

Approximation and Model Reduction for the Stochastic Kinetics of Reaction Networks

Bronstein, Leo
(2020)

DOI (TUprints): <https://doi.org/10.25534/tuprints-00013433>

Lizenz:



CC-BY-NC-ND 4.0 International - Creative Commons, Namensnennung, nicht kommerziell, keine Bearbeitung

Publikationstyp: Dissertation

Fachbereich: 18 Fachbereich Elektrotechnik und Informationstechnik

Quelle des Originals: <https://tuprints.ulb.tu-darmstadt.de/13433>

Approximation and Model Reduction for the Stochastic Kinetics of Reaction Networks

Dem Fachbereich Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt

zur Erlangung des Grades
Doctor rerum naturalium (Dr. rer. nat.)
vorgelegte Dissertation

von

Leo Bronstein, M.Sc.

Erstgutachter: Prof. Dr. techn. Heinz Koepl
Zweitgutachter: Prof. Dr. Peter Sollich

Darmstadt 2019

Bronstein, Leo: *Approximation and Model Reduction for the Stochastic Kinetics of Reaction Networks*
Darmstadt, Technische Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUprints: 2020
URN: urn:nbn:de:tuda-tuprints-134330
Tag der mündlichen Prüfung: 03.12.2019

Veröffentlicht unter CC BY-NC-ND 4.0 International
<https://creativecommons.org/licenses/>

Abstract

The mathematical modeling of the dynamics of cellular processes is a central part of systems biology. It has been realized that noise plays an important role in the behavior of these processes. This includes not only intrinsic noise, due to “random” molecular events within the cell, but also extrinsic noise, due to the varying environment of a cellular (sub-)system. These environmental effects and their influence on the system of interest have to be taken into account in a mathematical model.

The thesis at hand deals with the (exact or approximate) reduced or marginal description of cellular subsystems when the environment of the subsystem is of no interest, and also with the approximate solution of the forward problem for biomolecular reaction networks in general. These topics are investigated across the hierarchy of possible models for reaction networks, from continuous-time Markov chains to stochastic differential equations to ordinary differential equation models.

The first contribution is the derivation of moment closure approximations via a variational approach. The resulting viewpoint sheds light on the problems usually associated with moment closure, and allows one to correct some of them. The full probability distributions obtained from the variational approach are used to find approximate descriptions of heterogeneous rate equations with log-normally distributed extrinsic noise. The variational method is also extended to the approximation of multi-time joint distributions. Finally, the general form of moment equations and cumulant equations for mass-action kinetics is derived in the form of a diagrammatic technique.

The second contribution is the investigation of the use of the Nakajima-Zwanzig-Mori projection operator formalism for the treatment of heterogeneous kinetics. Cumulant expansions in terms of partial cumulants are used to obtain approximate convolutional forward equations for the process of interest, with the heterogeneous reaction rates or the environment marginalized out. The performance of the approximation is investigated numerically for simple linear networks.

Finally, extending previous work, a marginal description of the subsystem of interest on the process level, for fully bi-directionally coupled reaction networks, is obtained by means of stochastic filtering equations in combination with entropic matching. The resulting approximation is interpreted as an orthogonal projection of the full joint master equation, making it conceptually similar to the projection operator formalism. For mass-action kinetics, a product-Poisson ansatz for the filtering distribution leads to the simplest possible marginal process description, which is investigated analytically and numerically.

Zusammenfassung

Die mathematische Modellierung der Dynamik von biologischen Prozessen in Zellen ist ein zentraler Teil der Systembiologie. Es hat sich herausgestellt, dass Stochastizität eine wichtige Rolle im Verhalten dieser Prozesse spielt. Dabei ist nicht nur intrinsisches Rauschen, verursacht durch "zufällige" molekulare Ereignisse, von Bedeutung, sondern auch extrinsisches Rauschen, welches durch Variabilität in der Umgebung des (Sub-)Systems entsteht. Diese Umgebungseffekte und ihr Einfluss auf das interessierende System müssen in einem mathematischen Modell berücksichtigt werden.

Die vorliegende Arbeit behandelt die (exakte oder approximative) reduzierte (marginalisierte) Beschreibung von zellulären Subsystemen, wenn die Umgebung des Subsystems nicht von Interesse ist. Außerdem wird allgemein das Vorwärtsproblem für biomolekulare Reaktionsnetzwerke behandelt. Diese Fragestellungen werden über die ganze Modellhierarchie von Reaktionsnetzwerken betrachtet, von Markovketten in kontinuierlicher Zeit über stochastische Differenzialgleichungen bis hin zu gewöhnlichen Differenzialgleichungen.

Im ersten Teil der Arbeit werden Moment Closure Approximationen über einen Variationsansatz hergeleitet. Dieser Blickwinkel erlaubt es, einige der Probleme, die gewöhnlich mit Moment Closure assoziiert werden, zu verstehen und teilweise auch zu beheben. Die Wahrscheinlichkeitsverteilungen, die der Variationsansatz liefert, werden benutzt um die approximativen Lösungen von heterogener Dynamik von Reaktionsratengleichungen mit Log-Normal verteilter Heterogenität zu bestimmen. Die Variationsmethode wird auch auf die Approximation von gemeinsamen Verteilungen mehrerer Zeitpunkte verallgemeinert. Zuletzt wird die allgemeine Form von Kumulanten- bzw. Momentengleichungen durch eine Diagrammtechnik beschrieben.

Im zweiten Teil der Arbeit wird die Verwendung des Projektionsoperator-Formalismus von Nakajima, Zwanzig und Mori für die Behandlung von heterogener Reaktionskinetik untersucht. Kumulantenentwicklungen in partiellen Kumulanten werden verwendet, um approximative Vorwärtsgleichungen mit Konvolutionsterm für den marginalisierten Prozess zu erhalten. Die resultierende Approximation wird analytisch und numerisch untersucht.

Im letzten Teil wird, als Verallgemeinerung von existierenden Arbeiten, die marginale Beschreibung eines Subsystems für beliebige, bidirektional gekoppelte Reaktionsnetzwerke entwickelt. Hierfür werden stochastische Filtergleichungen mit Entropic Matching kombiniert. Die resultierende Approximation wird als orthogonale Projektion der Mastergleichung des vollen Prozesses interpretiert, wodurch die hergeleitete Methode konzeptionell ähnlich zum Projektionsoperator-Formalismus wird. Für Systeme mit Massenwirkungskinetik wird mit einem Produkt-Poisson-Ansatz die einfachste Form des approximativen marginalen Prozesses analytisch und numerisch untersucht.

Acknowledgments

I sincerely thank my supervisor Prof. Heinz Koepl for the opportunity to work in the highly interdisciplinary environment of his lab. His approach to research, which is characterized by an open mindset towards new ideas, methods and research directions, has been very encouraging and has had a strong influence on me.

I thank my collaborators Jascha Diemer, Johannes Falk, Maleen Hanst, Christopher Schneider, Christian Wildner, Sikun Yang and Christoph Zechner for the fruitful work on our joint projects.

I would also like to thank the other members of the lab, namely Bastian Alt, Derya Altintan, Sara Al-Sayed, Kilian Heck, Wasiur Rahman Khuda Bukhsh, Lukas Köhs, Nikita Kruk, Francois-Xavier Lehr, Dominik Linzner, Tim Prangemeier, Adrian Šošić, Nurgazy Sulaimanov and Klaus-Dieter Voss for the cooperative environment at the lab during my time there. I would like to thank Christine Cramer and Markus Baier for their regular help on administrative or technical problems.

Most importantly, I thank my parents for their continuous support throughout my academic career, without which the completion of this thesis would not have been possible.

Contents

1	Introduction	8
2	Preliminaries	11
2.1	Mathematical modeling of biomolecular kinetics	11
2.1.1	Continuous-time Markov chain models	12
2.1.2	Stochastic differential equation models	14
2.2	Modeling heterogeneity	15
2.2.1	Modeling heterogeneity using the CIR process	16
2.3	Approximate solutions of CME and CFPE	17
2.3.1	Stochastic simulation	18
2.3.2	Moment equations	18
3	A variational approach to moment closure	22
3.1	Variational moment closure and entropic matching	23
3.1.1	Variational moment closure	23
3.1.2	Entropic matching	25
3.1.3	Zero-information moment closure	28
3.1.4	Non-uniqueness of the distributional ansatz and minimum relative entropy	29
3.2	Examples of variational moment closure and entropic matching	29
3.2.1	Product-Poisson entropic matching for the CME	30
3.2.2	The finite state projection algorithm for the CME	30
3.2.3	Gaussian entropic matching for the Fokker-Planck equation	31
3.2.4	Log-normal entropic matching for the CFPE	32
3.3	Analysis of moment-closure schemes and the Poisson correction	32
3.3.1	Poisson mixtures	33
3.3.2	A single-species system	34
3.3.3	Other failure modes	38
3.4	Multi-time joint distributions	39
3.4.1	Multi-time moment closure and entropic matching	39
3.4.2	Consistency	41
3.5	Log-normal entropic matching for heterogeneous kinetics	42
3.5.1	Derivation	43
3.5.2	Analytical results	43
3.5.3	Numerical results	44
3.6	A diagrammatic technique for cumulant equations	46
3.6.1	The diagrammatic rules	47
3.6.2	Derivation	53

4	The projection operator formalism for marginalization of heterogeneous reaction kinetics	57
4.1	The projection operator formalism across the model hierarchy	57
4.1.1	Mean equation for the projected CME	59
4.2	Cumulant expansions for heterogeneous rate constants	62
4.2.1	Marginalization via Green's functions	63
4.2.2	Cumulant expansions	66
4.2.3	Example: Birth-death process with CIR-distributed decay rate . .	69
4.2.4	Example: Simple gene expression with CIR-distributed heterogeneity	70
4.2.5	Example: A case with a conservation relation	72
5	The marginal process framework for bi-directionally coupled reaction networks	74
5.1	Setting	75
5.2	Outline of the method	75
5.3	The marginal process and the filtering equation	79
5.3.1	The marginal process	79
5.3.2	The filtering equation	81
5.3.3	Application to reaction networks	83
5.3.4	Example: A case with finite-dimensional filtering equations	85
5.4	Finite-dimensional approximations of the filtering equation	86
5.4.1	Moment equations	86
5.4.2	Entropic matching	87
5.4.3	Example: The totally asymmetric exclusion process	88
5.5	The product-Poisson marginal process	90
5.5.1	The product-Poisson closure	90
5.5.2	Explicit representation of marginal rates	93
5.5.3	Limitations of the product-Poisson closure	94
5.6	Auxiliary-variable master equations for the marginal process with static heterogeneity	95
5.6.1	Marginal CME for static heterogeneity	95
5.6.2	Marginal SDE for static heterogeneity	97
5.7	Filtering equation and entropic matching as projection operations	100
6	Summary and Outlook	102
A	Differential equations for the solution of the cumulant expansions	104
B	The marginal simulation algorithm	105
C	List of Acronyms	107
	Bibliography	108

Chapter 1

Introduction

The quantitative understanding of the behavior of biological cells requires the consideration of dynamical properties of the cell. Advances in experimental techniques make it possible to observe single cells over a period of time, obtaining (possibly multivariate) measurements of molecule abundances. It has become apparent that stochastic fluctuations can play a major role in the behavior of these systems [1, 2]. This involves not only *intrinsic noise*, i.e. fluctuations due to “random” molecular events, but also *extrinsic noise* (or *heterogeneity*), i.e. differences in the environment between cells or fluctuations of the environment over time. Understanding of the underlying dynamics requires effective mathematical methods for the treatment of these noise sources.

The biomolecular reaction networks that describe subsystems of a cell are generally modeled via stochastic processes. The mathematical analysis of these stochastic process models has in recent years emerged as an important field within systems biology. A variety of open problems exists, a central one being the efficient solution of the forward problem for the dynamics of a given model of a cellular subsystem: Given the model and initial conditions, how does the probability distribution describing our state of knowledge of the system state evolve over time? That question can be approached from at least two different directions.

The first approach is stochastic simulation [3], which yields complete trajectories of the process under study. All properties of the reaction network can then in principle be computed as sample averages from the trajectories. This includes multi-time correlations and sensitivities to parameter values [4]. For large reaction networks, however, simulation-based approaches can be computationally very expensive. This is particularly true when the inverse problem is considered, where parameters of the model are to be determined from experimental data. Various approaches for the reduction of the computational cost of stochastic simulation have been developed [5, 6, 7].

The other main approach is the approximate computation of time-marginal probability distributions of the stochastic process, or even of just a finite number of moments. Methods of this type include the finite state projection algorithm for the approximate determination of the marginal probability distributions [8], van Kampen’s system size expansion [9] and moment closure approximations [10, 11].

Irrespective of which of these two approaches is used, it is very common to be interested only in a subpart of the full system that one is modeling. It is thus natural to search for reduced descriptions. Mathematically, this amounts to the marginalization of nuisance variables. A fundamental aspect of such a marginalization is the appearance of memory: The resulting marginal stochastic process is no longer Markovian. One can now proceed to either derive (possibly approximate) expressions for these memory terms [12, 13, 14,

15], or to derive approximate Markovian descriptions, based for instance on time-scale separation or abundance separation [16, 17, 18, 19, 20, 21, 22].

A related biological phenomenon that has received a lot of attention recently is *extrinsic noise* or *heterogeneity*: The observation that a population of cells will often exhibit fluctuations of molecule abundances that are larger than what would be expected due to intrinsic noise only. Mathematically, taking these effects into account amounts to the modeling of the cellular environment, either explicitly or implicitly via random parameters of the model. In both cases, the mathematical treatment of heterogeneity will be very similar to the treatment of subsystem dynamics described above.

The thesis at hand deals with methods of model reduction and approximation of stochastic process models of biomolecular reaction networks. The outline is as follows:

Chapter 2: The stochastic process models that are used throughout the thesis are introduced. It is explained how heterogeneity can be included in these models. Standard methods for the solution of the forward problem are reviewed.

Chapter 3: A building block for some of the model reduction techniques discussed later are moment closure approximations. Moment closures have been known as ad-hoc approximations, and a number of pathological behaviors have been observed. Here, a variational justification for a subset of moment closures is provided and used to solve the problem of divergences at low system sizes. A related approximation method, entropic matching, is also discussed and generalized. These techniques are extended to approximations for multi-time joint distributions. Variational moment closure is applied to the treatment of heterogeneous dynamics of deterministically modeled reaction networks with log-normal heterogeneity. Finally, a diagrammatic method for the derivation of moment or cumulant equations of arbitrary order is presented.

Chapter 4: A method for model reduction and marginalization that has been employed in many different contexts is the projection operator formalism. Here, this method is applied to biomolecular reaction networks, primarily for the approximate treatment of heterogeneous rate constants using cumulant expansions. The result is an approximate convolutional Kolmogorov-forward equation for the marginal dynamics of the network. The performance of the approximation is investigated for simple linear reaction networks.

Chapter 5: A framework for principled model reduction for reaction networks is presented. Extending the work in [12], marginal process and filtering equations for a general, bi-directionally coupled reaction network are derived. Using results from Chapter 3, a principled approximation of the filtering equation is applied. The resulting approximation is investigated both analytically and numerically. For the case of static heterogeneity, auxiliary-variable master equations for the marginal process are derived. Finally, the interpretation of the approximate marginal process in terms of a projection of the joint master equation is discussed.

Related publications

This thesis is based on the following publications:

Chapter 2 includes content from

[23] Leo Bronstein, Christoph Zechner and Heinz Koepl. Bayesian inference of reaction kinetics from single-cell recordings across a heterogeneous cell population. *Methods*, 85:22-35, 2015.

Chapter 3 is based on

[24] Leo Bronstein and Heinz Koepl. A variational approach to moment-closure approximations for the kinetics of biomolecular reaction networks. *Journal of Chemical Physics*, 148(1):014105, 2018.

[25] Leo Bronstein and Heinz Koepl. A diagram technique for cumulant equations in biomolecular reaction networks with mass-action kinetics. In *2016 IEEE 55th Conference on Decision and Control*, pp. 5857–5862, IEEE, 2016.

Chapter 4 includes material from

[26] Johannes Falk, Leo Bronstein, et al. Context in synthetic biology: Memory effects of environments with mono-molecular reactions. *Journal of Chemical Physics*, 150(2):024106, 2019.

Chapter 5 is based on

[27] Leo Bronstein and Heinz Koepl. Marginal process framework: A model reduction tool for Markov jump processes. *Physical Review E*, 97(6):062147, 2018.

Other publications produced by the author during his time as a doctoral student, not explicitly discussed in the present thesis, are:

[28] Christopher Schneider, Leo Bronstein, et al. ROC'n'Ribo: Characterizing a riboswitching expression system by modeling single-cell data. *ACS Synthetic Biology*, 6(7):1211–1224, 2017.

[29] Leo Bronstein and Heinz Koepl. Scalable inference using PMCMC and parallel tempering for high-throughput measurements of biomolecular reaction networks. In *2016 IEEE 55th Conference on Decision and Control*, pp. 770–775, IEEE, 2016.

Chapter 2

Preliminaries

In this chapter, basic concepts used throughout the thesis are reviewed. The stochastic process models of biomolecular kinetics, including methods for modeling heterogeneity, are introduced and illustrated on some simple examples from biology. Standard methods for the solution of the forward problem for biomolecular kinetics are reviewed.

