth rate should satisfy $a(0) = A$ and $a(\infty) = a_\infty \in [0, A)$. We consider the monotonically decreasing Hill form

$$a(y) = a_\infty + \frac{A - a_\infty}{1 + (y/L)^m}, \tag{60}$$

where the growth rate $a(y) \in [a_\infty, A]$ interpolates nonlinearly between the minimum and maximum growth rates, $a_\infty$ and $A$, respectively, with the characteristic protein concentration associated with diminished growth is $L$.

The differentiation is modulated by the level of TF via $h(x)$ which takes the Hill form

$$h(x) = \frac{x^n}{K^n + x^n}, \tag{61}$$

for a characteristic concentration $K$ and a Hill coefficient $n$.

The governing PDE takes the form

$$\frac{\partial}{\partial t}g(x,t) = Ag(x,t) - \Lambda(t)g(x,t) - fu(t)h(x)g(x,t) - \frac{\partial}{\partial x}[(\alpha - \beta x)g(x,t)] + \frac{1}{2\Omega}\frac{\partial^2}{\partial x^2}[(\alpha + \beta x)g(x,t)], \tag{62a}$$

$$\frac{\partial}{\partial t}p(y,t) = a(y)p(y,t) - \Lambda(t)p(y,t) - \frac{\partial}{\partial y}[(\mu - \nu y)p(y,t)] + \frac{1}{2\Omega}\frac{\partial^2}{\partial y^2}[(\mu + \nu y)p(y,t)], \tag{62b}$$

for a maximal differentiation rate $f$, subject to the boundary conditions

$$(\alpha - \beta x)g(x,t) - \frac{1}{2\Omega}\frac{\partial}{\partial x}[(\alpha + \beta x)g(x,t)] = 0, \qquad \text{at } x = 0 \text{ and as } x \to \infty, \tag{62c}$$

$$(\mu - \nu x)p(y,t) - \frac{1}{2\Omega}\frac{\partial}{\partial y}[(\mu + \nu x)p(y,t)] = \begin{cases} fu(t)\int_0^\infty h(x)g(x,t)\,\mathrm{d}x, & \text{at } y = 0, \\ 0, & \text{as } y \to \infty. \end{cases} \tag{62d}$$

The turbidostat dilution rate is

$$\Lambda(t) = A\int_0^\infty g(x,t)\,\mathrm{d}x + \int_0^\infty a(y)p(y,t)\,\mathrm{d}y. \tag{62e}$$

We impose the initial conditions

$$g(x,0) = g_0(x), \qquad p(0) = 0, \tag{62f}$$

where $g_0(x)$ is the stationary solution of (62a) with no light input $u(t) = 0$. This choice reproduces typical lab conditions where the optogenetic strains are cultivated in the dark for several cell generations.

We seek to maximise total production of the protein of interest, given by

$$J_{\text{PDE}} = \int_0^\infty yp(y,T)\,\mathrm{d}y + \int_0^T \Lambda(t) \int_0^\infty yp(y,t)\,\mathrm{d}y\,\mathrm{d}t. \tag{62g}$$

Using the shorthands

$$X^i(t) := \int_0^\infty x^i g(x,t)\,\mathrm{d}x, \qquad Y^i(t) := \int_0^\infty y^i p(y,t)\,\mathrm{d}y, \tag{63}$$

and assuming that the functions $h(x)$ and $a(y)$ have polynomial approximations

$$h(x) = \sum_{j=0}^{N(h)} h_j x^j, \qquad a(y) = \sum_{j=0}^{N(a)} a_j y^j, \tag{64}$$

we use the components derived in section 3 to obtain the moment hierarchy approximation

$$\begin{aligned}
\dot{X}^i(t) &= (A - \Lambda(t))X^i(t) - fu(t) \sum_{j=0}^{N(h)} h_j X^{j+i}(t) \\
&\quad + i(\alpha X^{i-1}(t) - \beta X^i(t)) + \frac{i(i-1)}{2\Omega}(\alpha X^{i-2}(t) + \beta X^{i-1}(t)), && \text{(65a)} \\
X^i(0) &= \left(\frac{\alpha}{\beta}\right)^i + \mathcal{O}(\Omega^{-1}), && \text{(65b)} \\
\dot{Y}^i(t) &= \sum_{j=0}^{N(a)} a_j Y^{j+i}(t) - \Lambda(t)Y^i(t) + \delta_{i0} fu(t) \sum_{j=0}^{N(h)} h_j X^j(t) \\
&\quad + i(\mu Y^{i-1}(t) - \nu Y^i(t)) + \frac{i(i-1)}{2\Omega}(\mu Y^{i-2}(t) + \nu Y^{i-1}(t)), && \text{(65c)} \\
Y^i(0) &= 0, && \text{(65d)}
\end{aligned}$$

for

$$\Lambda(t) = AX^0(t) + \sum_{j=0}^{N(a)} a_j Y^j(t), \tag{65e}$$

where $\delta_{i0}$ is the Kronecker Delta function, that takes the value one when $i = 0$ and zero otherwise. For $i = 0, 1$, we understand that terms in equations (65a) and (65c) with a zero coefficient vanish (so no negative indices are encountered). In other words,

$$\dot{X}^0(t) = (A - \Lambda(t))X^0(t) - fu(t) \sum_{j=0}^{N(h)} h_j X^j(t), \qquad X^0(0) = 1, \tag{65f}$$

$$\dot{X}^1(t) = (A - \Lambda(t))X^1(t) - fu(t) \sum_{j=0}^{N(h)} h_j X^{j+1}(t) + (\alpha X^0(t) - \beta X^1(t)), \qquad X^1(0) = \frac{\alpha}{\beta}, \tag{65g}$$

and similarly for $\dot{Y}^0(t)$ and $\dot{Y}^1(t)$.

We seek to maximise the objective

$$J_{\text{MC}} = Y^1(T) + \int_0^T \Lambda(t)Y^1(t)\,\mathrm{d}t. \tag{65h}$$

The initial conditions in (65) may be determined consistently by seeking the stationary solution when $u(t) = 0$ and $Y^i(0) = 0$. We note that $Y^i(t) = 0$ for all time $t$, therefore, $\Lambda = AX^0$, and, for $i = 0$, we have $A(1-X^0)X^0 = 0$. In light of the conservation law $X^0(t) + Y^0(t) = 1$, it follows that $X^0 = 1$. For $i = 1$, we find that $X^1 = \alpha/\beta$. For $i \geq 2$ we derive the recurrence relation

$$X^i = \frac{\alpha}{\beta}\left(1 + \frac{i-1}{2\Omega}\right) X^{i-1} + \frac{\alpha}{\beta}\frac{i-1}{2\Omega} X^{i-2}. \tag{66}$$

In the limit as $\Omega \to \infty$, the leading-order solution is given by $X^i = (\alpha/\beta)^i$.

## 7.1 Influence of heterogeneity

Before proceeding to discuss the technical aspects of polynomial approximations, numerical stability, and further results, we highlight an analytical result. Applying the quotient rule, and then using (65f) and (65g), we find that the mean protein level obeys the dynamics

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{X^1(t)}{X^0(t)}\right) &= \frac{X^1(t)}{X^0(t)}\left(\frac{\dot{X}^1(t)}{X^1(t)} - \frac{\dot{X}^0(t)}{X^0(t)}\right) \\ &= \alpha - \beta\frac{X^1(t)}{X^0(t)} \underbrace{- fu(t)\sum_{j=0}^{N(h)} h_j \frac{X^1(t)}{X^0(t)}\left(\frac{X^{j+1}}{X^1(t)} - \frac{X^j(t)}{X^0(t)}\right)}_{\text{Heterogeneous contribution}}, \end{aligned} \tag{67a}$$

$$\frac{X^1(0)}{X^0(0)} = \frac{\alpha}{\beta}. \tag{67b}$$

There are contributions to the mean dynamics that go beyond the commonly used homogeneous birth–death model (see, e.g., Ref. [23]). These come from the higher-order moments when $u(t) > 0$ and reflect the heterogeneity of the distribution that, upon light induction, acts to bias the distribution due to state-dependent differentiation. In fact, we see that the first term in the sum, for $j = 0$, vanishes. This reflects the fact that, if $h(x) = h_0$ is constant, the differentiation is not state-dependent and thus there is no additional contribution to the mean dynamics.

## 7.2 Function expansions and approximations

In deriving the moment hierarchy, we assumed a polynomial approximation of the non-polynomial functions $a(y)$ and $h(x)$ defined in (60) and (61). In the present example, the functions $a(y)$ and $h(x)$ are, up to addition or multiplication by a constant, of Hill form: $\xi^n/(K^n + \xi^n)$. The Hill functional form is rational assuming $n \in \mathbb{N}$, in that it is the ratio of two polynomial forms. In [27] it is shown that a moment hierarchy may be derived without the need for polynomial approximation by multiplying the governing equations by the denominators of all rational functions before multiplying by $x^i$ and integrating.