A word on notation: The expectation of a function $\phi(x)$ with respect to a distribution $p(x)$ is written as $\langle \phi(x) \rangle_p$. The argument is dropped, $\langle \phi \rangle_p$, if there is no risk of confusion. If p is parameterized, say p_t or p_θ , we also write $\langle \phi \rangle_t$ or $\langle \phi \rangle_\theta$ or similar. The subscript is also dropped, $\langle \phi \rangle$, if there is no risk of confusion. In certain cases, to distinguish between expectations over different sets of variables, the notation $\mathbb{E}[\cdot]$ will also be used.

2.1 Mathematical modeling of biomolecular kinetics

The mathematical results in this thesis were developed for the treatment of quantitative questions of the biomolecular kinetics inside cells. While many of the mathematical results are applicable in various other contexts where reaction network models (or more general Markov processes) are appropriate, the results are nevertheless mostly formulated in the language of biomolecular reaction networks and illustrated on examples from biology. Therefore, some basic biological facts, and their relation to the models used, are presented in this section. For more information on the biological background, the reader is referred to [30], while further details on the modeling of the relevant biological processes can be found in [31].

We will consider highly simplified and idealized descriptions of certain small subsystems of a cell. These subsystems will be defined by a set of chemical species, i.e. types of molecules, and a set of reactions. For example, one of the simplest models of expression of a protein might take into account two types of molecules: mRNA molecules and protein molecules. Our descriptions will ignore any spatial effects and operate on the abundances of the chemical species only. Thus, for the simple protein expression model, the system state would be given by the total amount of mRNA and protein in the cell. The system state can be changed by a number of different reactions. For the simple model mentioned above, for instance, we could have:

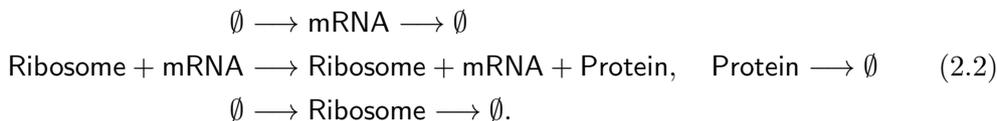
Example 2.1.1.



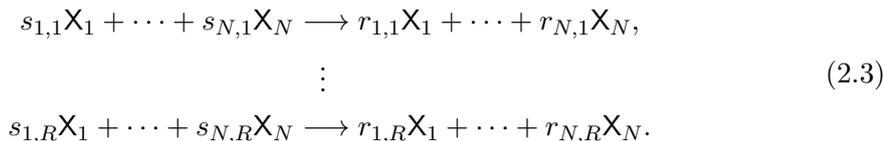
This model includes the creation and degradation of mRNA. Proteins are produced at a rate dependent on the amount of mRNA, and are also degraded.

It is important to take into account that, while we have chosen a particular subsystem of the cell for modeling, that subsystem does in fact not exist in isolation and will generally interact, often significantly, with other parts of the cell.

Example 2.1.2. In the simple gene expression model (2.1), the translation of a protein from an mRNA molecule will be performed by a ribosome. The number of free ribosomes will generally fluctuate, so that a model taking this into account might take the form



Generally, we will be interested in a number of chemical species X_1, \dots, X_N . The state of the system under consideration is given by the vector of copy-numbers $\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{N}_0^N$ of each chemical species. The system state can change via a reaction, during which a certain number of molecules of each species is consumed, and another number of molecules of each species produced. We always assume a finite number R of possible reactions, which are written as



This notation is meant to convey that for the j -th reaction, s_{nj} molecules of species X_n are consumed and r_{nj} molecules produced. The coefficients s_{nj} and r_{nj} are thus non-negative integers, and are referred to as substrate coefficients and product coefficients, respectively.

2.1.1 Continuous-time Markov chain models

Since the system state is discrete, the time evolution has to be in terms of discrete jumps. A Markov assumption then leads us to consider continuous-time Markov chains (CTMC). For a reaction network in state \mathbf{x} , reaction j fires at a rate $h_j(\mathbf{x})$ and leads to a change of the system state by the stoichiometric change vector

$$\boldsymbol{\nu}_j = (\nu_{1j}, \dots, \nu_{Nj}) = (r_{1j} - s_{1j}, \dots, r_{Nj} - s_{Nj}).$$

This means that for the process $\mathbf{X}(t)$ describing the system state, in a small time-interval Δt , the probability of reaction j to fire conditional on the current state \mathbf{x} of the process is given by

$$\Pr(\mathbf{X}(t + \Delta t) = \mathbf{x} + \boldsymbol{\nu}_j \mid \mathbf{X}(t) = \mathbf{x}) = h_j(\mathbf{x})\Delta t + o(\Delta t)$$

for each $j = 1, \dots, R$. Correspondingly, the probability of no reaction during this interval is

$$\Pr(\mathbf{X}(t + \Delta t) = \mathbf{x} \mid \mathbf{X}(t) = \mathbf{x}) = 1 - \Delta t \sum_{j=1}^R h_j(\mathbf{x}) + o(\Delta t).$$

The probability of two or more reactions firing within a sufficiently small time interval is of order $o(\Delta t)$.

In all concrete examples that we consider, we assume mass-action kinetics, so that the h_j are given by

$$h_j(\mathbf{x}) = h_j(\mathbf{x}, \Omega) = c_j \lambda_j(\mathbf{x}, \Omega) = \Omega c_j \prod_{n=1}^N \frac{(x_n)^{s_{nj}}}{\Omega^{s_{nj}}}, \quad (2.4)$$

where $(x)_n = x(x-1)\cdots(x-n+1)$ denotes the falling factorial and c_j is the reaction rate constant. Here we have also introduced the system size Ω , typically the volume of the cell. It is useful to analyze the behavior of the system in terms of the system size, because one expects that under suitable conditions, as $\Omega \rightarrow \infty$, the system dynamics can be reasonably assumed to be deterministic. When the dependence on the system size is of no interest, we implicitly set $\Omega = 1$.

Of particular importance are a subclass of all systems with mass-action kinetics for which all reactions are at most of order two, i.e., for which $\sum_n s_{nj} \leq 2$ for all $j = 1, \dots, R$. These reaction types and their corresponding mass-action rates are shown in Table 2.1. It should also be noted that reactions of order higher than two can be considered as approximations to sequences of at most bi-molecular reactions [32].

Order	Reaction type	Reaction rate in state \mathbf{x}
Zerth-order	$\emptyset \xrightarrow{c} [\dots]$	$c\Omega$
First-order	$X_n \xrightarrow{c} [\dots]$	cx_n
Second-order	$X_n + X_m \xrightarrow{c} [\dots]$ for $n \neq m$	$cx_n x_m / \Omega$
Second-order	$2X_n \xrightarrow{c} [\dots]$	$cx_n(x_n - 1) / \Omega$

Table 2.1: The four different reaction types for a reaction network with mass-action kinetics and at most bi-molecular reactions. The right-hand side of the reactions, indicated by $[\dots]$, can be arbitrary.

The motivation for mass-action kinetics comes from simple combinatorics: For a reaction of the type $X_n + X_m \rightarrow [\dots]$ with $n \neq m$, for instance, there are $x_n x_m$ different possibilities for one molecule of each type to collide, so that the rate should be proportional to $x_n x_m$. For a reaction of the type $2X_n \rightarrow [\dots]$, there are only $x_n(x_n - 1)/2$ possibilities of two molecules of the same type to collide, so that the rate should be proportional to $x_n(x_n - 1)$. Note that these reaction rates have the property that for any reaction that would reduce the number of molecules of any one species to below zero, the corresponding reaction rate is zero.

For a system modeled by a CTMC, the time-marginal probabilities $p_t(\mathbf{x}) = P(\mathbf{X}(t) = \mathbf{x})$ of the system state at some time t are governed by a Kolmogorov-forward equation, in this context called the master equation (ME) or more specifically the chemical master equation (CME),

$$\frac{dp_t(\mathbf{x})}{dt} = \sum_{j=1}^R \{h_j(\mathbf{x} - \boldsymbol{\nu}_j)p_t(\mathbf{x} - \boldsymbol{\nu}_j) - h_j(\mathbf{x})p_t(\mathbf{x})\}. \quad (2.5)$$

When the initial distribution $p_0(\mathbf{x})$ at time 0 is given, the marginal distributions at any later time t can, in principle, be computed using the CME. Note that the CME is an infinite-dimensional system of ordinary differential equations, so that a direct numerical solution is not possible. The most straightforward approach to deal with this problem is to truncate the state-space, leading to the finite state projection algorithm [8]. Unfortunately, the number of variables in the truncated system grows exponentially in the number of chemical species, so that this approach is only feasible for small systems.

2.1.2 Stochastic differential equation models

For some purposes, it is sufficient to consider an approximation of the CTMC model based on a continuous description in terms of concentrations $\mathbf{x} \in \mathbb{R}_{\geq 0}^N$. It has been shown that under certain assumptions, the behavior of the CTMC can be approximated by a stochastic differential equation (SDE), the so-called chemical Langevin equation (CLE) [33], given by

$$d\mathbf{X}(t) = \sum_{j=1}^R h_j(\mathbf{X}(t)) \boldsymbol{\nu}_j dt + \sum_{j=1}^R \sqrt{h_j(\mathbf{X}(t))} \boldsymbol{\nu}_j dW_j(t), \quad (2.6)$$

with independent Wiener processes W_1, \dots, W_R . Thus, there is one driving Wiener process for each reaction of the network.

The marginal probability density is then governed by a Fokker-Planck equation (FPE), in this context called the chemical Fokker-Planck equation (CFPE)

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = - \sum_{n=1}^N \frac{\partial [a_n(\mathbf{x}) p_t(\mathbf{x})]}{\partial x_n} + \frac{1}{2} \sum_{n,m=1}^N \frac{\partial^2 [B_{nm}(\mathbf{x}) p_t(\mathbf{x})]}{\partial x_n \partial x_m}, \quad (2.7)$$

with coefficients

$$\begin{aligned} a_n(\mathbf{x}) &= \sum_{j=1}^R h_j(\mathbf{x}) \nu_{nj}, & n = 1, \dots, N, \\ B_{nm}(\mathbf{x}) &= \sum_{j=1}^R h_j(\mathbf{x}) \nu_{nj} \nu_{mj}, & n, m = 1, \dots, N. \end{aligned} \quad (2.8)$$

Both CME and CFPE have the general form

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = \mathcal{L} p_t(\mathbf{x}), \quad (2.9)$$

where \mathcal{L} is the appropriate (Kolmogorov-forward) evolution operator of either the CME or the CFPE. In this thesis, results will often be based on (2.9) and will then be valid for either description. The hierarchy of models considered in this thesis is shown in Figure 2.1.

Note that in (2.9), we have used the partial derivative notation for time, even though, in the case of the CME, this might have been written as an ordinary derivative. We will do this throughout the thesis when considering equations of the form (2.9). Also, equations of this form will be interchangeably referred to as (Kolmogorov-)forward equations or as evolution equations.

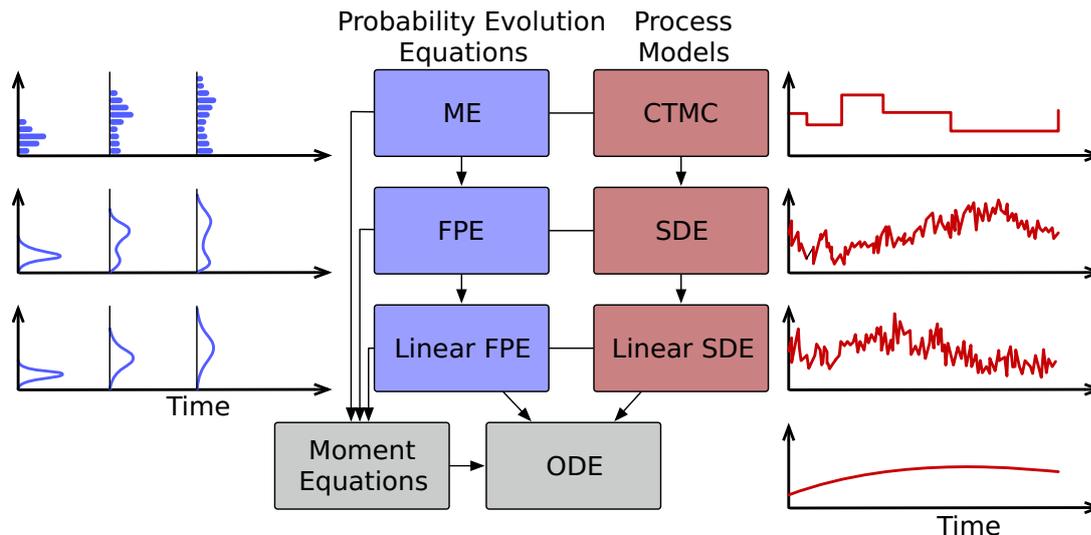


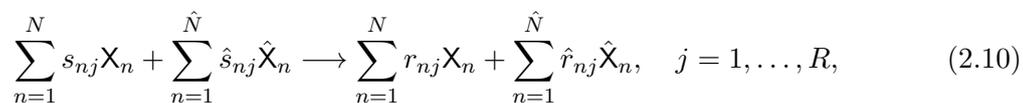
Figure 2.1: The hierarchy of models considered in this thesis. On the right-hand side (red), we have the stochastic process models considered, starting from “exact” CTMC models to SDE models (and the special case of linear SDEs) to ODE models such as the reaction rate equations. On the left-hand side (blue), we have the corresponding forward equations: The master equation and the Fokker-Planck equation (with the special case of a linear Fokker-Planck equation). Moment equations can be derived from each of the (stochastic) models of the hierarchy.

2.2 Modeling heterogeneity

As explained in the introduction, one of the main topics of this thesis is the mathematical treatment of marginal and heterogeneous dynamics.

Cell-to-cell variability or heterogeneity has in recent years emerged as a major component of the behavior of cell populations. It has been demonstrated [34] that understanding the effects of heterogeneity on observed bulk dynamics is critical for the correct interpretation of experimental results. The mathematical treatment of heterogeneity, however, has to date not received much attention in the context of reaction networks.

The most direct method of modeling heterogeneity is to explicitly include the environment of the system of interest into the model. Thus, we consider a reaction network of the form



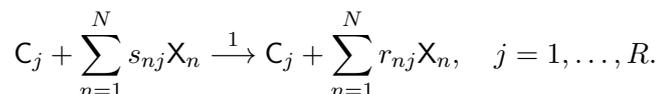
so that, in addition to the chemical species present in the system of interest (2.3), the *subnet*, we have additional species $\hat{X}_1, \dots, \hat{X}_{\hat{N}}$, the *environment*. The process $\mathbf{X} = (X_1, \dots, X_N)$ then describes the state of the subnet species, while $\hat{\mathbf{X}} = (\hat{X}_1, \dots, \hat{X}_{\hat{N}})$ describes the state of the environment species. These species will interact with the species of the subnet via reactions, and might also have reactions among themselves. If

such an approach for the modeling of heterogeneity is chosen, interest will generally focus on the *marginal* subsystem dynamics of the process \mathbf{X} . Note that the process $(\mathbf{X}, \hat{\mathbf{X}})$ is then assumed to be Markovian. For the particularly important case of mass-action kinetics (2.4), the reaction rates factorize into a product of a term depending on the subnet state only and a term depending on the environment state only,

$$h_j(\mathbf{x}, \hat{\mathbf{x}}) = c_j f_j(\mathbf{x}) \hat{f}_j(\hat{\mathbf{x}}), \quad j = 1, \dots, R. \quad (2.11)$$

Often, no knowledge of the source of the heterogeneity will be available. In such cases, it is desirable to have available simple ways of modeling heterogeneity that allow one to probe the dependence of the system dynamics on the heterogeneity. One simple method of doing this, and the one employed in this thesis, is the use of random reaction rate constants c_1, \dots, c_R entering the reaction rates $h_j(\mathbf{x}) = c_j \lambda(\mathbf{x})$. In the simplest case, these are static and assumed to follow a simple distribution such as a Gamma-distribution [35] or a log-normal distribution. It is also useful to assume a simple stochastic process model for the rate constants, an example being the Cox-Ingersoll-Ross (CIR) process [12]. The justification for such a model will be discussed further in the next section.

Whether the heterogeneous rate constants are assumed static or dynamic, it is in principle possible to include them in the state space of the random process so as to consider the process (\mathbf{c}, \mathbf{X}) . The description is then mathematically similar to the explicit modeling of heterogeneity via the process $(\mathbf{X}, \hat{\mathbf{X}})$, although the meaning of the variables is different. For mass-action kinetics specifically, it is useful to note that augmentation leads to a system that is again of mass-action form, if new “virtual” species C_1, \dots, C_R are introduced:



When a description in terms of the augmented process (\mathbf{c}, \mathbf{X}) is chosen and \mathbf{X} is modeled as a CTMC, the state space of the process is a product of a continuous and a discrete state space. The corresponding forward operator will then have both continuous and discrete parts, which however does not cause any difficulties.

In both cases, the inclusion of additional variables increases the dimensionality of the state space, which makes the (approximate) solution of the forward problem much more difficult. Techniques producing reduced descriptions for the marginal dynamics of the process \mathbf{X} are one of the main topics of this thesis.