Avoiding the need for approximation is appealing, however, this approach suffers several major drawbacks. First, it is inapplicable if the nonlinear propensity is not rational, as would be the case for non-integer $n$. Second, the resulting moment equations obtain a coupling of the $i$th moment to the $(i+n)$th moment. For large (integer) $n$, such as the example we use from the literature where $n = 10$, to resolve the $i = 0$ moment requires accurate resolution of the $i + n = 10$ moment. This high-order coupling can require the accurate resolution of an infeasibly large number of moments.

Third, for integer $n$, it is still not directly applicable to the example at hand. To see this, note that even when multiplying (62a) and (62b) by the denominators of $h(x)$ and $a(y)$, respectively, the inhomogeneous boundary condition (62d), describing the recombination from grower to producer, remains with the nonlinear form $h(x)$. This can be remedied by recasting the PDE in the full state space where both growers, $g(x, y, t)$, and producers, $p(x, y, t)$, are distributed over the full $(x, y)$-state space. The initial conditions (62f) are extended to $g(x, y, 0) = g_0(x)\delta(y)$ for the Dirac delta function and $p(x, y, 0) = 0$ remains. The boundary conditions (62c) and (62d) become zero normal flux in both dimensions for each variable. Crucially, the inhomogeneous term from the boundary condition (62d) becomes a bulk term in the PDE. Integrals in the dilution rate (62e) and the objective (62g) must be extended over the entire state space. This reformulation allows the nonlinearity at the boundary to be placed in the bulk and the approach is restored. Nevertheless, the cost is that the moment hierarchy returns to two dimensions, and we must now multiply the equation for $p$ by both denominators of $h(x)$ and $a(y)$. Thus, returning to the second point, the $\boldsymbol{i} = (0, 0)$ moment will be coupled to the $\boldsymbol{i} = (n, m)$ moment. In our example, this is of degree $n + m = 20$, and we see that for system dimension larger than one the problem is exacerbated. The requisite system size is not smaller than a feasible discretisation of the full PDE model.

We note that combining this technique with a low-order rational approximation — either by reducing the Hill coefficient $n$ or seeking a low-order Padé approximant for general nonlinear functions — might be a viable approach. Nevertheless, for the aforementioned reasons, we do not pursue this technique here any further and instead we turn our attention to polynomial approximation of the nonlinear propensities.

Perhaps the first thing that comes to mind is a truncated Taylor expansion of a function around a given point. Despite being real analytic for $n \in \mathbb{N}$ (the function is not analytic at zero for non-integer $n$), these functions have complex singularities on the circle of radius $K$, since the denominator vanishes at $\xi = -K\zeta$ for any $\zeta^n = -1$ which is of unit modulus. Therefore, the radius of convergence of any power series will be limited by how close the singularities are to the expansion point. We see this graphically in fig. 4a, where we need to expand around several different points to approximate the Hill function on the domain $[0, 1]$. There is wild deviation from the

Hill function at distances of more than approximately 0.1 from the expansion point. We infer that a fixed Taylor series will not provide a sufficiently accurate polynomial approximation.

A related method studied in the literature extends this idea by not taking a single fixed expansion point, but rather expanding around the mean of the distribution [1, 4, 9, 15, 16, 17]. The appeal of this approach is that, for distributions concentrated around their mean, the truncated Taylor expansion should constitute a close polynomial approximation of the Hill function. This approach may be implemented by representing the Taylor expansion is a power series form by using the binomial theorem

$$h(x) \approx \sum_{j=0}^{N(h)} \frac{1}{j!} h^{(j)} \left( \frac{X^1(t)}{X^0(t)} \right) \left( x - \frac{X^1(t)}{X^0(t)} \right)^j =: \sum_{j=0}^{N(h)} h_j x^j. \tag{68}$$

This proved to be numerically unstable for all approximation degrees $N(h)$, possibly owing to the large polynomial coefficients $h_j$.

To avoid having to translate the Taylor expansion form into a power series representation, we reformulate the moment hierarchy in terms of central moments, as described in full in Appendix B. This is achieved by defining moments centred at the distribution mean, and deriving the associated central moment equations. This formulation produced acceptable forward simulation results for sufficiently low Taylor truncation order, however, we were unable to obtain convergent results for the optimal control problem.

The truncated Taylor approximation at a fixed point is only locally accurate. While this motivated expanding at the mean, perhaps part of the difficulty in obtaining convergent optimisation results is due to the fact that the polynomial approximations are not "stationary". These two potential shortcomings motivate the consideration of "stationary" polynomial approximations that are accurate over a sufficiently large region of the state space (of course, polynomial approximations diverge, and thus can not accurately approximate bounded functions, for example, over the entire state space).

Given a closed interval, which we take to be $[0, 1]$, without loss of generality, and a polynomial degree $N(h)$, there exists a polynomial that best approximates the Hill function (in the $L^\infty$ sense). Minimax algorithm, such as Remez's algorithm, may be used to determine such a polynomial. A close approximation of the optimal polynomial is the Chebyshev series that interpolates the function at the Chebyshev points of the first kind. This approximation converges to the optimal approximation as the degree tends to infinity. For extended discussion and demonstration of these points, we refer the reader to Ref. [41]. Indeed, for the Hill function, we see that the convergence is rapid (fig. 4d). Nevertheless, for degree $N(h) = 8$, the approximations when varying the Hill parameter $K$ show large oscillations (fig. 4e; this is not Runge's phenomenon, since the sampling points are not equispaced). The magnitude of these oscillations, and the fact that the approximation can change sign when the original function cannot, renders them practically challenging.

To remedy this, we turn to Bernstein polynomials [8, Ch. 6], which can provide a constructive proof of Weierstrass' theorem and exhibit favourable analytical properties. In particular, the Bernstein approximants of a function converge to the function *and its derivatives*. Moreover, the approximant lies between the function's extrema and its derivatives are uniformly bounded in proportion to the function's derivatives. Approximants preserve monotonicity and convexity. These properties demonstrate how the approximant mimics important features of the original function structure. In particular, it will not suffer from the oscillations and sign changes of its Chebyshev counterpart. However [8], "There is a price that must be paid for these beautiful approximation properties: the convergence of Bernstein polynomials is very slow", as illustrated in fig. 4b. Nevertheless, for various values of $K$, the approximants preserve the qualitative features of the Hill function (fig. 4c).

## 7.3 Numerical Results

In all of the numerical simulations in this section, we use the default parameter values

$$N(a) = N(h) = 4, \quad A = 1, \quad a_\infty = 0, \quad f = 4, \quad m = n = 10, \quad K = 0.5, \quad \alpha = 0.5, \quad \beta = 1, \quad \mu = 0.5, \quad \nu = 1. \tag{69}$$

First, in fig. 5, we illustrate the moments of the forward problem solved when applying the light control profile

$$u(t) = \begin{cases} 1, & t \leq 2, \\ 0, & t > 2. \end{cases} \tag{70}$$

The comparison shows that the moments of the full distribution (coloured curves) are somewhat different from the moments obtained by solving the moment equations (65) using Bernstein approximants with $N(h) = N(a) = 3$ and the zero moment closure. It is worth keeping in mind that the full distribution is found using a numerical method that itself contains some numerical error. Nevertheless, the approximation captures much of the qualitative

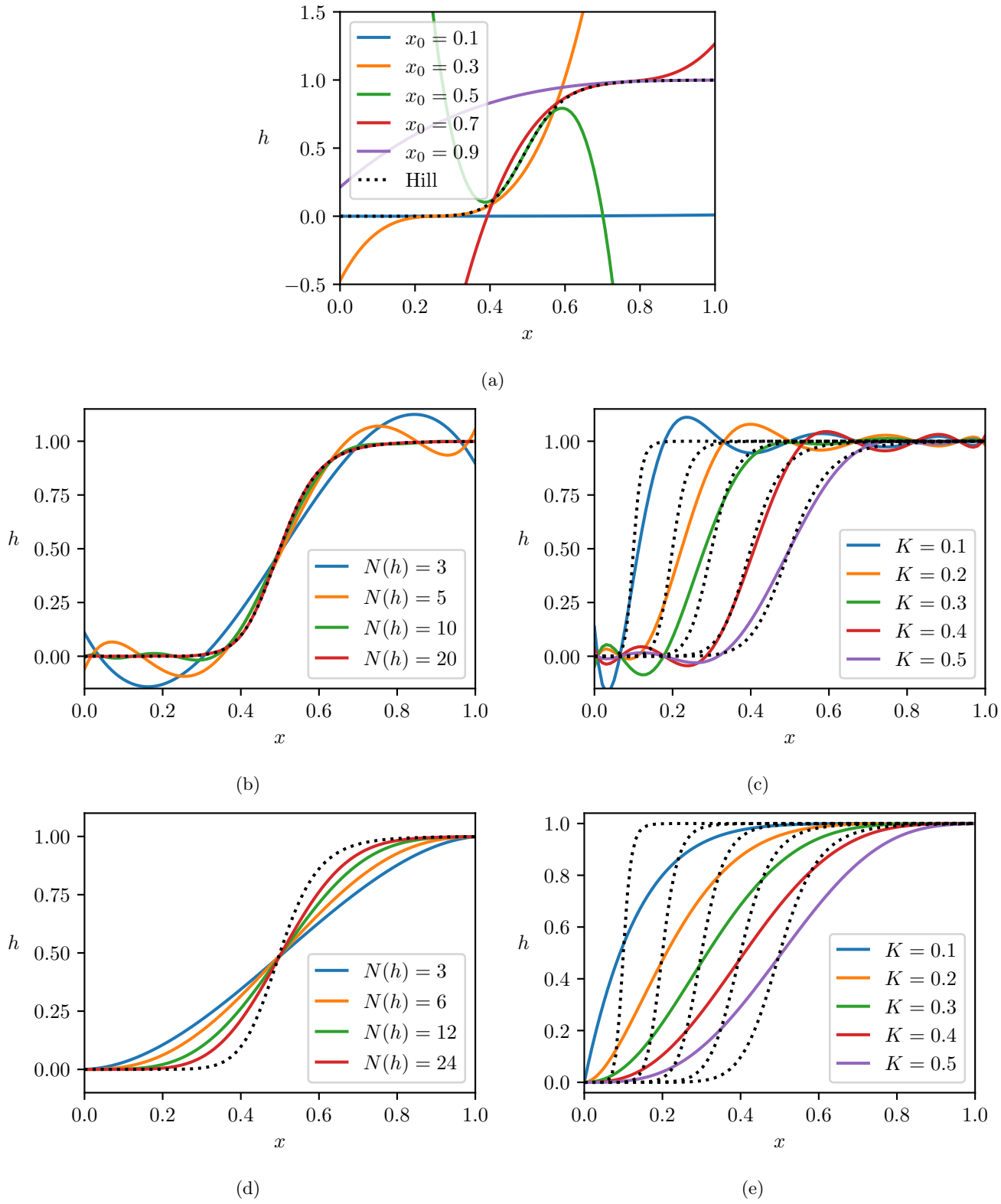
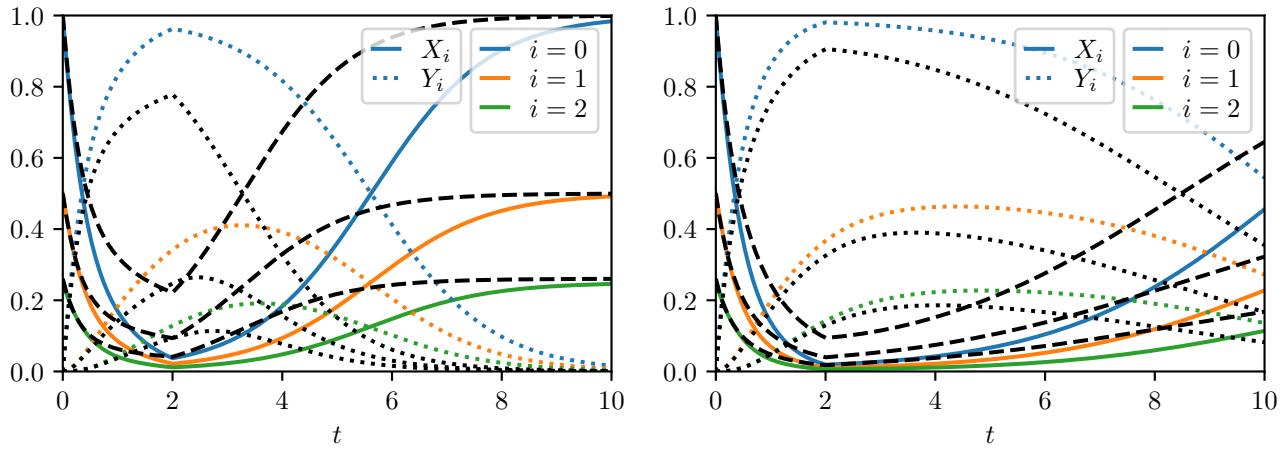



Figure 4: (a) Series expansions of the Hill function $h(x)$ in (61) for $n = 10$ and $K = 0.5$ around different points $x = x_0$ for $N(h) = 3$. (b,c) Chebyshev interpolants for different degrees $N(h)$ and Hill parameters $K$. (d,e) Bernstein approximants for different degrees $N(h)$ and Hill parameters $K$.

behaviour of the moments and is very similar to the central Taylor approximation (see fig. 7 in Appendix B). This raises the question, how does this discrepancy manifest in the associated optimal control?

To solve the optimal control problem numerically, we implement a multiple-shooting direct numerical approach based on an Euler discretisation and encoded in CasADi [3], a framework that performs automatic differentiation to produce a nonlinear program equipped with gradient information and solved by the IPOPT solver [42].




(a) $L = 0.1$.

(b) $L = 0.5$.

Figure 5: Moments of the full distribution (coloured curves) governed by (62) (solved using the discretisation from Ref. [23]) compared to solution of the moment equations (black curves) governed by (65) and using Bernstein approximants with $N(h) = N(a) = 3$ and a zero moment closure for $N = 15$ moments. The black dashed curves correspond to moments $X^i(t)$ while the black dotted curves correspond to $Y^i(t)$. The light profile (70) was employed, for various Hill parameters $L$ and the default parameters in (69).

It is not straightforward to adapt the zero-information closure to provide gradient information for use in optimal control. Moreover, the zero-information closure seems too slow to be relevant for use in optimal control. Perhaps surprisingly, we were not able to obtain good forward simulation results using the zero-information closure. For all of these reasons, to solve the optimal control problem (65) we are left with the zero and central closures, as well as Bernstein polynomial approximants and Chebyshev interpolants. Remarkably, despite the poor performance of the zero closure for the examples in sections 5 and 6, and despite the Bernstein approximants being suboptimal approximants (in the $L^\infty$ sense) this combination was the only one to exhibit robust convergence. We note that using the Chebyshev interpolants resulted in diverging iterations for all parameter values and closure methods we tried. On the other hand, using Bernstein approximants, the central closure did sometimes converge to a local optimum. In these cases, the result closely matched the local optimum reached using the zero closure. Nevertheless, the central moment closure was significantly more fragile, and small changes in the numbers of moments, the number of time steps, and the initial control guess could result in diverging iterations. Therefore, we illustrate results using Bernstein approximants and the zero moment closure.

To quantify the performance of the optimal control, we introduce the relative performance measure

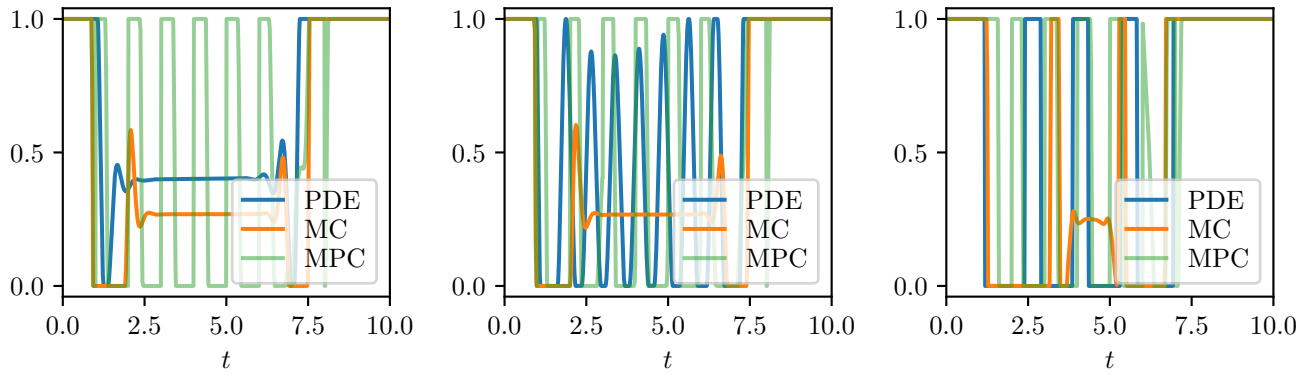
$$\mathcal{J}_{\mathrm{MC}} := \frac{J_{\mathrm{PDE}}(u_{\mathrm{MC}})}{J_{\mathrm{PDE}}(u_{\mathrm{PDE}})}, \tag{71}$$

denoting the ratio of the objective functions $J_{\mathrm{PDE}}$ of the PDE model (62) when applying different controls; $u_{\mathrm{MC}}$ denoting the optimal control of the moment-closure system (65) and $u_{\mathrm{PDE}}$ denoting the optimal control of the PDE system (62). When the PDE model is considered the ground truth, the performance measure (71) indicates how well the moment-closure model acts as a surrogate model, with $J_{\mathrm{MC}} = 1$ being an error-free replacement.