2.2.1 Modeling heterogeneity using the CIR process

Generally speaking, modeling heterogeneity by means of random rate constants will be done without having any detailed knowledge of the source or properties of the heterogeneity. It is therefore necessary to postulate a general-purpose model. Since fluctuations of a rate constant $c(t)$ are presumably due to a reaction network governing the underlying dynamics, we choose the simplest possible reaction network that will result in a non-trivial stationary distribution, which is a birth-death process. Since we have no knowledge of the absolute abundance of any chemical species that might be involved in

the fluctuations of the rate constant, we choose a continuous description in terms of the CLE and arrive at

$$dc(t) = (a - bc(t))dt + \sqrt{a}dW_1(t) + \sqrt{bc(t)}dW_2(t), \quad (2.12)$$

where a and b are, respectively, the rate constants of creation and decay reactions. This is a scalar SDE (interpreted in the Itô sense) for the time-dependent reaction rate constant c , driven by two independent Wiener processes W_1 and W_2 . To simplify the equation, we can instead consider

$$dc(t) = (a - bc(t))dt + \sqrt{a + bc(t)}dW(t),$$

which has the same Fokker-Planck equation as (2.12) but is driven by a single Wiener process W . This can be further simplified by a change of variables $c(t) \rightarrow c(t) - a/b$. Finally, we choose a more general parametrization in terms of three variables and arrive at the Cox-Ingersoll-Ross (CIR) process [36]

$$dc(t) = \gamma(\mu - c(t))dt + \sigma\sqrt{\frac{2\gamma}{\mu}c(t)}dW(t),$$

which was mentioned above. At stationarity, this process has Gamma-distributed single-time marginals,

$$\text{Gamma}(c \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} c^{\alpha-1} e^{-\beta c} \quad \text{with} \quad \alpha = \frac{\mu^2}{\sigma^2}, \quad \beta = \frac{\mu}{\sigma^2}.$$

Mean and covariance at stationarity are given by

$$\mathbb{E}[c(t)] = \mu \quad \text{and} \quad \mathbb{E}[(c(t) - \mu)(c(t') - \mu)] = \sigma^2 e^{-\gamma|t-t'|},$$

so that the introduction of an extra parameter (in addition to the two parameters of the birth-death process) allows us to separately control mean, variance and the decay of correlations over time. A CIR process with these parameters will be denoted $\text{CIR}(\mu, \sigma^2, \gamma)$. The case of static (Gamma-distributed) heterogeneity is recovered when $\gamma = 0$. A Gamma distribution for static heterogeneity has also been used in [35], where it was shown that a Gamma distribution is particularly convenient for process-level marginalization for CTMC models. Another useful aspect of the CIR process is that the transition distribution over a finite time interval is known (a non-central chi-square distribution) [36]. This makes it possible to simulate the process on a discrete time grid exactly, instead of relying on the Euler-Maruyama algorithm.

The arguments above lead us to consider the CIR process as a general-purpose model for heterogeneous rate constants. Further justification comes from the fact that several marginalization problems can be solved in closed form when CIR-process (or, in the static case, Gamma-distributed) heterogeneity is assumed, as we demonstrate in Chapter 4.

2.3 Approximate solutions of CME and CFPE

The solution of the CME or of the CFPE for a biomolecular reaction network can typically not be obtained in closed form, and numerical approaches are computationally

infeasible for larger systems. One established technique for obtaining approximate solutions is the linear noise approximation [37], which however assumes sufficiently large system sizes. In biological systems, where the abundance of certain chemical species can be very low, the linear noise approximation might become inappropriate. The main alternative is the use of moment equations in combination with moment closure, a method also employed in many other fields [38]. Moment closure approximations have often been referred to as ad-hoc approximations, and it has been shown [39, 40] that they can exhibit unphysical behavior. Finally, it is possible to simulate trajectories from the stochastic process under study. We now briefly describe the above approaches.

2.3.1 Stochastic simulation

A simple and generally applicable approach to investigate the dynamics of a reaction network is by simulation of trajectories of the underlying stochastic process. For a CTMC, this can be done without any approximation. The resulting method is known as the stochastic simulation algorithm (SSA) [3]. To simulate a trajectory of the process starting from state \mathbf{x}_0 at time 0 until time T , one follows the procedure described in Algorithm 1.

Algorithm 1 Stochastic simulation algorithm

```

Set  $t \leftarrow 0, \mathbf{x} \leftarrow \mathbf{x}_0$ .
while  $t < T$  do
  Sample index  $k \in \{1, \dots, R\}$  from discrete distribution with (unnormalized)
  weights  $h_1(\mathbf{x}), \dots, h_R(\mathbf{x})$ .
  Sample waiting time  $w$  from exponential distribution with rate  $\sum_{j=1}^R h_j(\mathbf{x})$ .
  Set  $t \leftarrow t + w, \mathbf{x} \leftarrow \mathbf{x} + \nu_k$ .
end while

```

For large reaction networks or fast reactions, SSA can be very expensive. One approach to reduce the computational burden while retaining the advantages of a simulation-based approach is to employ hybrid system models [41], where only a part of the molecular species are modeled via discrete molecule counts, while other species are modeled via concentrations and simulated using a stochastic differential equation model. Various other approaches to speed up stochastic simulation have also been proposed [5, 6, 7].

2.3.2 Moment equations

In general, obtaining the marginal probability distributions $p_t(\mathbf{x})$ of a stochastic process will be neither feasible nor necessarily desirable. Instead, a useful description of the marginal distributions can be given via a number of expectations $\langle \phi_1 \rangle_t, \dots, \langle \phi_K \rangle_t$ of some functions $\phi_1(\mathbf{x}), \dots, \phi_K(\mathbf{x})$ of the system state. These moments satisfy a system of ordinary differential equations (ODE): Multiplying the evolution equation (2.9) by $\phi_k(\mathbf{x})$ and summing (or, in the case of the CFPE, integrating) over all \mathbf{x} , we obtain

$$\frac{d}{dt} \langle \phi_k \rangle_t = \frac{d}{dt} \sum_{\mathbf{x}} \phi_k(\mathbf{x}) p_t(\mathbf{x}) = \sum_{\mathbf{x}} \phi_k(\mathbf{x}) \mathcal{L} p_t(\mathbf{x}) = \sum_{\mathbf{x}} p_t(\mathbf{x}) \mathcal{L}^\dagger \phi_k(\mathbf{x}) = \left\langle \mathcal{L}^\dagger \phi_k \right\rangle_t. \quad (2.13)$$

Here \mathcal{L}^\dagger is the adjoint of the operator \mathcal{L} with respect to the bilinear form $(p, \phi) := \sum_{\mathbf{x}} p(\mathbf{x})\phi(\mathbf{x})$. It acts on functions $\phi(\mathbf{x})$ of the system state and is given by

$$\mathcal{L}^\dagger \phi(\mathbf{x}) = \sum_{j=1}^R h_j(\mathbf{x}) \{ \phi(\mathbf{x} + \boldsymbol{\nu}_j) - \phi(\mathbf{x}) \} \quad (2.14)$$

for the CME (2.5) and by

$$\mathcal{L}^\dagger \phi(\mathbf{x}) = \sum_{n=1}^N a_n(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x_n} + \frac{1}{2} \sum_{n,m=1}^N B_{nm}(\mathbf{x}) \frac{\partial^2 \phi(\mathbf{x})}{\partial x_n \partial x_m} \quad (2.15)$$

for the CFPE (2.7). Unfortunately, the system of equations (2.13), for $k = 1, \dots, K$, will in general not be closed: The right-hand side will depend on moments other than $\langle \phi_1 \rangle, \dots, \langle \phi_K \rangle$. Attempting to include evolution equations for those moments into the system will introduce a dependence on yet other moments, and so on. The system is infinite-dimensional.

Example 2.3.1. Consider the following reaction network, modeled using mass-action kinetics:



Most commonly, equations for the mean abundances are sought. Here we find

$$\begin{aligned} \frac{d}{dt} \langle x_1 \rangle &= c_1 - c_2 \langle x_1 \rangle + c_3 \langle x_2 \rangle, \\ \frac{d}{dt} \langle x_2 \rangle &= -c_3 \langle x_2 \rangle + c_2 \langle x_1 \rangle - 2c_4 \langle x_2(x_2 - 1) \rangle. \end{aligned} \quad (2.17)$$

This is not a closed system, because the right-hand side involves $\langle x_2^2 \rangle$. We can try to include the variable $\langle x_2^2 \rangle$ into the system of ODEs. The ODEs for all second-order moments read

$$\begin{aligned} \frac{d}{dt} \langle x_1^2 \rangle &= c_1(2 \langle x_1 \rangle + 1) - 2c_2 \langle x_1^2 \rangle + c_2 \langle x_1 \rangle + 2c_3 \langle x_1 x_2 \rangle + c_3 \langle x_2 \rangle, \\ \frac{d}{dt} \langle x_1 x_2 \rangle &= c_1 \langle x_2 \rangle + c_2(\langle x_1^2 \rangle - \langle x_1 x_2 \rangle - \langle x_1 \rangle) + c_3(\langle x_2^2 \rangle - \langle x_1 x_2 \rangle - \langle x_2 \rangle) \\ &\quad - 2c_4 \langle x_1 x_2(x_2 - 1) \rangle, \\ \frac{d}{dt} \langle x_2^2 \rangle &= c_2 \langle x_1(2x_2 + 1) \rangle + c_3 \langle x_2(1 - 2x_2) \rangle - 4c_4 \langle x_2(x_2 - 1)^2 \rangle, \end{aligned} \quad (2.18)$$

which in turn involves the third-order moments $\langle x_2^3 \rangle$ and $\langle x_1 x_2^3 \rangle$.

Note that when $c_4 = 0$, i.e. the last reaction in (2.16) is absent, the system of first-order moment equations *is* closed. Similarly, the system of first- and second-order equations, in that case, is closed. This is the case because when $c_4 = 0$, the system is

linear, in the sense that the hazard functions of all reactions are (affine-)linear functions of the copy numbers, as is seen from the first two rows of Table 2.1.

In general, however, in order to obtain a closed system, some approximation is required. There exist two popular methods, which will be discussed now. We discuss them in the context of equations for “standard” algebraic moments for systems with mass-action kinetics, for which the first- and second-order moment equations have the form

$$\frac{d}{dt} \langle x_n \rangle = \sum_{j=1}^R \langle h_j(\mathbf{x}) \rangle \nu_{nj}, \quad n = 1, \dots, N \quad (2.19)$$

and

$$\frac{d}{dt} \langle x_n x_m \rangle = \sum_{j=1}^R \{ \langle h_j(\mathbf{x}) x_m \rangle \nu_{nj} + \langle h_j(\mathbf{x}) x_n \rangle \nu_{mj} + \langle h_j(\mathbf{x}) \rangle \nu_{nj} \nu_{mj} \}, \quad n, m = 1, \dots, N. \quad (2.20)$$

The system size expansion

The system size expansion [9] is traditionally seen as a systematic expansion of the CME around the deterministic behavior of the system for $\Omega \rightarrow \infty$. Here we instead briefly summarize a recent derivation [42] that operates directly on the moment equations, because that derivation is conceptually similar to the moment closure approximations treated in this thesis. It is based on a series expansion ansatz for the moments in terms of $\Omega^{-1/2}$:

$$\begin{aligned} \frac{\langle x_n \rangle_t}{\Omega} &= \mu_n^{(0)}(t) + \mu_n^{(1)}(t) \Omega^{-1/2} + \mu_n^{(2)}(t) \Omega^{-1} + \dots, & n = 1, \dots, N, \\ \frac{\langle x_n x_m \rangle_t}{\Omega^2} &= S_{nm}^{(0)}(t) + S_{nm}^{(1)}(t) \Omega^{-1/2} + S_{nm}^{(2)}(t) \Omega^{-1} + \dots, & n, m = 1, \dots, N, \end{aligned} \quad (2.21)$$

and similarly for higher-order moments. These equations are then simply inserted into the moment equations (2.19) and (2.20) and analyzed order by order in $\Omega^{-1/2}$, taking into account the scaling of mass-action kinetics in terms of Ω , as given by (2.4). It turns out that the equations close automatically, so that we obtain, for instance,

$$\frac{d}{dt} \mu_n^{(0)} = \sum_{j=1}^R c_j \left\{ \prod_{m=1}^N (\mu_m^{(0)})^{s_{mj}} \right\} \nu_{nj}, \quad n = 1, \dots, N. \quad (2.22)$$

at order Ω^0 . These are the so-called reaction rate equations (RRE), describing the deterministic behavior of the system in the limit $\Omega \rightarrow \infty$.

Example 2.3.2. Consider the reaction network (2.16). The reaction rate equations are given by

$$\begin{aligned} \dot{\mu}_1^{(0)} &= c_1 - c_2 \mu_1^{(0)} + c_3 \mu_2^{(0)}, \\ \dot{\mu}_2^{(0)} &= c_2 \mu_1^{(0)} - c_3 \mu_2^{(0)} - 2c_4 (\mu_2^{(0)})^2. \end{aligned}$$

Unlike the first-order moment equations (2.17), this is evidently a finite-dimensional (closed) system of ODEs.

Moment closure

The assumption of a system size sufficiently large for the application of the system size expansion is not always justified. An alternative approach is the use of a closure ansatz to obtain a finite-dimensional set of equations. The most popular variants are based on an ansatz for the form of the distribution of the solution of the CME. Among these, the ansatz used the most is second-order zero-cumulant closure, where all cumulants of order higher than second are assumed to vanish, as they do for a (multivariate) Gaussian distribution. This implies that third-order moments can be expressed in terms of first- and second-order moments as

$$\langle x_l x_m x_n \rangle = \langle x_l x_m \rangle \langle x_n \rangle + \langle x_m x_n \rangle \langle x_l \rangle + \langle x_n x_l \rangle \langle x_m \rangle - 2 \langle x_l \rangle \langle x_m \rangle \langle x_n \rangle. \quad (2.23)$$

For example, if this approach is applied to the system (2.17) and (2.18), we obtain the closed set of equations

$$\begin{aligned} \frac{d}{dt} \langle x_1 \rangle &= c_1 - c_2 \langle x_1 \rangle + c_3 \langle x_2 \rangle, \\ \frac{d}{dt} \langle x_2 \rangle &= -c_3 \langle x_2 \rangle + c_2 \langle x_1 \rangle - 2c_4 \langle x_2(x_2 - 1) \rangle, \\ \frac{d}{dt} \langle x_1^2 \rangle &= c_1(2 \langle x_1 \rangle + 1) - 2c_2 \langle x_1^2 \rangle + c_2 \langle x_1 \rangle + 2c_3 \langle x_1 x_2 \rangle + c_3 \langle x_2 \rangle, \\ \frac{d}{dt} \langle x_1 x_2 \rangle &= c_1 \langle x_2 \rangle + c_2(\langle x_1^2 \rangle - \langle x_1 x_2 \rangle - \langle x_1 \rangle) + c_3(\langle x_2^2 \rangle - \langle x_1 x_2 \rangle - \langle x_2 \rangle) \\ &\quad + 2c_4 \langle x_1 x_2 \rangle - 2c_4(2 \langle x_1 x_2 \rangle \langle x_2 \rangle + \langle x_2^2 \rangle \langle x_1 \rangle - 2 \langle x_1 \rangle \langle x_2 \rangle^2), \\ \frac{d}{dt} \langle x_2^2 \rangle &= c_2 \langle x_1(2x_2 + 1) \rangle + c_3 \langle x_2(1 - 2x_2) \rangle + 4c_4 \langle 2x_2^2 - x_2 \rangle - 4c_4 \langle x_2(x_2 - 1)^2 \rangle. \end{aligned}$$

Other variants have been discussed in [43, 44], among other works. A different approach, based on obtaining lower and upper bounds for the moments, has recently been developed [45, 46]. Approaches of this latter type are not further discussed in this thesis.

The moment closure approach seems somewhat ad hoc, and it has in fact often been criticized on these grounds. The behavior of zero-cumulant closures in terms of the system size Ω has been discussed in [47]. Conditions for the validity of moment closures have been investigated in [39, 40]. One of the contributions of this thesis is the principled derivation of moment closures, which we present in the next chapter.

Chapter 3

A variational approach to moment closure

This chapter deals with moment closure approximations for CME, CFPE or more generally Kolmogorov-forward equations. We provide a variational derivation of moment closure, which exhibits it as a principled approximation and helps us to understand some of the failure modes typically observed. We also extend entropic matching [48] to general Markov processes and arbitrary approximating distributions and show it to be a special case of variational moment closure. While variational moment closure is a principled approximation, only a subset of all possible closure schemes can be justified by it. On the one hand, we demonstrate that this subset does not suffer from some of the problems often attributed to ad-hoc closure schemes. For this purpose, we introduce mixtures of independent Poisson distributions as a general and useful class of closure distributions. On the other hand, some of the problems attributed to moment closure are not resolved by the variational approach (and are also present in entropic matching). However, our new variational interpretation of moment closure does provide an intuitive explanation for these failure modes. While we present our results in the context of biomolecular kinetics, they are valid more generally for approximations of other Markov processes.