Before diving into the optimal control results, we introduce the closed-loop control $u_{\mathrm{MPC}}$ and its associated performance measure $\mathcal{J}_{\mathrm{MPC}}$ defined in analogy with (71). The open-loop control is susceptible to modeling and measurement error, and not robust to real-time changes in the system. Solving the closed-loop problem is challenging as it requires us to solve the Hamilton-Jacobi-Bellman PDE [19]. One popular method to close the loop while avoiding this problem, and thereby remedy these problems to some extent, is to resolve the open-loop problem upon receiving updated state information. This control algorithm is called the receding-horizon control, or model-predictive control (MPC), which is the term we adopt here. In the simulations here, we update the state information in the MPC control at each time unit (modeling an hourly update in the bioreactor system).

In fig. 6 we compare the PDE optimal controls with both the open-loop and closed-loop optimal controls for the moment system. We see that the open-loop controls qualitatively reproduce much of the PDE control structure. There are singular arcs figs. 6a and 6b, and for larger $L$ in figs. 6c and 6d, the moment-closure control reproduces the isolated pulses of the PDE control. This result is a qualitative improvement over the low-order moment-closure approximations studied in Ref. [23] where these features were not faithfully reproduced.

The open-loop performance incurs error of less than 5%. The closed-loop controls find corrections that guarantee performance to within less than 2% error. Again comparing to the errors of the low-order moment-



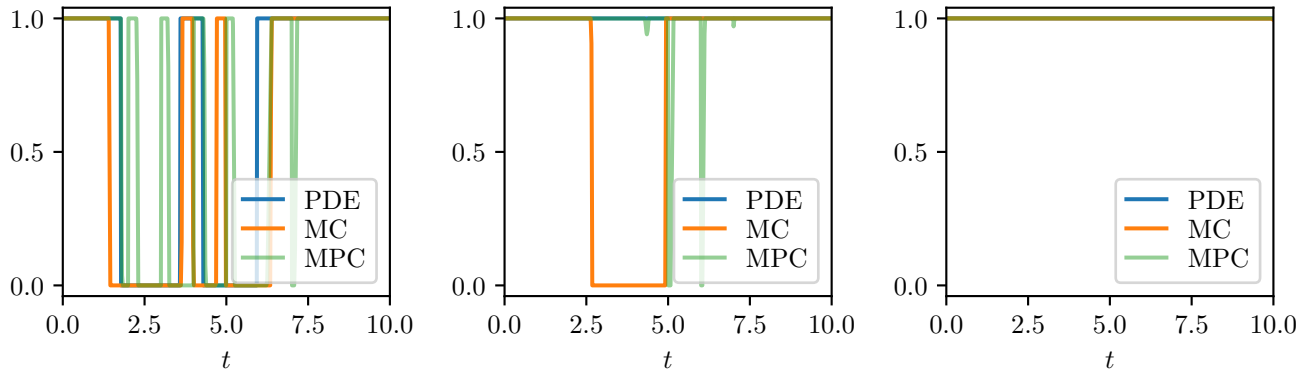

(a) $L = 0.1$: $\begin{pmatrix} \mathcal{J}_{\mathrm{MC}} \\ \mathcal{J}_{\mathrm{MPC}} \end{pmatrix} = \begin{pmatrix} 0.952 \\ 0.997 \end{pmatrix}$.




(b) $L = 0.2$: $\begin{pmatrix} \mathcal{J}_{\mathrm{MC}} \\ \mathcal{J}_{\mathrm{MPC}} \end{pmatrix} = \begin{pmatrix} 0.974 \\ 0.998 \end{pmatrix}$.




(c) $L = 0.3$: $\begin{pmatrix} \mathcal{J}_{\mathrm{MC}} \\ \mathcal{J}_{\mathrm{MPC}} \end{pmatrix} = \begin{pmatrix} 0.985 \\ 0.983 \end{pmatrix}$.




(d) $L = 0.4$: $\begin{pmatrix} \mathcal{J}_{\mathrm{MC}} \\ \mathcal{J}_{\mathrm{MPC}} \end{pmatrix} = \begin{pmatrix} 0.989 \\ 0.996 \end{pmatrix}$.




(e) $L = 0.5$: $\begin{pmatrix} \mathcal{J}_{\mathrm{MC}} \\ \mathcal{J}_{\mathrm{MPC}} \end{pmatrix} = \begin{pmatrix} 0.992 \\ 1.000 \end{pmatrix}$.




(f) $L = 0.6$: $\begin{pmatrix} \mathcal{J}_{\mathrm{MC}} \\ \mathcal{J}_{\mathrm{MPC}} \end{pmatrix} = \begin{pmatrix} 1.000 \\ 1.000 \end{pmatrix}$.

Figure 6: Optimal controls of the moment problem (65) with $N = 25$ moments, default parameters listed in (69), and various Hill parameters $L$. The optimal controls presented are the open-loop (MC) and closed-loop (MPC) controls for the PDE system (62), whose optimal control (PDE) is detailed in Ref. [23]. For $L \leq 0.2$ the PDE optimal control problems use an additional penalty term $\epsilon \int_0^T (\mathrm{d}u/\mathrm{d}t)^2 \, \mathrm{d}t$ in the cost function with $\epsilon = 10^{-4}$ to regularise high-frequency oscillations.

closure approximations in Ref. [23], which are up to 25%, these results demonstrate significant improvement. This observation — that the heterogeneity-preserving higher-order moment system restores near-optimal performance where the homogeneous moment-closure approximation fell short — drives home the importance of heterogeneity, as it demonstrates that the poor homogeneous performance was not the result of using moment equations, but of neglecting the heterogeneity.

# 8 Conclusions

Beginning with a modular class of PDE models describing the evolution of cellular populations, we derive and study an associated moment hierarchy approximation. By employing a family of benchmark processes where the moment hierarchies involve no approximation (and for which we have analytical solutions to which to compare), we explain analytically why the zero and central moment-closure approximations fail. We demonstrate how maximum entropy distributions provide a successful moment-closure technique for these systems.

Given this failure of the zero and central closures, that has been demonstrated previously in the discrete master-equation setting, it might be tempting to abandon their use. However, other methods tend to be prohibitively costly when it comes to optimisation. We provide a practically relevant model of a microbial consortium where only the simplest closure (the zero moment-closure) provides robust convergence to an accurate solution. Furthermore, this example includes rate functions of non-polynomial form. These forms do not produce a closed hierarchy of moment equations, and previous studies have ignored this additional complexity by not treating such cases. We discuss several systematic polynomial approximation methods that allows us to recover the moment hierarchy. Bernstein polynomials, while not the closest approximants, faithfully preserve features of the original non-polynomial function affording robust convergence in the optimisation routine. On the other hand, the closest

approximation typically introduces undesirable oscillations and results in moment equations for which the same optimisation routine diverges wildly.

The purpose of studying high-order moment-closure approximations rather than the full PDE model is to, on the one hand, retain the heterogeneous dynamics that are required to ensure optimal process control, while, on the other hand, enabling a dimension reduction. For this reason it is particularly useful to identify systems where the simple moment closure produces good results. In contrast to more sophisticated closure approaches, it is straightforward to implement and computationally cheap. Although it may require a particularly high-order system (we used 25 moments for this example), the optimal control is still calculated in a number of seconds on a standard laptop computer, where the optimal control of the PDE model is at least an order of magnitude slower.

Another benefit to the moment-closure approach is that, since the state space has been incorporated analytically into the framework, no discretisation is required in the state space, which can introduce non-negligible error into the numerical results. For the optimal controls studied here, the use of the polynomial approximation is a source of error absent in the PDE formulation. However, the PDE formulation studied in Ref. [23] uses finite-differences in the state space, which introduce a new source of error via numerical diffusion.

# Appendix

# A Moment hierarchy approximation

In this appendix, we detail how the components introduced in section 2 contribute to the moment equations.

## A.1 Copy-number population growth

In the copy-number paradigm, we multiply (3) by $\boldsymbol{x}^{\boldsymbol{i}}$ and integrate over the state space to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}X_k^{\boldsymbol{i}}(t) = -\sum_{|\boldsymbol{j}|\leq N(g_k)} g_{k,\boldsymbol{j}}(t)X_k^{\boldsymbol{i}+\boldsymbol{j}}(t) + 2\int_{\mathbb{R}_+^d} \boldsymbol{x}^{\boldsymbol{i}} \int_{\hat{\boldsymbol{x}}\geq\boldsymbol{x}} g_k(\hat{\boldsymbol{x}},t)p_k(\hat{\boldsymbol{x}},t)\theta_k(\boldsymbol{x}\mid\hat{\boldsymbol{x}})\,\mathrm{d}\hat{\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} + (\text{Non-growth sources}), \tag{72}$$

where, changing the order of integration, and leveraging self-similarity (5), we see that

$$\begin{aligned}\int_{\mathbb{R}_+^d} \boldsymbol{x}^{\boldsymbol{i}} \int_{\hat{\boldsymbol{x}}\geq\boldsymbol{x}} g_k(\hat{\boldsymbol{x}},t)p_k(\hat{\boldsymbol{x}},t)\theta_k(\boldsymbol{x}\mid\hat{\boldsymbol{x}})\,\mathrm{d}\hat{\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} &= \int_{\mathbb{R}_+^d} g_k(\hat{\boldsymbol{x}},t)p_k(\hat{\boldsymbol{x}},t)\int_{\boldsymbol{0}\leq\boldsymbol{x}\leq\hat{\boldsymbol{x}}} \boldsymbol{x}^{\boldsymbol{i}}\theta_k(\boldsymbol{x}\mid\hat{\boldsymbol{x}})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\hat{\boldsymbol{x}}\\ &= \int_{\mathbb{R}_+^d} \hat{\boldsymbol{x}}^{\boldsymbol{i}} g_k(\hat{\boldsymbol{x}},t)p_k(\hat{\boldsymbol{x}},t)\int_{\boldsymbol{0}\leq\boldsymbol{y}\leq\boldsymbol{1}} \boldsymbol{y}^{\boldsymbol{i}}\Theta_k(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\hat{\boldsymbol{x}}\\ &= \Theta_k^{\boldsymbol{i}}\sum_{|\boldsymbol{j}|\leq N(g_k)} g_{k,\boldsymbol{j}}(t)X_k^{\boldsymbol{i}+\boldsymbol{j}}(t).\end{aligned} \tag{73}$$

for

$$\Theta_k^{\boldsymbol{i}} := \int_{\boldsymbol{0}\leq\boldsymbol{y}\leq\boldsymbol{1}} \boldsymbol{y}^{\boldsymbol{i}}\Theta_k(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{74}$$

The contributions in (20) follow.

Some moments $\Theta_k^{\boldsymbol{i}}$ may be calculated in terms of lower-order moments using symmetry:

$$\begin{aligned}\Theta_k^{\boldsymbol{i}} = \int_{\boldsymbol{0}\leq\boldsymbol{y}\leq\boldsymbol{1}} \boldsymbol{y}^{\boldsymbol{i}}\Theta_k(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_{\boldsymbol{0}\leq\boldsymbol{y}\leq\boldsymbol{1}} \boldsymbol{y}^{\boldsymbol{i}}\Theta_k(\boldsymbol{1}-\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} &= \int_{\boldsymbol{0}\leq\boldsymbol{y}\leq\boldsymbol{1}} (\boldsymbol{1}-\boldsymbol{y})^{\boldsymbol{i}}\Theta_k(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\\ &= \int_{\boldsymbol{0}\leq\boldsymbol{y}\leq\boldsymbol{1}} \left(\prod_{j=1}^d (1-y_j)^{i_j}\right)\Theta_k(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\\ &= \int_{\boldsymbol{0}\leq\boldsymbol{y}\leq\boldsymbol{1}} \left(\prod_{j=1}^d \left[\sum_{k=0}^{i_j}\binom{i_j}{k}(-1)^k y_j^k\right]\right)\Theta_k(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\\ &= (-1)^{|\boldsymbol{i}|}\Theta_k^{\boldsymbol{i}} + \sum_{\boldsymbol{j}<\boldsymbol{i}} c_{\boldsymbol{j}}\Theta_k^{\boldsymbol{j}},\end{aligned} \tag{75}$$

where $\boldsymbol{j} < \boldsymbol{i}$ denotes the set of multi-indices $\{\boldsymbol{j} \mid \forall m : j_m \leq i_m,\ \exists m : j_m < i_m\}$ and $c_{\boldsymbol{j}}$ are integer coefficients. Therefore, for odd orders $|\boldsymbol{i}|$, we have

$$\Theta_k^{\boldsymbol{i}} = \frac{1}{2} \sum_{\boldsymbol{j} < \boldsymbol{i}} c_{\boldsymbol{j}} \Theta_k^{\boldsymbol{j}}. \tag{76}$$

## A.2 Continuum dynamics

The contributions stemming from the Fokker–Planck terms (13) can be derived by expressing the operator in divergence form, namely,

$$\frac{\partial}{\partial t} p_k(\boldsymbol{x}, t) = -\nabla \cdot \boldsymbol{\Phi}_k(\boldsymbol{x}, t) + (\text{other sources}), \tag{77}$$

for the flux $\boldsymbol{\Phi}_k(\boldsymbol{x}, t)$ given by

$$\boldsymbol{\Phi}_k(\boldsymbol{x}, t) = \boldsymbol{e}_k \left\{ r_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) - \frac{1}{2\Omega} \boldsymbol{e}_k \cdot \nabla \left[ r_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) \right] \right\}. \tag{78}$$

By application of the divergence theorem, and using the zero normal flux boundary condition (14), it follows that there are no zeroth-order contributions ($\boldsymbol{i} = \boldsymbol{0}$). This reflects the fact that the process describes a change in the continuum species, but not the population mass. By application of Green's identities, we find, for $|\boldsymbol{i}| > 0$, that

$$\begin{aligned}
&- \int_{\mathbb{R}_+^d} \boldsymbol{x}^{\boldsymbol{i}} \nabla \cdot \boldsymbol{\Phi}_k(\boldsymbol{x}, t) \, \mathrm{d}\boldsymbol{x} \\
&\quad = \int_{\mathbb{R}_+^d} \left( \nabla \boldsymbol{x}^{\boldsymbol{i}} \cdot \boldsymbol{e}_k \right) \left\{ r_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) - \frac{1}{2\Omega} \boldsymbol{e}_k \cdot \nabla \left[ r_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) \right] \right\} \mathrm{d}\boldsymbol{x} \\
&\quad \approx \int_{\mathbb{R}_+^d} \left( \sum_{\ell=1}^{d} i_\ell e_{k,\ell} \boldsymbol{x}^{\boldsymbol{i} - \boldsymbol{v}_\ell} \right) r_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) + \frac{1}{2\Omega} \left( r_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) \boldsymbol{e}_k \cdot \nabla \left( \nabla \boldsymbol{x}^{\boldsymbol{i}} \cdot \boldsymbol{e}_k \right) \right) \mathrm{d}\boldsymbol{x} \\
&\quad = \sum_{\ell=1}^{d} i_\ell e_{k,\ell} \int_{\mathbb{R}_+^d} \boldsymbol{x}^{\boldsymbol{i} - \boldsymbol{v}_\ell} r_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) \, \mathrm{d}\boldsymbol{x} + \frac{1}{2\Omega} \sum_{\ell=1}^{d} \sum_{m=1}^{d} i_\ell (i_m - \delta_{\ell m}) e_{k,\ell} e_{k,m} \int_{\mathbb{R}_+^d} \boldsymbol{x}^{\boldsymbol{i} - \boldsymbol{v}_\ell - \boldsymbol{v}_m} r_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) \, \mathrm{d}\boldsymbol{x} \\
&\quad \approx \sum_{\ell=1}^{d} i_\ell e_{k,\ell} \sum_{|\boldsymbol{j}| \leq N(r_k)} r_{k,\boldsymbol{j}}(t) X_k^{\boldsymbol{i} - \boldsymbol{v}_\ell + \boldsymbol{j}}(t) + \frac{1}{2\Omega} \sum_{\ell=1}^{d} \sum_{m=1}^{d} i_\ell (i_m - \delta_{\ell m}) e_{k,\ell} e_{k,m} \sum_{|\boldsymbol{j}| \leq N(r_k)} r_{k,\boldsymbol{j}}(t) X_k^{\boldsymbol{i} - \boldsymbol{v}_\ell - \boldsymbol{v}_m + \boldsymbol{j}}(t),
\end{aligned} \tag{79}$$

where $\delta_{\ell m}$ is the Kronecker Delta function, that takes the value one when $\ell = m$ and zero otherwise, $e_{k,\ell}$ denotes the $\ell$th component of the vector $\boldsymbol{e}_k$, and $\boldsymbol{v}_\ell$ denotes the vector of zeros except for the $\ell$th component whose value is one. The first approximation stems from ignoring boundary contributions (these are small if the density is small at the boundaries since the term has a coefficient of $\mathcal{O}(1/\Omega)$), while the second approximation stems from the polynomial approximation. The form in (25) follows.