As a first application of moment closures, we consider the problem of obtaining approximate marginal distributions for heterogeneous reaction kinetics (modeled deterministically via the RRE) with log-normally distributed rate constants. It is demonstrated that, by choosing a log-normal ansatz for the joint distribution over rate constants and concentrations, very good approximations can be obtained.

Another aspect of moment equations and moment closures that seems to have not received much attention is the general form of the moment equations, even in the case of mass-action kinetics. Here, a diagrammatic technique is developed allowing one to derive moment (and cumulant) equations of arbitrary order and for arbitrary mass-action reactions in a very simple and transparent way.

We begin by deriving and connecting variational moment closure and entropic matching in Section 3.1. We demonstrate these methods on several examples in Section 3.2. In Section 3.3 we analyze the deficiencies of moment closure, where we also suggest possible solutions. In Section 3.4 we generalize variational moment closure and entropic matching to the approximation of multi-time joint distributions. Variational moment closure is applied to log-normal heterogeneous reaction kinetics in Section 3.5. Finally, the diagrammatic technique for moment and cumulant equations is developed in Section 3.6. This chapter includes material from [24] and [25].

3.1 Variational moment closure and entropic matching

3.1.1 Variational moment closure

In this section, we derive the usual moment closure equations by a variational argument. We consider an evolution equation

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = \mathcal{L}p_t(\mathbf{x}), \quad (3.1)$$

which might be, for instance, the CME or CFPE. Our goal is to approximate the solution of (3.1) at each time t by a member of a parametric family of probability distributions $p_{\boldsymbol{\theta}}(\mathbf{x})$. The parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_L)$ ranges in some open subset of \mathbb{R}^L . Since the solution of (3.1) depends on the time t , a full approximate solution is given by a curve $\boldsymbol{\theta}(t)$. Thus, the time-dependence in the solution $p_t(\mathbf{x})$ of (3.1) is contained in the time-dependence of the parameters $\boldsymbol{\theta}(t)$. Of course, the exact solution will in general not be a member of the chosen parametric family, so we require some means of measuring the approximation error in order to define in which sense the approximation is to be performed.

To obtain the moment-closure equations, this is done in the following way: Begin by choosing a collection of moment functions $\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_K(\mathbf{x}))$. For example, to obtain the usual moment equations of second order, one would choose $K = N + N(N + 1)/2$ monomial moment functions

$$\begin{aligned} \phi_n(\mathbf{x}) &= x_n, & n &= 1, \dots, N, \\ \phi_{nm}(\mathbf{x}) &= x_n x_m, & n, m &= 1, \dots, N, \quad n \leq m. \end{aligned} \quad (3.2)$$

Each of these moment functions can be used to measure the distance between two distributions $p(\mathbf{x})$ and $q(\mathbf{x})$ via the difference between their means $\langle \phi_k(\mathbf{x}) \rangle_{p(\mathbf{x})} - \langle \phi_k(\mathbf{x}) \rangle_{q(\mathbf{x})}$. To turn this difference into a meaningful distance measure, we map it through a function $C : \mathbb{R} \rightarrow [0, \infty)$ to arrive at

$$E_k(p, q) = C \left(\langle \phi_k(\mathbf{x}) \rangle_{p(\mathbf{x})} - \langle \phi_k(\mathbf{x}) \rangle_{q(\mathbf{x})} \right). \quad (3.3)$$

As will be seen below, the precise form of this function is not relevant, as long as it satisfies $C(0) = C'(0) = 0$ and $C''(0) > 0$. For simplicity, we will show the derivation for $C(x) = x^2/2$.

We will derive an ordinary differential equation for the parameter vector $\boldsymbol{\theta}(t)$. To do this, assume that at some time-point t , we have an approximation $p_{\boldsymbol{\theta}(t)}(\mathbf{x})$ of the solution of (3.1) available. Allowing this distribution to evolve a short time Δt using the evolution operator \mathcal{L} of the Markov process, the result

$$p(\mathbf{x}) = p_{\boldsymbol{\theta}(t)}(\mathbf{x}) + \Delta t \mathcal{L}p_{\boldsymbol{\theta}(t)}(\mathbf{x}) + O(\Delta t^2) \quad (3.4)$$

will in general no longer belong to the parametric family. We can try to determine a new approximation $p_{\boldsymbol{\theta}(t+\Delta t)}(\mathbf{x})$ from the parametric family by choosing $\boldsymbol{\theta}(t + \Delta t)$ to minimize (simultaneously) the errors

$$E_k(p_{\boldsymbol{\theta}(t)} + \Delta t \mathcal{L}p_{\boldsymbol{\theta}(t)}, p_{\boldsymbol{\theta}(t+\Delta t)}), \quad k = 1, \dots, K.$$

Of course, choosing such a $\boldsymbol{\theta}(t + \Delta t)$ simultaneously for all error functions will in general not be possible. However, we are in fact only interested in the limit $\Delta t \rightarrow 0$ (in which $p_{\boldsymbol{\theta}(t)} + \Delta t \mathcal{L} p_{\boldsymbol{\theta}(t)} \rightarrow p_{\boldsymbol{\theta}(t)}$) in order to obtain an ordinary differential equation for the parameters $\boldsymbol{\theta}$. Writing for brevity $\boldsymbol{\theta} = \boldsymbol{\theta}(t)$, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(t + \Delta t)$, the resulting equations for $\hat{\boldsymbol{\theta}}$ are

$$\begin{aligned} 0 &= \frac{\partial E_k(p_{\boldsymbol{\theta}} + \Delta t \mathcal{L} p_{\boldsymbol{\theta}}, p_{\hat{\boldsymbol{\theta}}})}{\partial \hat{\theta}_i} \\ &= \left[\langle \phi_k \rangle_{\boldsymbol{\theta}} - \langle \phi_k \rangle_{\hat{\boldsymbol{\theta}}} + \Delta t \langle \mathcal{L}^\dagger \phi_k \rangle_{\boldsymbol{\theta}} \right] \frac{\partial \langle \phi_k \rangle_{\hat{\boldsymbol{\theta}}}}{\partial \hat{\theta}_i}. \end{aligned} \quad (3.5)$$

By $\langle \cdot \rangle_{\boldsymbol{\theta}}$ we denote an expectation taken with respect to $p_{\boldsymbol{\theta}}$.

Dividing (3.5) by Δt and taking the limit, we obtain

$$0 = \left[- \sum_{l=1}^L \frac{\partial \langle \phi_k \rangle_{\boldsymbol{\theta}}}{\partial \theta_l} \dot{\theta}_l + \langle \mathcal{L}^\dagger \phi_k \rangle_{\boldsymbol{\theta}} \right] \frac{\partial \langle \phi_k \rangle_{\boldsymbol{\theta}}}{\partial \theta_i}.$$

We now assume that the matrix

$$F_{kl}(\boldsymbol{\theta}) = \frac{\partial \langle \phi_k \rangle_{\boldsymbol{\theta}}}{\partial \theta_l} = \left\langle \phi_k \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_l} \right\rangle_{\boldsymbol{\theta}} \quad (3.6)$$

is invertible. In particular, the number K of moment functions and the dimension L of the parameter vector $\boldsymbol{\theta}$ are equal and $\nabla_{\boldsymbol{\theta}} \langle \phi_k \rangle_{\boldsymbol{\theta}} \neq \mathbf{0}$ for each k . We then obtain

$$\dot{\boldsymbol{\theta}} = F(\boldsymbol{\theta})^{-1} \langle \mathcal{L}^\dagger \boldsymbol{\phi} \rangle_{\boldsymbol{\theta}}, \quad (3.7)$$

which are the moment-closure equations when parameterized in terms of $\boldsymbol{\theta}$, and when using the distributional ansatz $p_{\boldsymbol{\theta}}$ to close the equations. When we instead introduce the parameters $\boldsymbol{\mu} = \langle \boldsymbol{\phi} \rangle_{\boldsymbol{\theta}}$, we have

$$\dot{\mu}_k = \sum_{l=1}^K \frac{\partial \mu_k}{\partial \theta_l} \dot{\theta}_l = \sum_{l=1}^K \frac{\partial \langle \phi_k \rangle_{\boldsymbol{\theta}}}{\partial \theta_l} \dot{\theta}_l = \sum_{l=1}^K F_{kl}(\boldsymbol{\theta}) \dot{\theta}_l$$

so that we obtain the moment-closure equations in their usual form

$$\dot{\boldsymbol{\mu}} = \langle \mathcal{L}^\dagger \boldsymbol{\phi} \rangle_{\boldsymbol{\mu}}. \quad (3.8)$$

Note again that we could have replaced $C(x) = x^2/2$ by any other function as long as $C(0) = C'(0) = 0$ and $C''(0) > 0$, because we only used the error functions E_k in the limit $\Delta t \rightarrow 0$.

The moment-closure equations can be seen, as is evident from our derivation, as a “greedy” algorithm (to borrow terminology from computer science). The parameter vector $\boldsymbol{\theta}(t)$ evolves to minimize the approximation error (as measured by (3.3)) after an infinitesimal time-step dt , irrespective of the effect this might have on the approximation quality at a later time. This property will help us to understand some of the failure modes of moment-closure approximations in Section 3.3. We note that a different (more complicated) variational justification for moment closure has been given by Eyink [49].

The two ingredients for a moment-closure approximation are thus a choice of parametric distribution $p_{\theta}(\mathbf{x})$ and a choice of moment functions $\phi(\mathbf{x})$. We stress that the space on which these are defined has to be adapted to the Markov process under investigation. Thus, if one is considering the CME (2.5), moment functions and ansatz distribution should be defined on \mathbb{N}_0^N , whereas if one is considering the CFPE, they should be defined on $\mathbb{R}_{\geq 0}^N$. In Section 3.3, we demonstrate that at least some of the problems usually attributed to moment-closure techniques can be explained by a failure to take this into account when choosing a distributional ansatz. For the purpose of this thesis, we will call any moment-closure approach that can be seen as an instance of variational moment closure *principled*. Thus, a moment closure scheme is principled if (i) ansatz distribution and moment functions are defined on the correct state space and (ii) the number K of moment functions and the number L of ansatz distribution parameters are equal and the matrix (3.6) is invertible. Any approach which cannot be justified in this way will be called not principled (or ad-hoc), although of course there might exist other justifications for it. Note also that moment closure, when justifiable via the variational approach, provides an approximation to the full solution of the CME or CFPE. It does not merely provide lower-order moments, as is often asserted for ad-hoc moment-closure approximations.

Our derivation does not make use of the fact that $p_{\theta}(\mathbf{x})$ is a probability distribution. Thus it is possible to allow $p_{\theta}(\mathbf{x})$ to take negative values or to not sum to one. Using parametric families $p_{\theta}(\mathbf{x})$ which can become negative in some (perhaps negligible) part of the state space provides one with more flexibility when choosing an appropriate ansatz. Also note that the derivation above remains valid when the moment functions depend on the variational parameters, i.e. $\phi(\mathbf{x}) = \phi_{\theta}(\mathbf{x})$, which will be necessary to establish the connection to entropic matching. In this case, the error functions (3.3) are defined using the parameter value $\theta(t)$ when deriving the evolution equation from time t to $t + \Delta t$.

3.1.2 Entropic matching

While the approach described in Section 3.1.1 is very flexible, it does not provide any indication of how to choose the ansatz distribution $p_{\theta}(\mathbf{x})$ and the moment functions $\phi(\mathbf{x})$. Here we extend entropic matching [48], which is an approximation method based on information-theoretic considerations, to arbitrary ansatz distributions and processes. As we will see, entropic matching turns out to be a special case of variational moment closure, which provides a natural choice of moment functions $\phi(\mathbf{x})$ for any choice of distribution $p_{\theta}(\mathbf{x})$. The relationships between the various approximations which we derive here and in the following sections are depicted graphically in Figure. 3.1.

The idea underlying entropic matching is the same as was demonstrated in our derivation of the moment-closure equations, which in fact was inspired by entropic matching. The difference lies in the distance measure used. From an information-theoretic point of view, there are strong arguments [50] for using the relative entropy

$$D[p \parallel q] = \left\langle \ln \frac{p(\mathbf{x})}{q(\mathbf{x})} \right\rangle_{p(\mathbf{x})}$$

as a general-purpose distance measure to a distribution $p(\mathbf{x})$ when approximating it by a distribution $q(\mathbf{x})$. We now follow the same approach as in the derivation of variational

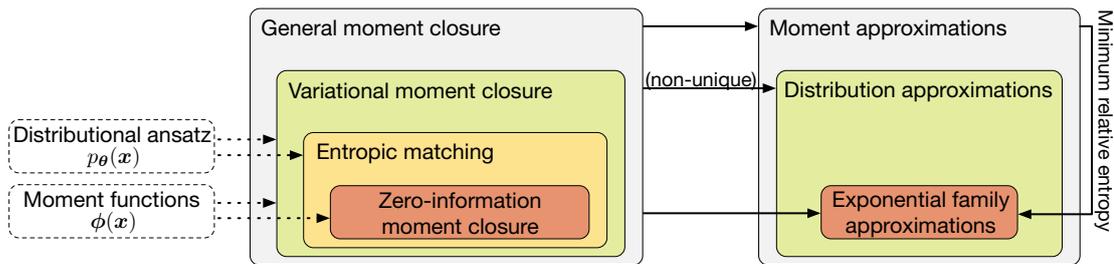


Figure 3.1: Relation between general (ad-hoc) moment closure, variational moment closure, entropic matching and zero-information moment closure. Dashed arrows indicate the building blocks for each approximation: Variational moment closures are based on a choice of moment functions $\phi(\mathbf{x})$ and a choice of ansatz distribution $p_{\theta}(\mathbf{x})$. Entropic matching requires only a choice of ansatz distribution, while ZI moment closure requires only a choice of moment functions. In this sense, entropic matching and ZI moment closure are dual to each other. Full arrows indicate the type of approximation provided by each approximation method: Ad-hoc closures provide only approximations of low-order moments, whereas variational moment closures (and the special cases entropic matching and ZI moment closure) provide approximations of the full distribution. However, ad-hoc closures can provide approximations of the full distributions via minimum relative entropy.

moment closure, using the relative entropy as distance measure. We again try to find an approximate solution to (3.1) within a parametric family $p_{\theta}(\mathbf{x})$. Assume that, at some time t , an approximating distribution is available and specified by the parameters $\theta(t)$. A small time Δt later, the distribution is again given by (3.4). Entropic matching proceeds by approximating this distribution by $p_{\theta(t+\Delta t)}(\mathbf{x})$, where $\theta(t+\Delta t)$ is chosen to minimize the relative entropy $D[p \parallel p_{\theta(t+\Delta t)}]$ between $p(\mathbf{x})$ and $p_{\theta(t+\Delta t)}(\mathbf{x})$. Writing again $\theta = \theta(t)$, $\hat{\theta} = \theta(t+\Delta t)$ and using (3.4), the relative entropy is given, up to order one in Δt , by

$$\begin{aligned} D[p \parallel p_{\hat{\theta}}] &= \left\langle \ln \frac{p_{\theta} + \Delta t \mathcal{L} p_{\theta}}{p_{\hat{\theta}}} \right\rangle_{p(\mathbf{x})} \\ &= \left\langle \ln \frac{p_{\theta}}{p_{\hat{\theta}}} \right\rangle_{\theta} + \Delta t \left[\left\langle \frac{\mathcal{L} p_{\theta}}{p_{\theta}} \right\rangle_{\theta} + \left\langle \frac{\mathcal{L} p_{\theta}}{p_{\theta}} \ln \frac{p_{\theta}}{p_{\hat{\theta}}} \right\rangle_{\theta} \right] \\ &= D[p_{\theta} \parallel p_{\hat{\theta}}] + \Delta t \left[\left\langle \frac{\mathcal{L} p_{\theta}}{p_{\theta}} \right\rangle_{\theta} + \left\langle \frac{\mathcal{L} p_{\theta}}{p_{\theta}} \ln \frac{p_{\theta}}{p_{\hat{\theta}}} \right\rangle_{\theta} \right]. \end{aligned}$$

The relative entropy between two members of a parametric family with parameters $\theta, \hat{\theta}$ is given, to second order in $\hat{\theta} - \theta$, by

$$D[p_{\theta} \parallel p_{\hat{\theta}}] = \frac{1}{2} (\hat{\theta} - \theta)^{\dagger} G(\theta) (\hat{\theta} - \theta), \quad (3.9)$$

where $G(\theta) = [G_{kl}(\theta)]$ is the Fisher information matrix of the parametric distribution

at parameter value $\boldsymbol{\theta}$, given by

$$G_{kl}(\boldsymbol{\theta}) = \left\langle \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_k} \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_l} \right\rangle_{\boldsymbol{\theta}}.$$

Using this, we obtain

$$0 = \nabla_{\hat{\boldsymbol{\theta}}} D[p \parallel p_{\hat{\boldsymbol{\theta}}}] = G(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \Delta t \left\langle \nabla_{\hat{\boldsymbol{\theta}}} \ln p_{\hat{\boldsymbol{\theta}}} \frac{\mathcal{L} p_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}}} \right\rangle_{\hat{\boldsymbol{\theta}}}.$$

Dividing by Δt and taking the limit, we obtain

$$G(\boldsymbol{\theta})\dot{\boldsymbol{\theta}} = \left\langle (\nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}) \frac{\mathcal{L} p_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}}} \right\rangle_{\boldsymbol{\theta}} = \left\langle \mathcal{L}^{\dagger} \nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}}$$

so that the final evolution equation for the parameters $\boldsymbol{\theta}$ reads

$$\dot{\boldsymbol{\theta}} = G(\boldsymbol{\theta})^{-1} \left\langle \mathcal{L}^{\dagger} \nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}}. \quad (3.10)$$

We can now see how we can obtain the entropic matching equation (3.10) within the variational approach to moment closure: We choose the moment functions $\phi_k = \partial \ln p_{\boldsymbol{\theta}} / \partial \theta_k$. Then

$$F_{kl}(\boldsymbol{\theta}) = \left\langle \phi_k \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_l} \right\rangle_{\boldsymbol{\theta}} = \left\langle \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_k} \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_l} \right\rangle_{\boldsymbol{\theta}} = G_{kl}(\boldsymbol{\theta})$$

so that equations (3.10) and (3.7) agree. Note that, in general, we here require the more general form of variational moment closure in which the moment functions depend on the variational parameters $\boldsymbol{\theta}$. It is also worth noting that, although the relative entropy is non-symmetric in its arguments, the final evolution equation for the parameters would be the same had we tried to minimize $D[p_{\boldsymbol{\theta}(t+\Delta t)} \parallel p]$ instead. This is because infinitesimally, the relative entropy is given by (3.9), which is symmetric.