The bursty production (15) similarly has no zeroth-order contribution. To see this, note that the zeroth-order terms sum all possible jumps — for each $\boldsymbol{x}$ add all the valid jumps *to* $\boldsymbol{x}$ — which may be expressed from the dual perspective — for each $\boldsymbol{x}$ add all the valid jumps *from* $\boldsymbol{x}$. Then, utilising (16), we find that

$$\begin{aligned}
&\|\boldsymbol{e}_k\| \int_{\mathbb{R}_+^d} \int_{\boldsymbol{x} - z\boldsymbol{e}_k \in \mathbb{R}_+^d} f_k(\boldsymbol{x} - z\boldsymbol{e}_k, t) p_k(\boldsymbol{x} - z\boldsymbol{e}_k, t) Q_k(\boldsymbol{x} - z\boldsymbol{e}_k, z\|\boldsymbol{e}_k\|) \, \mathrm{d}z \, \mathrm{d}\boldsymbol{x} \qquad (80) \\
&\qquad = \int_{\mathbb{R}_+^d} f_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) \int_0^\infty \|\boldsymbol{e}_k\| Q_k(\boldsymbol{x}, z\|\boldsymbol{e}_k\|) \, \mathrm{d}z \, \mathrm{d}\boldsymbol{x} \\
&\qquad = \int_{\mathbb{R}_+^d} f_k(\boldsymbol{x}, t) p_k(\boldsymbol{x}, t) \, \mathrm{d}\boldsymbol{x}. \qquad (81)
\end{aligned}$$

For the higher order calculation, we perform a preliminary binomial expansion for vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and the multi-index $\boldsymbol{i}$:

$$(\boldsymbol{x} + \boldsymbol{y})^{\boldsymbol{i}} = \prod_{j=1}^{d} (x_j + y_j)^{i_j} = \prod_{j=1}^{d} \left( \sum_{\ell=0}^{i_j} \binom{i_j}{\ell} x_j^\ell y_j^{i_j - \ell} \right) = \sum_{\boldsymbol{j} \leq \boldsymbol{i}} b_{\boldsymbol{i},\boldsymbol{j}} \boldsymbol{x}^{\boldsymbol{j}} \boldsymbol{y}^{\boldsymbol{i} - \boldsymbol{j}}, \tag{82}$$

for some integer coefficients $b_{\boldsymbol{i},\boldsymbol{j}}$, where $b_{\boldsymbol{i},\boldsymbol{i}} = 1$. Then, at higher order, again appealing to the dual perspective, we evaluate the integral:

$$\begin{aligned}
&\int_{\mathbb{R}^d_+} \boldsymbol{x}^{\boldsymbol{i}} \|\boldsymbol{e}_k\| \int_{\boldsymbol{x}-z\boldsymbol{e}_k \in \mathbb{R}^d_+} f_k(\boldsymbol{x}-z\boldsymbol{e}_k,t) p_k(\boldsymbol{x}-z\boldsymbol{e}_k,t) Q(\boldsymbol{x}-z\boldsymbol{e}_k, z\|\boldsymbol{e}_k\|)\,\mathrm{d}z\,\mathrm{d}\boldsymbol{x} \\
&\quad = \int_{\mathbb{R}^d_+} \int_{\boldsymbol{x}-z\boldsymbol{e}_k \in \mathbb{R}^d_+} (\boldsymbol{x}-z\boldsymbol{e}_k + z\boldsymbol{e}_k)^{\boldsymbol{i}} \|\boldsymbol{e}_k\| f_k(\boldsymbol{x}-z\boldsymbol{e}_k,t) p_k(\boldsymbol{x}-z\boldsymbol{e}_k,t) Q_k(\boldsymbol{x}-z\boldsymbol{e}_k, z\|\boldsymbol{e}_k\|)\,\mathrm{d}z\,\mathrm{d}\boldsymbol{x} \\
&\quad = \sum_{\boldsymbol{j}\le\boldsymbol{i}} b_{\boldsymbol{i},\boldsymbol{j}} \boldsymbol{e}_k^{\boldsymbol{i}-\boldsymbol{j}} \int_{\mathbb{R}^d_+} \int_{\boldsymbol{x}-z\boldsymbol{e}_k \in \mathbb{R}^d_+} (\boldsymbol{x}-z\boldsymbol{e}_k)^{\boldsymbol{j}} z^{|\boldsymbol{i}-\boldsymbol{j}|} \|\boldsymbol{e}_k\| f_k(\boldsymbol{x}-z\boldsymbol{e}_k,t) p_k(\boldsymbol{x}-z\boldsymbol{e}_k,t) Q_k(\boldsymbol{x}-z\boldsymbol{e}_k, z\|\boldsymbol{e}_k\|)\,\mathrm{d}z\,\mathrm{d}\boldsymbol{x} \\
&\quad = \sum_{\boldsymbol{j}\le\boldsymbol{i}} b_{\boldsymbol{i},\boldsymbol{j}} \boldsymbol{e}_k^{\boldsymbol{i}-\boldsymbol{j}} \int_{\mathbb{R}^d_+} \boldsymbol{x}^{\boldsymbol{j}} f_k(\boldsymbol{x},t) p_k(\boldsymbol{x},t) \int_0^\infty z^{|\boldsymbol{i}-\boldsymbol{j}|} \|\boldsymbol{e}_k\| Q_k(\boldsymbol{x}, z\|\boldsymbol{e}_k\|)\,\mathrm{d}z\,\mathrm{d}\boldsymbol{x} \\
&\quad = \sum_{\boldsymbol{j}\le\boldsymbol{i}} b_{\boldsymbol{i},\boldsymbol{j}} \boldsymbol{e}_k^{\boldsymbol{i}-\boldsymbol{j}} \|\boldsymbol{e}_k\|^{-|\boldsymbol{i}-\boldsymbol{j}|} \int_{\mathbb{R}^d_+} \boldsymbol{x}^{\boldsymbol{j}} Q_{k,|\boldsymbol{i}-\boldsymbol{j}|}(\boldsymbol{x}) f_k(\boldsymbol{x},t) p_k(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x},
\end{aligned} \tag{83}$$

where we have defined, for $j \in \mathbb{N}_0$,

$$Q_{k,j}(\boldsymbol{x}) := \int_0^\infty y^j Q_k(\boldsymbol{x}, y)\,\mathrm{d}y. \tag{84}$$

Recalling that $b_{\boldsymbol{i},\boldsymbol{i}} = 1$, we deduce the form in (26).

# B Taylor-expansion central moment equations

In this appendix, we formulate a system of central moment equations of the consortium model (62) based on a truncated Taylor expansion of the nonlinear propensities. The idea is analogous to the non-central moment equations discussed in section 3, except the moments and polynomial approximations are centred about the distribution means.

We begin by defining the shifted moments

$$\hat{X}^i(t) := \int_0^\infty (x - m_x(t))^i g(x,t)\,\mathrm{d}x, \qquad \hat{Y}^i(t) := \int_0^\infty (y - m_y(t))^i p(y,t)\,\mathrm{d}y, \tag{85}$$

for arbitrary functions $m_x(t)$ and $m_y(t)$.

Next, we Taylor expand the nonlinear propensities in equations (62a) and (62b) about $m_x(t)$ and $m_y(t)$, respectively:

$$\begin{aligned}
h(x) &\approx \sum_{j=0}^{N(h)} \frac{h^{(j)}(m_x(t))}{j!} (x - m_x(t))^j =: \sum_{j=0}^{N(h)} \hat{h}_j(m_x(t))(x - m_x(t))^j, \\
a(y) &\approx \sum_{j=0}^{N(a)} \frac{a^{(j)}(m_y(t))}{j!} (y - m_y(t))^j =: \sum_{j=0}^{N(a)} \hat{a}_j(m_y(t))(y - m_y(t))^j.
\end{aligned} \tag{86}$$

Multiplying equations (62a) and (62b) by $(x - m_x(t))^i$ and $(y - m_y(t))^i$, respectively, and applying the same approximation techniques as developed for the non-shifted moment systems, we find that the shifted moments

satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{X}^i(t) = -i\hat{X}^{i-1}(t)\dot{m}_x(t) + A\hat{X}^i(t) - \Lambda(t)\hat{X}^i(t) - fu(t)\sum_{j=0}^{N(h)}\hat{h}_j(m_x(t))\hat{X}^{j+i}(t) \\ + i\left[\alpha\hat{X}^{i-1}(t) - \beta\left(\hat{X}^i(t) + m_x(t)\hat{X}^{i-1}(t)\right)\right] + \frac{i(i-1)}{2\Omega}\left[\alpha\hat{X}^{i-2}(t) + \beta\left(\hat{X}^{i-1}(t) + m_x(t)\hat{X}^{i-2}(t)\right)\right], \tag{87a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{Y}^i(t) = -i\hat{Y}^{i-1}(t)\dot{m}_y(t) + \sum_{j=0}^{N(a)}\hat{a}_j(m_y(t))\hat{Y}^{j+i}(t) - \Lambda(t)\hat{Y}^i(t) + (-m_y(t))^i fu(t)\sum_{j=0}^{N(h)}\hat{h}_j(m_x(t))\hat{X}^j(t) \\ + i\left[\mu\hat{Y}^{i-1}(t) - \nu\left(\hat{Y}^i(t) + m_y(t)\hat{Y}^{i-1}(t)\right)\right] + \frac{i(i-1)}{2\Omega}\left[\mu\hat{Y}^{i-2}(t) + \nu\left(\hat{Y}^{i-1}(t) + m_y(t)\hat{Y}^{i-2}(t)\right)\right], \tag{87b}$$

$$\Lambda(t) = A\hat{X}^0(t) + \sum_{j=0}^{N(a)}\hat{a}_j(m_y(t))\hat{Y}^j(t). \tag{87c}$$

The form of equations (87a) and (87b) is similar of the non-shifted moment formulation (65) with some perturbations due to the moments being shifted to a non-zero, time-varying point.

The initial conditions of system (87a) and (87b) are given by the steady state solution in the absence of light ($u(t) = 0$) with no producers. To determine these initial conditions, we need to specify the functions $m_x(t)$ and $m_y(t)$, which we take as the means of each distribution $g(x,t)$ over $x$ and $p(y,t)$ over $y$, namely

$$m_x(t) = \frac{X^1(t)}{\hat{X}^0(t)}, \qquad\qquad m_y(t) = \frac{Y^1(t)}{\hat{Y}^0(t)}, \tag{87d}$$

where $X^1(t)$ and $Y^1(0)$ are the non-shifted moments defined in (63). We thus refer to the moments as *central moments*. Note that the central and non-central zeroth-order moments coincide by definition: $X^0(t) = \hat{X}^0(t)$ and $Y^0(t) = \hat{Y}^0(t)$. Moreover, the first-order central moment vanishes identically by definition, since

$$\hat{X}^1(t) = \int_0^\infty \left(x - \frac{X^1(t)}{\hat{X}^0(t)}\right) g(x,t)\,\mathrm{d}x = X^1(t) - \frac{X^1(t)}{\hat{X}^0(t)}\hat{X}^0(t) = 0, \tag{87e}$$

and similarly $\hat{Y}^1(t) = 0$.

This formulation requires that we track the non-central first-order moments. The governing equations may be written in a manner consistent with the central moment formulation, namely

$$\frac{\mathrm{d}}{\mathrm{d}t}X^1(t) = AX^1(t) - \Lambda(t)X^1(t) - fu(t)\sum_{j=0}^{N(h)}\hat{h}_j(m_x(t))\left(\hat{X}^{j+i}(t) + \frac{X^1(t)}{\hat{X}^0(t)}\hat{X}^j(t)\right) + \alpha\hat{X}^0(t) - \beta X^1(t), \tag{87f}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y^1(t) = \sum_{j=0}^{N(a)}\hat{a}_j(m_y(t))\left(\hat{Y}^{j+1}(t) + \frac{Y^1(t)}{\hat{Y}^0(t)}\hat{Y}^j(t)\right) - \Lambda(t)Y^1(t) + \mu Y^0(t) - \nu Y^1(t). \tag{87g}$$

The right-hand sides of (87f) and (87g) depend only on $\{\hat{X}^0(t), \ldots, \hat{X}^N(t), X^1(t), \hat{Y}^0(t), \ldots, \hat{Y}^N(t), Y^1(t)\}$.

We may now determine the initial conditions by seeking the steady state of (87) with no light input and no initial producers, whereby $\hat{Y}^i = Y^1 = 0$ for all $i$. Setting time derivatives to zero, we find that

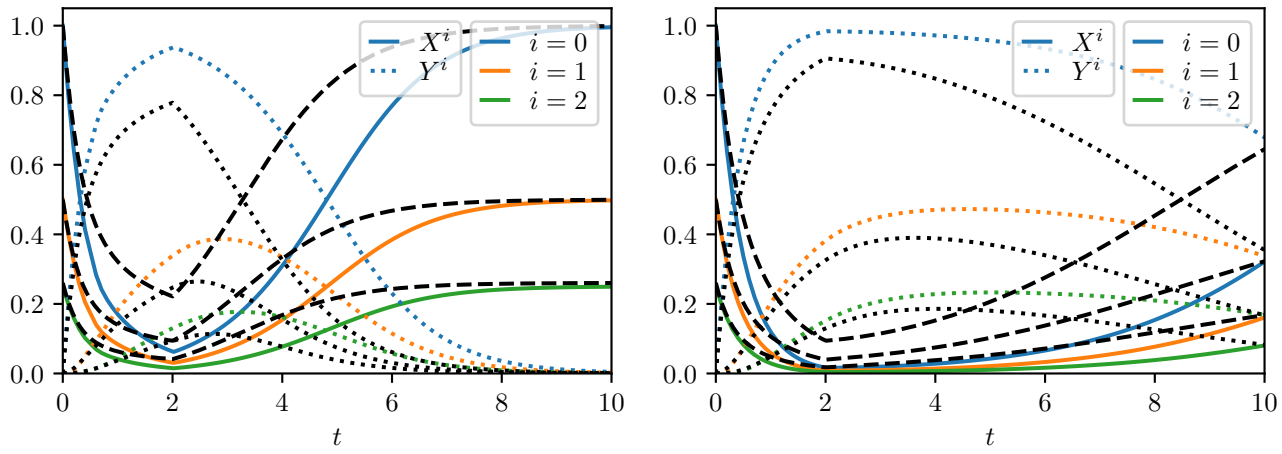
$$0 = A\hat{X}^i - \Lambda\hat{X}^i + i\left[\alpha\hat{X}^{i-1} - \beta\left(\hat{X}^i + \frac{X^1}{\hat{X}^0}\hat{X}^{i-1}\right)\right] + \frac{i(i-1)}{2\Omega}\left[\alpha\hat{X}^{i-2} + \beta\left(\hat{X}^{i-1} + \frac{X^1}{\hat{X}^0}\hat{X}^{i-2}\right)\right], \tag{88}$$

where $\Lambda = A\hat{X}^0$. For $i = 0$, we deduce that $0 = A\hat{X}^0(1 - \hat{X}^0)$, and hence $\hat{X}^0 = 1$, since the zero solution is not physically feasible for non-negative growth rates. It follows that $\Lambda = A$ and the growth terms vanish. For $i = 1$, we find that

$$X^1 = \frac{\alpha}{\beta}. \tag{89}$$

For $i \geq 2$, we deduce the recurrence relation

$$\hat{X}^i = \frac{i-1}{2\Omega}\left(\hat{X}^{i-1} + \frac{2\alpha}{\beta}\hat{X}^{i-2}\right), \tag{90}$$




(a) $L = 0.1$.

(b) $L = 0.5$.

Figure 7: Moments of the full distribution (coloured curves) governed by (62) (solved using the discretisation from Ref. [23]) compared to solution of the central moment equations (black curves) governed by (87) and using a central moment closure via truncating at $N = 10$ moments. The black dashed curves correspond to moments $X^i(t)$ while the black dotted curves correspond to $Y^i(t)$. The light profile (70) was employed, for various Hill parameters $L$ and the default parameters in (69).

which defines the initial conditions up to arbitrary order $N$ using the initial conditions $\hat{X}^1 = 0$ and $\hat{X}^0 = 1$.

Finally, the objective function (62g) takes the central-moment form

$$J_{\mathrm{MC}} = Y^1(T) + \int_0^T \Lambda(t) Y^1(t)\,\mathrm{d}t. \tag{91}$$

A "zero" moment closure in this formulation is equivalent to what we refer to as a central moment closure. For the sake of comparison to fig. 5, we plot the non-central moments, which are given by expressing the central moments (85) in terms of the non-central moments (63)

$$\hat{X}^i(t) = \int_0^\infty \left(x - \frac{X^1(t)}{\hat{X}^0(t)}\right)^i g(x,t)\,\mathrm{d}x = X^i(t) + \sum_{j=0}^{i-1} \binom{i}{j} X^j(t) \left(\frac{X^1(t)}{\hat{X}^0(t)}\right)^{i-j}. \tag{92}$$

The identity (92) may be inverted to give $X^i(t)$ in terms of $\hat{X}^i(t)$ (and lower-order moments $\{X^j(t)\}$ for $j < i$).