It turns out that (3.10) has appeared in the literature previously [51, 52], where it was derived by directly defining a projection using the Fisher information metric. This was done in the context of stochastic filtering equations, which can be considered to be a generalization of a Markovian evolution equation when observations of the stochastic process are included. Arguably, our derivation could be considered more principled because it does not postulate the use of the Fisher information, but rather starts out with the minimization of the relative entropy, for which as explained above there are strong arguments. Here we should also mention that more generally, in the context of filtering equations, moment closure approximations are well-known under the name ‘‘assumed density filtering’’ [51].

It is also interesting to connect entropic matching to a more advanced information-theoretic approximation for stochastic processes. Instead of approximating the marginal distributions of the solution of the forward equation (3.1), it is possible to approximate the distribution over trajectories of the stochastic process. This can be done by minimizing the relative entropy of an approximating stochastic process relative to the process of interest over the space of all trajectories. Such an approach has been carried out for a number of processes [53, 54, 55, 15] and one would expect such an approximation to

perform better than entropic matching. The latter can be understood (for the same reasons already given for variational moment closure) to be a “greedy” algorithm, which at each time point chooses parameters which minimize the relative entropy at an infinitesimally later time point while not taking into account any later time points. Additionally, entropic matching produces only approximations to single-time marginal distributions and does not provide information about multi-time correlations. This latter deficiency is addressed in Section 3.4. The advantage of entropic matching is that it only requires computations using marginal distributions as given by (3.10). Variational approximations on the process level, on the other hand, are presumably tractable only for a very small number of processes.

3.1.3 Zero-information moment closure

It is worthwhile to connect entropic matching with another approach to moment closure motivated by information-theoretic considerations [56, 57]. When choosing a distributional ansatz to complement a set of moment functions ϕ_1, \dots, ϕ_K , it seems reasonable to choose the maximum entropy distribution associated with these functions, given by

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{Z(\boldsymbol{\theta})} \exp \left\{ \sum_{k=1}^K \theta_k \phi_k(\mathbf{x}) \right\} p_0(\mathbf{x}). \quad (3.11)$$

This is consistent with having the moments $\langle \phi(\mathbf{x}) \rangle_{\boldsymbol{\theta}}$ of the distribution, and no other information available. The parameters $\boldsymbol{\theta}$ can be computed from the given moments $\langle \phi(\mathbf{x}) \rangle_{\boldsymbol{\theta}}$, and $Z(\boldsymbol{\theta})$ is a normalization constant. This approach has been termed zero-information (ZI) closure [57]. Here we have additionally introduced a “background” measure $p_0(\mathbf{x})$. This is useful since on an infinite state space, the entropy has in general to be replaced by the relative entropy with respect to some background measure. Equation (3.11) is then the distribution of minimum relative entropy to $p_0(\mathbf{x})$. Note that maximization of entropy has to be replaced by minimization of relative entropy because relative entropy is defined with the inverse sign.

Entropic matching and ZI moment closure can be seen to be somewhat dual to each other, as is illustrated in Figure 3.1: Entropic matching starts out with a choice of ansatz distribution and supplies a natural choice of moment functions, whereas ZI closure starts out with a choice of moment functions and supplies a natural choice of parametric distribution. Nevertheless, entropic matching is a generalization of ZI closure, as was first noticed for the case of stochastic differential equations [51]. To see this, we check that when entropic matching is applied with a distribution of the form (3.11), the result is equivalent to ZI moment closure. For a distribution of the form (3.11), we have $\partial \ln p_{\boldsymbol{\theta}}(\mathbf{x}) / \partial \theta_k = \phi_k(\mathbf{x}) - \langle \phi_k(\mathbf{x}) \rangle_{\boldsymbol{\theta}}$. For the matrix $F(\boldsymbol{\theta})$ from (3.6), we thus obtain

$$\begin{aligned} F_{kj}(\boldsymbol{\theta}) &= \left\langle \phi_k \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_j} \right\rangle_{\boldsymbol{\theta}} = \langle \phi_k (\phi_j - \langle \phi_j \rangle_{\boldsymbol{\theta}}) \rangle_{\boldsymbol{\theta}} \\ &= \langle (\phi_k - \langle \phi_k \rangle_{\boldsymbol{\theta}}) (\phi_j - \langle \phi_j \rangle_{\boldsymbol{\theta}}) \rangle_{\boldsymbol{\theta}} \\ &= G_{kj}(\boldsymbol{\theta}), \end{aligned}$$

i.e. $F(\boldsymbol{\theta})$ is equal to the Fisher information matrix $G(\boldsymbol{\theta})$. Similarly, we have

$$\left\langle \mathcal{L}^\dagger \phi_k \right\rangle_{\boldsymbol{\theta}} = \left\langle \mathcal{L}^\dagger (\phi_k - \langle \phi_k \rangle_{\boldsymbol{\theta}}) \right\rangle_{\boldsymbol{\theta}} = \left\langle \mathcal{L}^\dagger \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_k} \right\rangle_{\boldsymbol{\theta}},$$

so that (3.7) and (3.10) agree.

We mention in passing that by using the differential equations (3.7) for the parameters of (3.11), the computationally expensive explicit minimization of the relative entropy [57] does not have to be performed [56].

3.1.4 Non-uniqueness of the distributional ansatz and minimum relative entropy

Variational moment closure (and in particular its special case entropic matching) produces an approximation for the distribution of the solution of (3.1), as opposed to only approximations for a number of moments. Ad-hoc moment-closure schemes, on the other hand, are usually considered to only produce approximations for the moments. As explained above, a well-known method to reconstruct full probability distributions from moments is maximum entropy (or more generally minimum relative entropy), and this approach has been used previously in the context of moment equations [56, 58]. Here, we briefly point out why such an approach remains meaningful even when a variational moment-closure scheme is employed.

Consider a reaction network with polynomial (e.g. mass-action (2.4)) reaction kinetics and assume that we use the standard monomial moment functions (3.2) or their high-order analogues. Choosing a variational ansatz $p_{\boldsymbol{\theta}}(\boldsymbol{x})$, the moment equations (3.8) only ever require a finite number of relations between the moments up to a certain order. Thus, the distributional ansatz is not uniquely specified by the moment-closure equations derived from it. This implies that it is not a priori clear how the approximate solution of the CME should be reconstructed from the variational parameters $\boldsymbol{\theta}$ obtained as the solution to (3.7), and minimum relative entropy provides one possible answer. The minimum relative entropy distribution again has the form (3.11) (note however that the parameters $\boldsymbol{\theta}$ governing the minimum relative entropy distribution in (3.11) are not the same as the parameters of the moment equations (3.7)). As for ZI moment closure, we mention that an explicit minimization of relative entropy is not necessary [58], because differential equations for the parameters of the minimum relative entropy distribution can be derived [56]. These equations can then be solved simultaneously with the moment equations (3.7).

3.2 Examples of variational moment closure and entropic matching

In this section, we demonstrate the approximations derived in Section 3.1 on several examples. Since the standard moment equations using monomial moment functions (3.2) are well-known, we here focus on cases where interesting connections to other existing approximations can be established.

3.2.1 Product-Poisson entropic matching for the CME

A first interesting result is obtained by using a product-Poisson ansatz

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \prod_{n=1}^N e^{-\theta_n} \frac{\theta_n^{x_n}}{x_n!}$$

to approximately solve the CME for mass-action kinetics (2.4) via entropic matching. A product-form ansatz might seem to be very restrictive, but it has been shown that a certain class of networks actually has product-form distributions at stationarity [59, 60]. The class of networks for which a product-Poisson distribution at some initial time remains of this form for later times has also recently been characterized [61]. Using the backwards evolution operator (2.14) for the CME, we have

$$\begin{aligned} \mathcal{L}^\dagger \frac{\partial}{\partial \theta_n} \ln p_{\boldsymbol{\theta}}(\mathbf{x}) &= \mathcal{L}^\dagger \frac{\partial}{\partial \theta_n} \sum_{m=1}^N \{x_m \ln \theta_m - \theta_m - \ln x_m!\} \\ &= \frac{1}{\theta_n} \sum_{j=1}^R h_j(\mathbf{x}) \nu_{nj}. \end{aligned}$$

The factorial moments of a Poisson distribution are given by $\langle (x)_s \rangle_{\boldsymbol{\theta}} = \theta^s$, which allows us to evaluate the right-hand side of (3.10). The Fisher information matrix of a product Poisson distribution is diagonal, $G_{mn}(\boldsymbol{\theta}) = \delta_{mn} \theta_n^{-1}$. Combining these results, it turns out that the entropic matching equations are identical with the macroscopic reaction rate equations (RRE)

$$\dot{\boldsymbol{\theta}} = \sum_{j=1}^R c_j \theta_1^{s_{1j}} \cdots \theta_N^{s_{Nj}} \boldsymbol{\nu}_j. \quad (3.12)$$

This implies that the macroscopic rate equations have a meaning even at arbitrary low system sizes. Product-Poisson entropic matching is an instance of ZI moment closure, where however we have to choose a non-trivial background measure

$$p_0(\mathbf{x}) = \frac{e^{-N}}{x_1! \cdots x_N!}.$$

For some systems, certain molecular species might only exist in zero or one copy, such as a gene that can be either in the active or in the inactive state. In this case, an analogous result can be obtained by using a Bernoulli distribution instead of a Poisson distribution for the species in question.

3.2.2 The finite state projection algorithm for the CME

Variational moment closure encompasses approximations not usually considered as moment closure. For example, taking the ansatz

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{\mathbf{x}' \in \mathbb{X}} \delta_{\mathbf{x}, \mathbf{x}'} \theta_{\mathbf{x}'} \quad (3.13)$$

for some finite subset $\mathbb{X} \subset \mathbb{N}_0^N$, and the family of moment functions

$$\phi_{\mathbf{x}}(\mathbf{x}') = \delta_{\mathbf{x}, \mathbf{x}'}, \quad \mathbf{x} \in \mathbb{X},$$

one recovers the finite state projection algorithm [8] for the numerical solution of the CME on the finite subset \mathbb{X} of states. Here we have made use of the fact that within the variational moment closure framework, $p_{\theta}(\mathbf{x})$ does not necessarily have to sum to one. This is required because the finite state projection approach “leaks” probability into the part of the state space outside of \mathbb{X} . More generally, the moment functions $\delta_{\mathbf{x}, \mathbf{x}'}$ could be replaced by more general basis functions, leading to Galerkin-type approximations [62].

3.2.3 Gaussian entropic matching for the Fokker-Planck equation

Entropic matching was originally demonstrated [48] for the case of a stochastic differential equation (SDE)

$$d\mathbf{X}_t = \mathbf{a}(\mathbf{x})dt + S(\mathbf{x})d\mathbf{W}_t \quad (3.14)$$

with a Gaussian distributional ansatz, where however the diffusion matrix S was assumed to not depend on the state \mathbf{x} . Using our formula, we can extend this result to state-dependent diffusion matrices.

We consider the FPE with backwards evolution operator (2.15) corresponding to the (Itô) stochastic differential equation (3.14). The Gaussian ansatz with parameters $\theta = (\boldsymbol{\mu}, \Sigma)$ is given by

$$\ln p_{\theta}(\mathbf{x}) = -\frac{1}{2} \ln \det(2\pi\Sigma) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\dagger} \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}).$$

The Fisher information matrix of a Gaussian with respect to the parameters $(\boldsymbol{\mu}, \Sigma)$ is block diagonal, with one block corresponding to $\boldsymbol{\mu}$ and given by Σ^{-1} , and the other block corresponding to Σ and given by

$$\left\langle \frac{\partial^2 \ln p_{\theta}}{\partial \Sigma_{ij} \partial \Sigma_{kl}} \right\rangle = \frac{1}{2}(\Sigma^{-1})_{ik}(\Sigma^{-1})_{jl}.$$

We evaluate the right hand side of (3.10) and obtain

$$\begin{aligned} \left\langle \mathcal{L}^{\dagger} \frac{\partial \ln p_{\theta}}{\partial \mu_n} \right\rangle &= [\Sigma^{-1} \langle \mathbf{a}(\mathbf{x}) \rangle]_n, \\ \left\langle \mathcal{L}^{\dagger} \frac{\partial \ln p_{\theta}}{\partial \Sigma_{nm}} \right\rangle &= \left\langle \mathcal{L}^{\dagger} (\mathbf{x} - \boldsymbol{\mu})^{\dagger} \Sigma^{-1} E_{(nm)} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\rangle \\ &= \frac{1}{2} \left\langle \mathbf{a}(\mathbf{x})^{\dagger} \Sigma^{-1} E_{(nm)} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\rangle + \frac{1}{2} \left\langle (\mathbf{x} - \boldsymbol{\mu})^{\dagger} \Sigma^{-1} E_{(nm)} \Sigma^{-1} \mathbf{a}(\mathbf{x}) \right\rangle \\ &\quad + \frac{1}{2} (\Sigma^{-1} \langle B(\mathbf{x}) \rangle \Sigma^{-1})_{nm} \\ &= \frac{1}{2} (\langle A(\mathbf{x}) \rangle \Sigma^{-1})_{nm} + \frac{1}{2} (\Sigma^{-1} \langle A(\mathbf{x}) \rangle^{\dagger})_{nm} + \frac{1}{2} (\Sigma^{-1} \langle B(\mathbf{x}) \rangle \Sigma^{-1})_{nm} \end{aligned}$$

where $E_{(nm)}$ is a matrix with the entry “1” in the n -th row and m -th column, and zeros otherwise. We also used, in the last line, the equation

$$\left\langle u \frac{\partial v}{\partial x_m} \right\rangle + \left\langle v \frac{\partial u}{\partial x_m} \right\rangle = \sum_{n=1}^N \langle uv(\Sigma^{-1})_{mn}(x_n - \mu_n) \rangle$$

for averages of functions $u(\mathbf{x}), v(\mathbf{x})$ with respect to a Gaussian distribution with parameters $(\boldsymbol{\mu}, \Sigma)$. Multiplying by the inverse of the Fisher information matrix now yields

$$\begin{aligned} \dot{\boldsymbol{\mu}} &= \langle \mathbf{a}(\mathbf{x}) \rangle, \\ \dot{\Sigma} &= \langle A(\mathbf{x}) \rangle \Sigma + \Sigma \langle A(\mathbf{x}) \rangle^\dagger + \langle B(\mathbf{x}) \rangle, \end{aligned} \tag{3.15}$$

where $A(\mathbf{x})$ is the Jacobian of $\mathbf{a}(\mathbf{x})$, $B(\mathbf{x}) = S(\mathbf{x})S(\mathbf{x})^\dagger$ and the expectation values are taken with respect to the Gaussian approximation. When $B(\mathbf{x}) = B$ is state-independent, this agrees with the previous results [48].

It is interesting to note that our result agrees with a process-level approximation [15] (more precisely, with the single-time marginals) mentioned in Section 3.1.2. However, since those results were obtained using a complex form of the relative entropy involving auxiliary variables, it is not clear whether our result would also agree with a probabilistic approach using the standard, real relative entropy.

3.2.4 Log-normal entropic matching for the CFPE

For applications in chemical kinetics, a Gaussian ansatz is somewhat unsatisfactory because the ansatz distribution should be restricted to have support on $\mathbb{R}_{\geq 0}^N$. An alternative is to employ a multivariate log-normal distribution. It turns out that the form of the equations (3.15) for the parameters $\boldsymbol{\mu}$ and Σ (now parameterizing a log-normal distribution) remain the same. However, the coefficients $\mathbf{a}(\mathbf{x})$ and $B(\mathbf{x})$ in (3.15) have to be replaced by

$$\begin{aligned} \hat{a}_n(\mathbf{x}) &= a_n(e^{\mathbf{x}})e^{-x_n} - \frac{1}{2}e^{-2x_n}B_{nn}(e^{\mathbf{x}}), & n &= 1, \dots, N, \\ \hat{B}_{nm}(\mathbf{x}) &= B_{nm}(e^{\mathbf{x}})e^{-(x_n+x_m)}, & n, m &= 1, \dots, N, \end{aligned}$$

where for a vector \mathbf{x} , we write $e^{\mathbf{x}} = (e^{x_1}, \dots, e^{x_N})$. Note that when we express the entropic matching equations in terms of these modified coefficients, the expectations in (3.15) have to be computed with respect to a Gaussian distribution with parameters $(\boldsymbol{\mu}, \Sigma)$, even though these parameters govern a log-normal distribution. If we employ mass-action kinetics (2.4), the resulting expectations have closed form expressions. This will be used in Section 3.5 to obtain entropic matching equations for log-normal heterogeneity.

3.3 Analysis of moment-closure schemes and the Poisson correction

General, ad-hoc moment-closure schemes have been shown to exhibit some unphysical properties [39, 40], and attempts to understand these pathologies have so far been un-

successful. In this section, we use the variational interpretation of moment closure from Section 3.1 to analyze moment-closure schemes and some their failure modes.

As a first observation, recall that the finite state projection algorithm was shown to be an instance of variational moment closure in Section 3.2.2. This simple observation has the important implication that it is not the moment-closure concept as such that leads to unphysical behavior of the resulting approximation. The finite state projection algorithm will under mild conditions produce an approximation to the exact solution of the CME whose error can be made arbitrary small. Thus, it is the flexibility of the approximating ansatz distribution (and the choice of moment functions) which determines whether moment closure will produce an acceptable approximation. The same is true for the RRE, which were also shown to be an instance of ZI closure in Section 3.2.1.

One typical failure mode of ad-hoc moment closures is the divergence of the solutions, especially in the low copy-number regime. In order to address this problem, we recall that, as explained in Section 3.1.1, the distributional ansatz has to be supported on \mathbb{N}_0^N or $\mathbb{R}_{\geq 0}^N$, depending on whether one considers the CME or the CFPE. This is not difficult to achieve for the CFPE, but it is not obvious for many of the closure schemes introduced previously [44] for the CME. One might conjecture that the divergences observed at low copy numbers might be related to the failure to choose an ansatz distribution with the correct support. In order to investigate this, we first have to introduce a sufficiently flexible family of distributions with support on \mathbb{N}_0^N , which we do in the following section. The approach we present allows one to turn most existing closure methods for the CME into principled closures in the sense of Section 3.1. This property makes it particularly convenient to compare ad-hoc closure methods to their principled counterparts.

In this section, we focus on the conventional moment equations using the moment functions (3.2) or their higher-order analogues. We also restrict attention to mass-action kinetics (2.4).

3.3.1 Poisson mixtures

A flexible approach for defining analytically tractable distributions on \mathbb{N}_0^N is the mixing of independent distributions on \mathbb{N}_0 using a distribution on $\mathbb{R}_{\geq 0}^N$. In the context of chemical kinetics, a natural choice for the discrete mixing distributions is the Poisson distribution. The general form of the distributional assumption that we consider will thus be of the form

$$p(\mathbf{x} \mid \boldsymbol{\theta}) = \int_{\mathbb{R}_{\geq 0}^N} d\boldsymbol{\lambda} p(\boldsymbol{\lambda} \mid \boldsymbol{\theta}) e^{-(\lambda_1 + \dots + \lambda_N)} \frac{\lambda_1^{x_1} \dots \lambda_N^{x_N}}{x_1! \dots x_N!}, \quad (3.16)$$

where $p(\boldsymbol{\lambda} \mid \boldsymbol{\theta})$ is the mixing distribution governed by parameters $\boldsymbol{\theta}$. Using Poisson mixtures for the exact or approximate solution of the CME (the Poisson representation) is a well-known approach [63]. Here we use it to define a flexible family of distributions for moment closure. A log-normal mixture of Poisson distributions, used within Eyink's variational framework [49], was proposed previously and motivated using ideas from statistical physics [64]. It is however important to realize that this approach is simply moment closure using a log-normal-Poisson mixture ansatz.

A situation in which we will be particularly interested in is the case when $p(\boldsymbol{\lambda} \mid \boldsymbol{\theta})$ is a distribution that has previously been directly employed for moment closure. This is the case, for example, for the Gamma and multivariate log-normal distributions [44],

since these have support on $\mathbb{R}_{\geq 0}^N$ and can thus serve as mixing distributions. It is also the case for univariate zero-cumulant (ZC) closure of order two, although this is not immediately obvious. To perform moment closure, we have to obtain expressions for the expectation values of monomials in \mathbf{x} for $p(\mathbf{x} | \boldsymbol{\theta})$. Using conditional expectations, these are expressed as

$$\langle x_1^{\alpha_1} \cdots x_N^{\alpha_N} \rangle_{\boldsymbol{\theta}} = \langle \mathbb{E}[x_1^{\alpha_1} \cdots x_N^{\alpha_N} | \boldsymbol{\lambda}] \rangle_{p(\boldsymbol{\lambda} | \boldsymbol{\theta})}$$

where $\mathbb{E}[\cdot | \boldsymbol{\lambda}]$ denotes the expectation with respect to a product-Poisson distribution with mean $\boldsymbol{\lambda}$. For moments of order one to three, for instance, we obtain the equations

$$\begin{aligned} \langle x_n \rangle_{\boldsymbol{\theta}} &= \langle \lambda_n \rangle_{p(\boldsymbol{\lambda} | \boldsymbol{\theta})}, \\ \langle x_n x_m \rangle_{\boldsymbol{\theta}} &= \langle \lambda_n \lambda_m \rangle_{p(\boldsymbol{\lambda} | \boldsymbol{\theta})} + \delta_{nm} \langle x_n \rangle_{\boldsymbol{\theta}}, \\ \langle x_i x_n x_m \rangle_{\boldsymbol{\theta}} &= \langle \lambda_i \lambda_n \lambda_m \rangle_{p(\boldsymbol{\lambda} | \boldsymbol{\theta})} - 2\delta_{in} \delta_{nm} \langle x_i \rangle_{\boldsymbol{\theta}} + \delta_{mi} \langle x_i x_n \rangle_{\boldsymbol{\theta}} + \delta_{in} \langle x_n x_m \rangle_{\boldsymbol{\theta}} + \delta_{nm} \langle x_m x_i \rangle_{\boldsymbol{\theta}}. \end{aligned} \quad (3.17)$$

In general, all moments of a product-Poisson distribution are polynomials in the parameters $\boldsymbol{\lambda}$. This implies that if the moment-closure equations using some distribution $p(\boldsymbol{\lambda} | \boldsymbol{\theta})$ for polynomial reaction kinetics can be computed in closed form, then this will also be the case for the Poisson mixture with mixing distribution $p(\boldsymbol{\lambda} | \boldsymbol{\theta})$. Additionally, we see that the expressions in (3.17) defining the closure scheme are merely corrected by polynomials in lower-order moments when moving from a closure using $p(\boldsymbol{\lambda} | \boldsymbol{\theta})$ to a Poisson- $p(\boldsymbol{\lambda} | \boldsymbol{\theta})$ mixture closure. For this reason, we will use the term *Poisson correction* when referring to this case.

One might expect that the difference between a moment-closure scheme defined using a distribution on $\mathbb{R}_{\geq 0}^N$ and the corresponding Poisson-corrected closure might be more pronounced for low copy numbers of the chemical species. Indeed, many moment-closure schemes are known to diverge in this regime. In the following section, we empirically demonstrate that this problem can be explained by the failure to take into account the discreteness of molecule counts when choosing a moment-closure scheme.

3.3.2 A single-species system

We will investigate the properties of Poisson corrections on the single-species system



We concentrate on the standard second-order moment equations, which for this system are given by

$$\begin{aligned} \dot{\mu}_1 &= a\Omega + c(\mu_1 - \mu_2)/\Omega + b\mu_1, \\ \dot{\mu}_2 &= a\Omega + (2a\Omega + b - c/\Omega)\mu_1 + (2b + 3c/\Omega)\mu_2 - 2c\mu_3/\Omega. \end{aligned} \quad (3.19)$$

Here μ_1, μ_2, μ_3 are the non-centered moments of order one to three and Ω is the system size as defined in (2.4). A moment closure scheme is then given by specifying the third-order moment as a function of the first- and second-order moments, $\mu_3 = V(\mu_1, \mu_2)$. We focus on three popular ad-hoc closures: The zero-cumulant closure

$$V_{ZC}(\mu_1, \mu_2) = 3\mu_2\mu_1 - 2\mu_1^3, \quad (3.20)$$

the log-normal closure

$$V_{\text{LN}}(\mu_1, \mu_2) = (\mu_2/\mu_1)^3 \quad (3.21)$$

and the Gamma closure

$$V_{\text{G}}(\mu_1, \mu_2) = \mu_2(2\mu_2 - \mu_1^2)/\mu_1. \quad (3.22)$$

When applying the Poisson correction, a moment closure function $V(\mu_1, \mu_2)$ is transformed to

$$\hat{V}(\mu_1, \mu_2) = V(\mu_1, \mu_2 - \mu_1) + 3\mu_2 - 2\mu_1.$$

The result of applying these moment closures on the system is shown in Figure 3.2. We see that without Poisson correction, the ad-hoc closures diverge for sufficiently small system sizes. This does not happen with the Poisson correction. We do however see that the quality of the Poisson-corrected closures breaks down at sufficiently small sizes. For these findings to support our hypothesis, we have to verify that the three ad-hoc closures do not correspond to any distributions defined on \mathbb{N}_0 . While these closures are defined via distributions on $\mathbb{R}_{\geq 0}$ or (in the case of zero-cumulant closure) \mathbb{R} , we have to remember that they only specify a relation of the form $\mu_3 = V(\mu_1, \mu_2)$, and there might exist distributions on \mathbb{N}_0 for which the same relation holds. We now show to what extent this is the case.

We are interested in the following questions: (i) Given a pair (μ_1, μ_2) , under which conditions does there exist a distribution on \mathbb{N}_0 with these moments of first and second order? This defines the domain where a moment closure scheme should ideally be defined. It also corresponds to the domain of valid initial conditions for moment equations, a fact that has often been neglected in previous studies. (ii) Assuming that (μ_1, μ_2) does correspond to a distribution on \mathbb{N}_0 , and in addition given μ_3 , under which conditions does there exist a distribution on \mathbb{N}_0 with moments of orders one to three given by (μ_1, μ_2, μ_3) ? Existence questions of this type have recently been answered [65]. Obviously, if $\mu_1 = 0$, the distribution is degenerate and concentrated on 0. Thus, in the following, assume $\mu_1 > 0$. Then it turns out that the well-known condition $\mu_2 \geq \mu_1^2$ is not sufficient for the existence of a distribution with the prescribed moments on \mathbb{N}_0 . Instead, a necessary and sufficient condition is given by [65]

$$\mu_2 - \mu_1^2 \geq \{\mu_1\}(1 - \{\mu_1\}), \quad (3.23)$$

where for any real y , we write $\{y\} = y - \lfloor y \rfloor$ with $\lfloor y \rfloor$ the greatest integer smaller or equal to y , so that $\{y\}$ is the fractional part of y . Now assume that (3.23) holds, and that in addition we are given μ_3 . Then in order for there to exist a distribution on \mathbb{N}_0 with moments (μ_1, μ_2, μ_3) , a necessary and sufficient condition is [65]

$$\frac{\mu_3}{\mu_1} - \left(\frac{\mu_2}{\mu_1}\right)^2 \geq \left\{\frac{\mu_2}{\mu_1}\right\} \left(1 - \left\{\frac{\mu_2}{\mu_1}\right\}\right). \quad (3.24)$$

Additionally, if equality holds in (3.23), all higher-order moments are uniquely determined. In particular, μ_3 is then determined by requiring equality in (3.24),

$$\frac{\mu_3}{\mu_1} - \left(\frac{\mu_2}{\mu_1}\right)^2 = \left\{\frac{\mu_2}{\mu_1}\right\} \left(1 - \left\{\frac{\mu_2}{\mu_1}\right\}\right). \quad (3.25)$$

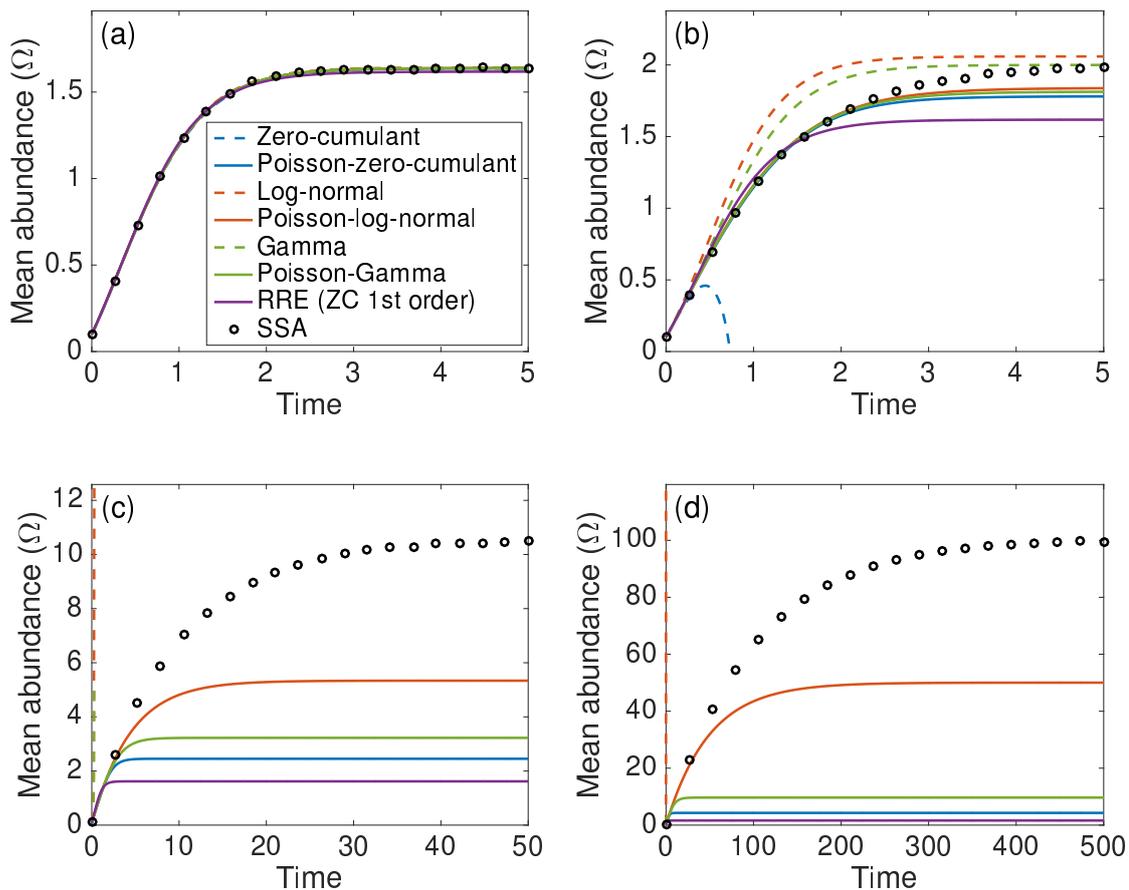


Figure 3.2: Moment closure predictions of mean abundance for (3.19) at various system sizes for zero-cumulant, log-normal and Gamma closures and their Poisson-corrections. RRE solutions are given for comparison. Abundances are shown in units of Ω . Exact means obtained from 10,000 SSA realizations. Parameters $a = b = c = 1$. The initial conditions correspond to a Poisson distribution with mean $\Omega/10$. (a) At $\Omega = 10$, all moment closures show negligible error. (b) At $\Omega = 1$, the zero-cumulant closure has diverged. (c) At $\Omega = 10^{-1}$, all closures without Poisson correction have diverged. (d) At $\Omega = 10^{-2}$, approximation quality reduces for all closures except Poisson-log-normal, but the Poisson-corrected closures do not diverge.

Any principled second-order moment closure V has to satisfy these conditions when μ_3 is replaced by $V(\mu_1, \mu_2)$. Since we are primarily interested in the low copy-number regime, we note that for $\mu_1 \leq 1$, the domain (3.23) of valid pairs (μ_1, μ_2) is simply characterized by $\mu_2 \geq \mu_1$. If we in fact have $\mu_2 = \mu_1$, the resulting equality constraint in (3.25) reduces to $\mu_3 = \mu_1$.