In fig. 7, we plot the moments of the forward problem solved using the central moment hierarchy and truncated Taylor expansion approximations (87), with the light control profile (70), in analogy with fig. 5. The moments are very similar to those obtained using the Bernstein polynomial approximations with the non-central moment hierarchy and a zero moment closure (fig. 5). The central system with Taylor approximations has a slightly closure agreement with the moments of the full PDE distribution.

The results were robust to the order $N$ of the moment truncation, but less so to the orders of the Taylor series truncation $N(h)$ and $N(a)$. From $N(h) = N(a) = 4$ there was a noticeable deterioration in performance due to the manifestation of numerical instabilities. This compromises the approximation: low-order polynomials must be used, however, these only provide good agreement with original non-polynomial function in a narrow vicinity of the mean. We were unable to find convergent results (for any polynomial degree) using this approach for the optimal control problem.

## References

[1] Angelique Ale, Paul Kirk, and Michael PH Stumpf. A general moment expansion method for stochastic kinetic models. *J. Chem. Phys.*, 138(17):174101, 2013.

[2] S. J. Altschuler and L. F. Wu. Cellular heterogeneity: do differences make a difference? *Cell*, 141(4):559–563, 2010.

[3] J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. CasADi: a software framework for nonlinear optimization and optimal control. *Math. Program. Comput.*, 11(1):1–36, 2019.

[4] A. Andreychenko, L. Mikeev, and V. Wolf. Model reconstruction for moment-based stochastic chemical kinetics. *ACM T. Model. Comput. S.*, 25(2):1–19, 2015.

[5] K. Brenner, L. You, and F. H. Arnold. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.*, 26(9):483–489, 2008.

[6] A. Brock, H. Chang, and S. Huang. Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.*, 10(5):336–342, 2009.

[7] G. D. Byrne and A. C. Hindmarsh. A polyalgorithm for the numerical solution of ordinary differential equations. *ACM Trans. Math. Softw.*, 1(1):71—-96, mar 1975.

[8] P. J. Davis. *Interpolation and approximation*. Dover Publications, 1975.

[9] S. Engblom. Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comput.*, 180(2):498–515, 2006.

[10] N. Friedman, L. Cai, and X. S. Xie. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys. Rev. Lett.*, 97:168302, Oct 2006.

[11] E. Hairer, S. P. Norsett, and G. Wanner. Solving ordinary differential equations i: Nonstiff problems. *SIAM Rev.*, 32(3):485, 1990.

[12] J. Harmand, C. Lobry, A. Rapaport, and T. Sari. *The chemostat: Mathematical theory of microorganism cultures*. John Wiley & Sons, USA, 2017.

[13] J. Harmand, C. Lobry, A. Rapaport, and T. Sari. *Optimal Control in Bioprocesses: Pontryagin's Maximum Principle in Practice*. John Wiley & Sons, USA, 2019.

[14] S. Huang. Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136(23):3853–3862, 2009.

[15] A. Kazeroonian, F. J. Theis, and J. Hasenauer. Modeling of stochastic biological processes with non-polynomial propensities using non-central conditional moment equation. *IFAC P. Vol.*, 47(3):1729–1735, 2014.

[16] E. Lakatos, A. Ale, P. D. W. Kirk, and M. P. H. Stumpf. Multivariate moment closure techniques for stochastic kinetic models. *J. Chem. Phys.*, 143(9):094107, 2015.

[17] C. H. Lee. A moment closure method for stochastic chemical reaction networks with general kinetics. *MATCH - Commun. Math. Co.*, 70(3):785–800, 2013.

[18] Z. Li, X. Wang, and H. Zhang. Balancing the non-linear rosmarinic acid biosynthetic pathway by modular co-culture engineering. *Metab. Eng.*, 54:1–11, 2019.

[19] D. Liberzon. *Calculus of variations and optimal control theory*. Princeton university press, 2011.

[20] Y. T. Lin and C. R. Doering. Gene expression dynamics with stochastic bursts: Construction and exact results for a coarse-grained model. *Phys. Rev. E*, 93:022409, Feb 2016.

[21] D. Lunz. On continuum approximations of discrete-state markov processes of large system size. *Multiscale Model. Sim.*, 19(1):294–319, 2021.

[22] D. Lunz, G. Batt, J. Ruess, and J. F. Bonnans. Beyond the chemical master equation: Stochastic chemical kinetics coupled with auxiliary processes. *PLOS Comput. Biol.*, 17(7):1–24, 07 2021.

[23] D. Lunz, J. F. Bonnans, and J. Ruess. Optimal control of bioproduction in the presence of population heterogeneity. https://hal.inria.fr/hal-03445175.

[24] D. L. Marchisio and R. O. Fox. Solution of population balance equations using the direct quadrature method of moments. *J. Aerosol Sci.*, 36(1):43–73, 2005.

[25] R. McGraw. Description of aerosol dynamics by the quadrature method of moments. *Aerosol Sci. Tech.*, 27(2):255–265, 1997.

[26] R. McGraw and D. L. Wright. Chemically resolved aerosol dynamics for internal mixtures by the quadrature method of moments. *J. Aerosol Sci.*, 34(2):189–209, 2003.

[27] P. Milner, C. S. Gillespie, and D. J. Wilkinson. Moment closure approximations for stochastic kinetic models with rational rate laws. *Math. Biosci.*, 231(2):99–104, 2011.

[28] I. Nåsell. An extension of the moment closure method. *Theor. Popul. Biol.*, 64(2):233–239, 2003.

[29] J. Paijmans, M. Bosman, P. R. ten Wolde, and D. K. Lubensky. Discrete gene replication events drive coupling between the cell cycle and circadian clocks. *P. Natl. Acad. Sci. USA*, 113(15):4063–4068, 2016.

[30] S. E. Pratsinis. Simultaneous nucleation, condensation, and coagulation in aerosol reactors. *J. Colloid Interf. Sci.*, 124(2):416–427, 1988.

[31] K. M. Rapp, J. P. Jenkins, and M. J. Betenbaugh. Partners for life: building microbial consortia for the future. *Curr. Opin. Biotech.*, 66:292–300, 2020.

[32] M. D. Rolfe, C. J. Rice, S. Lucchini, C. Pin, A. Thompson, A. D. S. Cameron, M. Alston, M. F. Stringer, R. P. Betts, J. Baranyi, M. W. Peck, and J. C. D. Hinton. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J. bacteriol.*, 194(3):686–701, 2012.

[33] D. Schnoerr, G. Sanguinetti, and R. Grima. Validity conditions for moment closure approximations in stochastic chemical kinetics. *J. Chem. Phys.*, 141(8):084103, 2014.

[34] D. Schnoerr, G. Sanguinetti, and R. Grima. Comparison of different moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.*, 143(18):185101, 2015.

[35] D. Schnoerr, G. Sanguinetti, and R. Grima. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J. Phys. A - Math. Theor.*, 50(9):093001, jan 2017.

[36] L. F. Shampine and M. W. Reichelt. The matlab ode suite. *SIAM J. Sci. Comput.*, 18(1):1–22, 1997.

[37] A. Singh and J. P. Hespanha. Approximate moment dynamics for chemically reacting systems. *IEEE T. Automat. Cont.*, 56(2):414–418, 2011.

[38] P. Smadbeck and Y. N. Kaznessis. A closure scheme for chemical master equations. *P. Natl. Acad. Sci. USA*, 110(35):14261–14265, 2013.

[39] S. L. Spencer, S. Gaudet, J. G. Albeck, J. M. Burke, and P. K. Sorger. Non-genetic origins of cell-to-cell variability in trail-induced apoptosis. *Nature*, 459(7245):428–432, 2009.

[40] D. G. Spiller, C. D. Wood, D. A. Rand, and M. R. H. White. Measurement of single-cell dynamics. *Nature*, 465(7299):736–745, 2010.

[41] L. N. Trefethen. *Approximation Theory and Approximation Practice*. SIAM, 2019.

[42] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.*, 106(1):25–57, March 2006.

[43] P. Whittle. On the use of the normal approximation in the treatment of stochastic processes. *J. Roy. Stat. Soc. B Met.*, 19(2):268–281, 1957.

[44] M. Yu, J. Lin, and T. Chan. A new moment method for solving the coagulation equation for particles in brownian motion. *Aerosol Sci. Tech.*, 42(9):705–713, 2008.

[45] K. Zhou, K. Qiao, S. Edgar, and G. Stephanopoulos. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.*, 33(4):377–383, 2015.