We proceed to apply these results to the three ad-hoc closure schemes presented above. The regions where (3.24) is not satisfied (while (3.23) is satisfied) are shown in Figure 3.3. For zero-cumulant closure, the inequality constraint (3.24) is not satisfied for most pairs (μ_1, μ_2) of region (3.23). For log-normal and Gamma closures, on the

other hand, all pairs (μ_1, μ_2) in the region shown do satisfy (3.24). We would expect that this is reflected in the behavior of these moment closures, and this is verified in Figure 3.3, where we show the phase plots for the moment equations (3.19) in the region of low copy numbers, after the application of moment closure. Zero-cumulant closure shows pathological behavior for a wide range of initial conditions (μ_1, μ_2) , whereas log-normal and Gamma closure are well-behaved for a large sub-domain of initial conditions. However, even for these closures, unphysical behavior does occur. To understand this, we have to take into account the equality constraint (3.25) for those pairs (μ_1, μ_2) for which equality holds in (3.23). Focusing on the low copy-number regime $\mu_1 \leq 1$, we are concerned with the line $\mu_2 = \mu_1$, and we observe from Figure 3.3 that it is precisely along

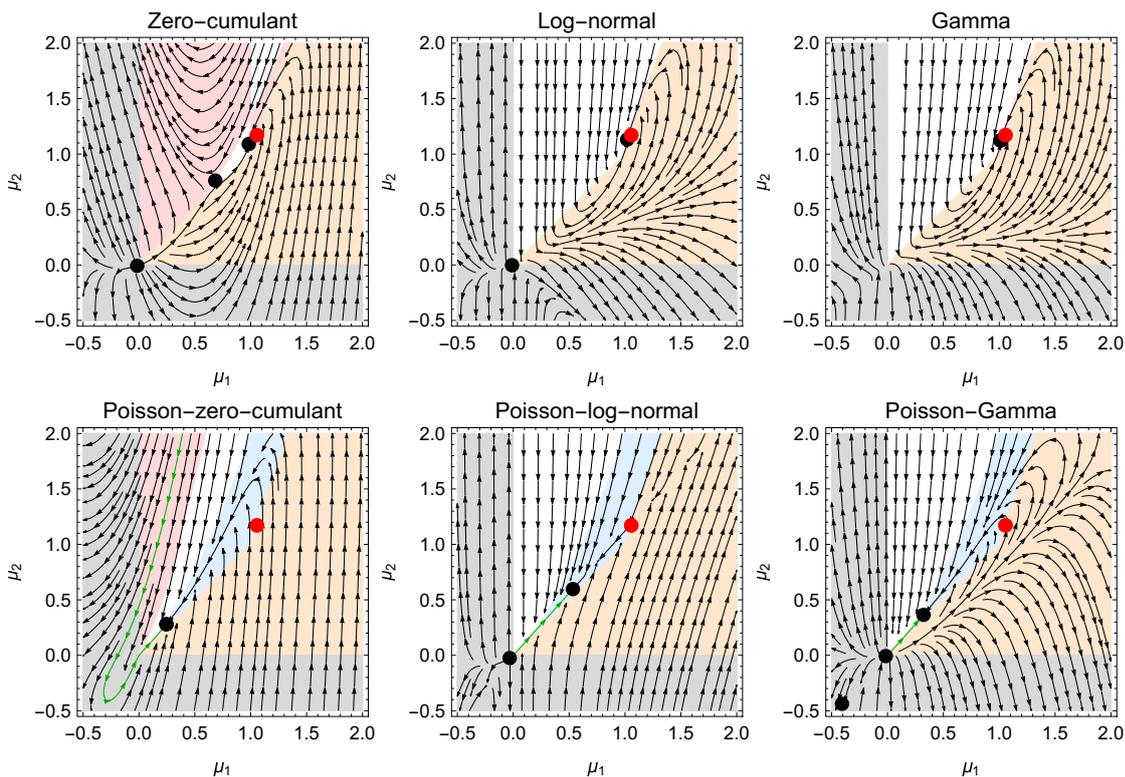


Figure 3.3: Phase-plot for (3.19) for zero-cumulant, log-normal and Gamma closures (top row) and their Poisson-corrected versions (bottom row). Red point: Moments of exact stationary distribution. Black point(s): Stationary solution(s) of the moment equations. Grey area: Region where $\mu_1 < 0$ or $\mu_2 < 0$, shown for better visibility of trajectories. Orange area: Region where (3.23) is violated. Red area: Region where (3.23) is valid but (3.24) is violated. Blue area: Region where (3.23) is valid but the distribution is sub-Poissonian. Only the white area corresponds to valid initial conditions for the moment equations. For the Poisson-corrected closures, the green phase curve runs along the domain boundary $\mu_1 = \mu_2$, preventing phase curves from entering the orange region as they do in the ad-hoc closures. Parameters were $a = b = c = 1$ and $\Omega = 0.1$.

this line that unphysical behavior occurs for log-normal and Gamma closures. Indeed, from (3.21) and (3.22), we find that $V_{\text{LN}}(\mu_1, \mu_2) = 1$ and $V_{\text{G}}(\mu_1, \mu_2) = (2 - \mu_1)\mu_1$ when $\mu_1 = \mu_2$, violating (3.25). In general, the pathologies of the ad-hoc closures seem to correspond quite closely to what one would expect.

We now investigate to what extent these pathologies are removed by applying a Poisson correction. We first have to determine whether the three ad-hoc closures (3.20), (3.21) and (3.22) would be compatible with distributions on $\mathbb{R}_{\geq 0}$, because they play the role of mixing distributions in (3.16). This is clear for the log-normal and Gamma closures. For the zero-cumulant closure, this is an instance of the truncated Stieltjes moment problem, and one can check that the relation $\mu_3 = V_{\text{ZC}}(\mu_1, \mu_2)$ is compatible with a distribution on $\mathbb{R}_{\geq 0}$ as long as $\mu_1^2 \leq \mu_2 \leq 2\mu_1^2$. This is a relatively small domain, and the behavior visible in Figure 3.3 is compatible with these findings: Poisson-correction does remove divergences in a relatively small domain. For log-normal and Gamma closures, we observe from Figure 3.3 that the divergences close to the line $\mu_1 = \mu_2$ are removed by applying a Poisson correction. Indeed, one immediately checks that, for $\mu_1 \leq 1$, we have $\hat{V}_{\text{LN}}(\mu_1, \mu_2) = \mu_1$ and $\hat{V}_{\text{GM}}(\mu_1, \mu_2) = \mu_1$, so that the Poisson-corrected versions do satisfy the equality constraint (3.25), as of course they have to.

Our findings suggest that the failure to choose distributions with a support adapted to the CME for moment closure provides a good explanation for the divergences often observed in the low copy-number regime, and that the Poisson correction prevents this from happening. One further observation from Figure 3.3 is that the stationary points of the Poisson-corrected closures show a larger error (relative to the exact stationary points) than for the ad-hoc closures. This is explained by the fact that the system (3.18) at stationarity has a sub-Poissonian distribution, a property that cannot be realized by a mixture of Poisson distributions. From this, we also see that unphysical behavior of the equations on the one hand, and approximation quality on the other hand are not necessarily correlated. For completeness, note that the opposite case of distributions with heavier tails than a Poisson distribution can be realized by using an appropriate mixing distribution.

3.3.3 Other failure modes

Apart from divergences, there are other important failure modes of moment closures. Here we briefly indicate how the variational derivation of moment closures makes these failure modes at least plausible. The CME, under mild conditions, will have a unique stationary distribution $p(\mathbf{x})$ where

$$\mathcal{L}p(\mathbf{x}) = 0.$$

One would then hope that any approximation should retain this property. It has however been observed [39, 40] that moment closures can possess either multiple stationary states (within a meaningful region of parameters space), or show sustained oscillations and thus fail to converge to a stationary state for certain initial conditions.

However, using our derivation of variational moment closures (or the conceptually analogous derivation of entropic matching), these properties are not surprising and have an intuitive explanation. Since these approximations can be seen as “greedy” algorithms, choosing the best approximation locally after each infinitesimal time-step, there is no

guarantee that the effects which lead to a single stationary distribution can be captured by the approximation. Thus, for instance, for a system with oscillatory trajectories, the fact that a unique stationary distribution exists is a consequence of the fact that different trajectories do not stay in phase as time progresses. If an approximation is performed after each infinitesimal time-step, however, this effect is not necessarily captured.

3.4 Multi-time joint distributions

In this section, we extend variational moment closure and entropic matching to multi-time joint distributions and investigate the conditions under which the resulting approximation is consistent with the single-time approximations derived in Section 3.1.

3.4.1 Multi-time moment closure and entropic matching

Our variational derivation of moment-closure approximations can be extended to an approximation for joint probability distributions at multiple time points. We here focus on two-time joint distributions, although the same approach extends to higher-order joint distributions.

Fix an initial time-point t_0 and consider the two-time joint distribution $p_{t_1, t_0}(\mathbf{x}^1, \mathbf{x}^0)$ for which we want to obtain an approximation for $t_1 > t_0$. In order to derive a variational approximation, we again choose an ansatz distribution $p_{\theta}(\mathbf{x}^1, \mathbf{x}^0)$, which now has to depend on two arguments. For example, any of the Poisson-mixture distributions introduced in Section 3.3 could be used, where now they have to be defined over a space of dimension $2N$. Similarly, we choose a set of moment functions $\phi(\mathbf{x}^1, \mathbf{x}^0) = (\phi_1(\mathbf{x}^1, \mathbf{x}^0), \dots, \phi_K(\mathbf{x}^1, \mathbf{x}^0))$ which now have to depend on two arguments. For example, in analogy to the standard moment functions (3.2), we could define the $K = 2N + N(N + 1) + N^2$ moment functions

$$\begin{aligned} \phi_n^0(\mathbf{x}^1, \mathbf{x}^0) &= x_n^0, & n &= 1, \dots, N, \\ \phi_n^1(\mathbf{x}^1, \mathbf{x}^0) &= x_n^1, & n &= 1, \dots, N, \\ \phi_{nm}^{00}(\mathbf{x}^1, \mathbf{x}^0) &= x_n^0 x_m^0, & n, m &= 1, \dots, N, n \leq m, \\ \phi_{nm}^{10}(\mathbf{x}^1, \mathbf{x}^0) &= x_n^1 x_m^0, & n, m &= 1, \dots, N, \\ \phi_{nm}^{11}(\mathbf{x}^1, \mathbf{x}^0) &= x_n^1 x_m^1, & n, m &= 1, \dots, N, n \leq m. \end{aligned} \tag{3.26}$$

These moment functions are again used to define distances between distributions $p(\mathbf{x}^1, \mathbf{x}^0)$ and $q(\mathbf{x}^1, \mathbf{x}^0)$,

$$E_k(p, q) = \frac{1}{2} \left[\langle \phi_k(\mathbf{x}^1, \mathbf{x}^0) \rangle_p - \langle \phi_k(\mathbf{x}^1, \mathbf{x}^0) \rangle_q \right]^2. \tag{3.27}$$

Here as in Section 3.1.1 we could again use a more general function $C : \mathbb{R} \rightarrow [0, \infty)$ instead of $(\cdot)^2/2$ to apply to the difference $\langle \phi_k \rangle_p - \langle \phi_k \rangle_q$ without changing the result. The derivation of the moment-closure approximation proceeds analogously to the case of single-time marginal distributions. Assuming again that we have available an approximation $p_{\theta(t_1, t_0)}(\mathbf{x}^1, \mathbf{x}^0)$ for some $t_1 \geq t_0$, the evolution of this distribution over a short time-interval Δt is given by

$$p(\mathbf{x}^1, \mathbf{x}^0) = p_{\theta}(\mathbf{x}^1, \mathbf{x}^0) + \Delta t p_{\theta}(\mathbf{x}^0) \mathcal{L}_1 p_{\theta}(\mathbf{x}^1 | \mathbf{x}^0) + o(\Delta t).$$

Here $p_{\boldsymbol{\theta}}(\mathbf{x}^0)$ and $p_{\boldsymbol{\theta}}(\mathbf{x}^1 | \mathbf{x}^0)$ are, respectively, the marginal and conditional distributions corresponding to $p_{\boldsymbol{\theta}}(\mathbf{x}^1, \mathbf{x}^0)$. The subscript “1” on the operator \mathcal{L}_1 indicates that it acts only on the argument \mathbf{x}^1 , and not on \mathbf{x}^0 . We try to approximate the distribution p by a member of the chosen parametric family by minimizing the distance functions (3.27). Writing again $\boldsymbol{\theta} = \boldsymbol{\theta}(t_1, t_0)$, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(t_1 + \Delta t, t_0)$, the resulting equations for the minima are

$$\begin{aligned} 0 &= \frac{\partial E_k(p_{\boldsymbol{\theta}} + \Delta t p_{\boldsymbol{\theta}}, p_{\hat{\boldsymbol{\theta}}})}{\partial \hat{\theta}_i} \\ &= \left[\langle \phi_k(\mathbf{x}^1, \mathbf{x}^0) \rangle_{\boldsymbol{\theta}} - \langle \phi_k(\mathbf{x}^1, \mathbf{x}^0) \rangle_{\hat{\boldsymbol{\theta}}} + \Delta t \langle \mathcal{L}_1^\dagger \phi_k(\mathbf{x}^1, \mathbf{x}^0) \rangle_{\boldsymbol{\theta}} \right] \frac{\partial \langle \phi_k(\mathbf{x}^1, \mathbf{x}^0) \rangle_{\hat{\boldsymbol{\theta}}}}{\partial \hat{\theta}_i}, \end{aligned}$$

where \mathcal{L}_1^\dagger again acts only on the argument \mathbf{x}^1 . Dividing by Δt and taking the limit, we obtain as in Section 3.1.1

$$\dot{\boldsymbol{\theta}} = F(\boldsymbol{\theta})^{-1} \left\langle \mathcal{L}_1^\dagger \phi(\mathbf{x}^1, \mathbf{x}^0) \right\rangle_{\boldsymbol{\theta}}, \quad (3.28)$$

assuming again that the matrix

$$F_{kj}(\boldsymbol{\theta}) = \left\langle \phi_k(\mathbf{x}^1, \mathbf{x}^0) \frac{\partial \ln p_{\boldsymbol{\theta}}(\mathbf{x}^1, \mathbf{x}^0)}{\partial \theta_j} \right\rangle_{\boldsymbol{\theta}}$$

is invertible. Note that (3.28) are equations for the evolution of $\boldsymbol{\theta} = \boldsymbol{\theta}(t_1, t_0)$ in t_1 , while t_0 is fixed. We can also re-parameterize the equations in terms of the moments $\boldsymbol{\mu} = \langle \phi \rangle_{\boldsymbol{\theta}}$ to obtain

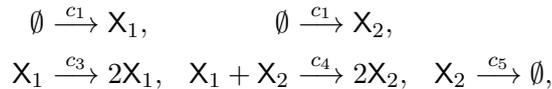
$$\dot{\boldsymbol{\mu}} = \left\langle \mathcal{L}_1^\dagger \phi(\mathbf{x}^1, \mathbf{x}^0) \right\rangle_{\boldsymbol{\mu}}. \quad (3.29)$$

We see that the moment-closure equations for two-time joint probability distributions are completely analogous to the case of single-time distributions (3.7). The same reasoning can be applied to obtain multi-time joint distributions for entropic matching. The derivation proceeds along similar lines, so that we only state the resulting equation

$$\dot{\boldsymbol{\theta}} = G(\boldsymbol{\theta})^{-1} \left\langle \mathcal{L}_1^\dagger \nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{x}^1, \mathbf{x}^0) \right\rangle_{\boldsymbol{\theta}}. \quad (3.30)$$

For completeness, note that computer implementations for the solution of (3.7) or (3.8) are easily reused for the solution of (3.28) or (3.29). This is because for a reaction system modeled by the CME or the CFPE, the multi-time moment closure equations (3.28) and (3.30) are in fact equivalent to the corresponding single-time equations (3.7) and (3.10) applied to an augmented system: In addition to the species and reactions of the original system (2.3), we introduce N species X_1^0, \dots, X_N^0 that do not participate in any reactions. Then the single-time equations applied to this $2N$ -dimensional system using an ansatz distribution $p_{\boldsymbol{\theta}}(\mathbf{x}^1, \mathbf{x}^0)$ and moment functions $\phi(\mathbf{x}^1, \mathbf{x}^0)$ will be equivalent to the two-time equations.

We proceed to demonstrate the two-time moment closure equations numerically. An interesting case for multi-time correlations are oscillatory systems. We use the Lotka-Volterra model



where we added input reactions for each species to prevent explosion and extinction events. Results for the two-time covariances $\langle x_m^1 x_n^0 \rangle_{\theta(t,0)} - \langle x_m^1 \rangle_{\theta(t,0)} \langle x_n^0 \rangle_{\theta(t,0)}$ are shown in Figure 3.4.

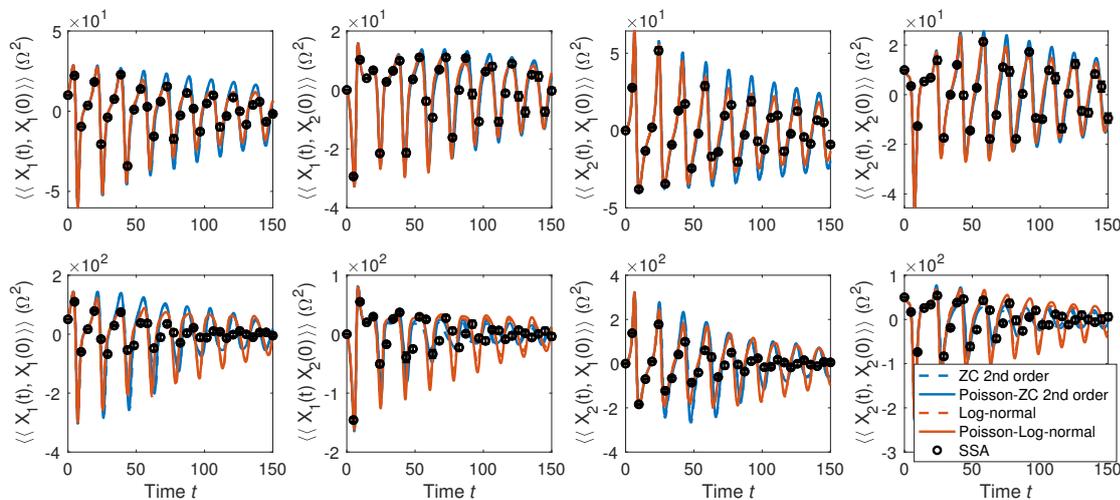


Figure 3.4: Two-time covariance functions $\langle X_1(t)X_1(0) \rangle - \langle X_1(t) \rangle \langle X_1(0) \rangle$, $\langle X_1(t)X_2(0) \rangle - \langle X_1(t) \rangle \langle X_2(0) \rangle$, $\langle X_2(t)X_1(0) \rangle - \langle X_2(t) \rangle \langle X_1(0) \rangle$ and $\langle X_2(t)X_2(0) \rangle - \langle X_2(t) \rangle \langle X_2(0) \rangle$ (corresponding to columns from left to right) as a function of time t . System size $\Omega = 5$ (top row) and $\Omega = 1$ (bottom row). The approximation error at the lower system size is presumably due to the inadequateness of a second-order closure, regardless of whether a Poisson correction is employed or not. Parameters values were $c_1 = 1, c_2 = 1, c_3 = 0.5, c_4 = 0.003, c_5 = 0.3$. Initial conditions were product-Poisson with means $\langle X_1(0) \rangle = \langle X_2(0) \rangle = 50\Omega$. Black circles correspond to the covariances computed from 100,000 SSA realizations. Error bars of SSA estimates were computed by dividing the SSA samples into 10 equally sized parts, computing the covariances estimates for each part and plotting ± 1 standard deviation of these estimates.

3.4.2 Consistency

An important question that arises is whether the equations (3.28) or (3.30) produce consistent approximations when applied jointly with the single-time approximations (3.7) or (3.10). The following condition should be satisfied: Starting from a single-time marginal distribution $p_{t_0}(\mathbf{x}^0)$ at time t_0 , we can use (3.7) to produce approximations to the single-time marginals $p_{t_1}(\mathbf{x}^1)$ for any $t_1 > t_0$. We can also use (3.28) to produce approximations to the two-time joint distributions $p_{t_1, t_0}(\mathbf{x}^1, \mathbf{x}^0)$, which implies a marginal distribution over \mathbf{x}^1 at time t_1 . These two marginal distributions at time t_1 should agree. Of course, this can only hold if the chosen ansatz distributions (which are defined over spaces of different dimensions) and the moment functions are compatible in a sense to be defined below. In general, the moment functions $\phi(\mathbf{x}^1, \mathbf{x}^0)$ can be grouped into functions

depending on both or on just one of the arguments,

$$\phi^1(\mathbf{x}^1), \quad \phi^{10}(\mathbf{x}^1, \mathbf{x}^0), \quad \phi^0(\mathbf{x}^0).$$

Consider a family $p_{\boldsymbol{\mu}}(\mathbf{x}^1, \mathbf{x}^0)$ parameterized in terms of the moments $\boldsymbol{\mu} = (\boldsymbol{\mu}^1, \boldsymbol{\mu}^{10}, \boldsymbol{\mu}^0)$,

$$\boldsymbol{\mu}^1 = \langle \phi^1 \rangle_{\boldsymbol{\mu}}, \quad \boldsymbol{\mu}^{10} = \langle \phi^{10} \rangle_{\boldsymbol{\mu}}, \quad \boldsymbol{\mu}^0 = \langle \phi^0 \rangle_{\boldsymbol{\mu}}.$$

Denote by $p_{\boldsymbol{\mu}}^1(\mathbf{x}^1)$ and $p_{\boldsymbol{\mu}}^0(\mathbf{x}^0)$ the resulting marginal distributions over \mathbf{x}^1 and \mathbf{x}^0 , respectively. We now assume that $p_{\boldsymbol{\mu}}^1(\mathbf{x}^1)$ can be parameterized in terms of $\boldsymbol{\mu}^1$ only, and similarly $p_{\boldsymbol{\mu}}^0(\mathbf{x}^0)$ in terms of $\boldsymbol{\mu}^0$ only, i.e.

$$p_{\boldsymbol{\mu}}^1(\mathbf{x}^1) = p_{\boldsymbol{\mu}^1}^1(\mathbf{x}^1), \quad p_{\boldsymbol{\mu}}^0(\mathbf{x}^0) = p_{\boldsymbol{\mu}^0}^0(\mathbf{x}^0).$$

This is the case, for instance, for the multivariate Gaussian and log-normal distributions, and then also for a log-normal mixture of independent Poissons as introduced in Section 3.3. Then we immediately see that

$$\dot{\boldsymbol{\mu}}^1 = \left\langle \mathcal{L}^\dagger \phi^1(\mathbf{x}^1) \right\rangle_{\boldsymbol{\mu}} = \left\langle \mathcal{L}^\dagger \phi^1(\mathbf{x}^1) \right\rangle_{\boldsymbol{\mu}^1} \quad (3.31)$$

because the moment functions ϕ^1 do not depend on \mathbf{x}^0 , and the marginal distribution $p_{\boldsymbol{\mu}}^1(\mathbf{x}^1)$ only depends on $\boldsymbol{\mu}^1$. This is a closed equation for the parameters of the marginal distribution $p_{\boldsymbol{\mu}^1}^1(\mathbf{x}^1)$. If, on the other hand, we use the parametric family $p_{\boldsymbol{\mu}^1}^1(\mathbf{x}^1)$ and the moment functions $\phi^1(\mathbf{x}^1)$ for the single-time moment equations (3.7), we also obtain (3.31). Thus, the two-time moment equations using the parametric family $p_{\boldsymbol{\mu}}(\mathbf{x}^1, \mathbf{x}^0)$ are consistent with the single-time moment equations using the parametric family $p_{\boldsymbol{\mu}^1}^1(\mathbf{x}^1)$ as required. We similarly see that

$$\dot{\boldsymbol{\mu}}^0 = \left\langle \mathcal{L}_1^\dagger \phi^0(\mathbf{x}^0) \right\rangle_{\boldsymbol{\mu}} = \mathbf{0},$$

because ϕ^0 does not depend on \mathbf{x}^1 . Thus, the approximation to the marginal distribution at time t_0 does not change (i.e. it remains equal to the initially supplied marginal distribution $p_{t_0}^0(\mathbf{x}^0)$), as one would expect.

Because of the consistency property explained above, our variational approximation is a viable alternative to the process-level variational approximations mentioned in Section 3.1.2. Our approximation is tractable for a wide variety of ansatz distributions, whereas closed-form process-level approximations can presumably only be obtained for a very restricted class of ansatz stochastic processes.

3.5 Log-normal entropic matching for heterogeneous kinetics

In this section, variational moment closure is used to treat heterogeneous reaction rate equations when the heterogeneous rate parameters follow a log-normal distribution. It is shown numerically and analytically that log-normal heterogeneity of rate constants appears to be propagated into log-normal distributions of species abundances with high accuracy for various systems.

3.5.1 Derivation

Our starting point is a general reaction network (2.3) with mass-action kinetics (2.4), modeled deterministically via the RREs (2.22). As explained in Section 2.2, one way to treat heterogeneity mathematically is to incorporate the random reaction rate constants c_1, \dots, c_R into the state space of the process. The resulting reaction network on an extended state space still has mass-action kinetics. We can then restrict our attention to the case of heterogeneous (i.e., random) initial conditions for mass-action kinetics, and assume fixed reaction rate constants with value 1. We denote the state of the augmented system by $\mathbf{y} = (y_1, \dots, y_M)$, so that we have $M = N + R$ and $\mathbf{y} = (x_1, \dots, x_N, c_1, \dots, c_R)$. The augmented system is then governed by an RRE for a mass-action system. Such an RRE is fully specified by the stoichiometric change vectors and substrate coefficients of the network. For the augmented system, we denote these by $\tilde{\mathbf{v}}_j = (\tilde{v}_{1j}, \dots, \tilde{v}_{Mj}) = (\mathbf{v}_j, \mathbf{0})$ and $\tilde{\mathbf{s}}_j = (\tilde{s}_{1j}, \dots, \tilde{s}_{Mj}) = (\mathbf{s}_j, \mathbf{e}_j)$, respectively. Here \mathbf{v}_j is the stoichiometric change vector and $\mathbf{s}_j = (s_{1j}, \dots, s_{Nj})$ the vector of substrate coefficients of the original (non-augmented) system as defined in (2.3), and \mathbf{e}_j is the j -th standard basis vector in \mathbb{R}^R .

In order to apply entropic matching, we consider the Liouville equation associated with the RRE of the augmented system, which is obtained by dropping the diffusion term in the corresponding CFPE (2.7) of the augmented system. We can immediately apply the results in Section 3.2.4 to perform log-normal entropic matching on this system. Denoting by $\boldsymbol{\mu}$ and Σ the parameters of the log-normal ansatz distribution for the augmented system, the equations for these parameters, using (3.15), take the form

$$\begin{aligned} \dot{\mu}_m &= \sum_{j=1}^R b_{jm}(\boldsymbol{\mu}, \Sigma) \tilde{v}_{mj}, \quad m = 1, \dots, M, \\ \dot{\Sigma} &= A\Sigma + \Sigma A^\dagger, \end{aligned} \tag{3.32}$$

where the matrix A has components

$$A_{mn} = \sum_{j=1}^R b_{jm}(\boldsymbol{\mu}, \Sigma) (\tilde{s}_{nj} - \delta_{nm}) \tilde{v}_{mj}, \quad m, n = 1, \dots, M.$$

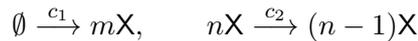
Here

$$b_{jm}(\boldsymbol{\mu}, \Sigma) = \exp \left\{ (\tilde{\mathbf{s}}_j - \tilde{\mathbf{e}}_m)^\dagger \boldsymbol{\mu} + \frac{1}{2} (\tilde{\mathbf{s}}_j - \tilde{\mathbf{e}}_m)^\dagger \Sigma (\tilde{\mathbf{s}}_j - \tilde{\mathbf{e}}_m) \right\},$$

$\tilde{\mathbf{e}}_m$ denotes the m -th standard basis vector in \mathbb{R}^M and δ_{nm} is the Kronecker delta.

3.5.2 Analytical results

We can investigate the performance of the approximation analytically on a simple class of one-dimensional systems, given by



for $m, n > 0$, with reaction rate equation

$$\dot{x} = mc_1 - c_2 x^n.$$

The solution at stationarity is $x_\infty = (mc_1/c_2)^{1/n}$. Assigning independent log-normal distributions $c_1 \sim \ln \mathcal{N}(\mu_1, \sigma_1^2)$ and $c_2 \sim \ln \mathcal{N}(\mu_2, \sigma_2^2)$, the distribution of the solution at stationarity is

$$x_\infty \sim \ln \mathcal{N}((\ln m + \mu_1 - \mu_2)/n, (\sigma_1^2 + \sigma_2^2)/n^2),$$

as follows from standard properties of log-normal distributions.

We now show that log-normal entropic matching reproduces this result. The rate equation in log-transformed variables $\xi = \ln x$, $\gamma_1 = \ln c_1$, $\gamma_2 = \ln c_2$ is

$$\dot{\xi} = \exp\{\ln m + \gamma_1 - \xi\} - \exp\{\gamma_2 + (n-1)\xi\}.$$

We denote the mean of ξ at stationarity by μ_x , and the covariance matrix at stationarity by

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{x1} & \Sigma_{x2} \\ \Sigma_{x1} & \sigma_1^2 & 0 \\ \Sigma_{x2} & 0 & \sigma_2^2 \end{bmatrix}.$$

Here we have already made use of the fact that c_1 and c_2 are independent, and that the distribution of (c_1, c_2) remains constant over time. Defining

$$u := \langle \exp\{\ln m + \gamma_1 - \xi\} \rangle = \exp\{\ln m + \mu_1 - \mu_x + (\Sigma_{xx} + \sigma_1^2 - 2\Sigma_{x1})/2\},$$

$$v := \langle \exp\{\gamma_2 + (n-1)\xi\} \rangle = \exp\{\mu_2 + (n-1)\mu_x + ((n-1)^2\Sigma_{xx} + \sigma_2^2 + 2(n-1)\Sigma_{x2})/2\},$$

the solution of (3.32) at stationarity can be seen to satisfy

$$\begin{aligned} u - v &= 0, \\ -(u + (n-1)v)\Sigma_{xx} + u\Sigma_{x1} - v\Sigma_{x2} &= 0, \\ -(u + (n-1)v)\Sigma_{x1} + u\sigma_1^2 &= 0, \\ -(u + (n-1)v)\Sigma_{x2} - v\sigma_2^2 &= 0. \end{aligned}$$

From this we obtain

$$\Sigma_{xx} = \frac{\sigma_1^2 + \sigma_2^2}{n^2}$$

and

$$\mu = \frac{\ln m + \mu_1 - \mu_2}{n},$$

which is the exact solution.

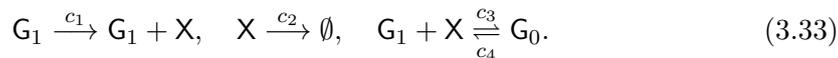
It is possible to extend this example to a certain special class of reaction systems that has a log-normal stationary distribution. This class is probably too restricted to be of any use, and it is clear that most reaction networks do not have a log-normal stationary distribution. Nevertheless, we demonstrate empirically in the following section that log-normal entropic matching can produce excellent approximations even for these systems.

3.5.3 Numerical results

We now investigate the performance of our approximation on two example systems: A gene expression model with negative feedback and a bistable system.

Gene expression with negative feedback

The first system we consider is a simple model of gene expression with negative feedback,



Here G_1 is the gene in the unbound state, G_0 the gene in the bound state, and X the protein produced. We consider both heterogeneous rate constants and random initial conditions and consider the quality of the approximation at various time points. The comparison between Monte Carlo simulations and log-normal entropic matching are displayed in Figure 3.5. We see very good agreement between exact (Monte Carlo) and approximate results. This holds for all three species and both during the transient regime and at stationarity.

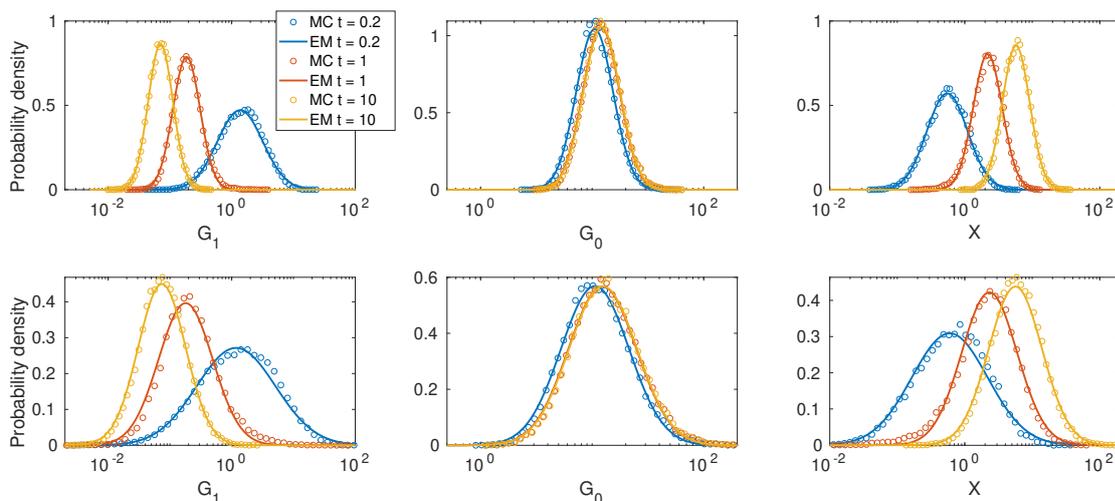
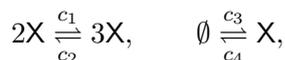


Figure 3.5: Comparison between log-normal entropic matching (EM, solid line) and Monte Carlo simulations (MC, circles, from 20,000 simulations) for gene expression with negative feedback (3.33). Parameter values and initial conditions were independent log-normally distributed. For the rate constants, $c_i \sim \ln \mathcal{N}(\mu_i, \sigma^2)$ with $\mu_1 = \ln 10$, $\mu_2 = \ln 0.1$, $\mu_3 = \ln 30$, $\mu_4 = \ln 1$. For the initial conditions at time 0, $G_1(0) \sim \ln \mathcal{N}(\ln 10, \sigma^2)$, $G_0(0) \sim \ln \mathcal{N}(\ln 2, \sigma^2)$ and $X(0) \sim \ln \mathcal{N}(\ln 1, \sigma^2)$. Top row corresponds to $\sigma^2 = 0.2$ and bottom row to $\sigma^2 = 0.7$. Colors encode the time-points $t = 0.2$, $t = 1$ and $t = 10$.

A bistable system

Here we consider the Schloegl system [66]



which is known to be bistable for certain values of the rate constants. Clearly, the unimodal log-normal distribution cannot capture the bimodal distributions that will occur for the Schloegl system. However, we can investigate whether the resulting approximation correctly captures the bulk of the mass of the true distribution (and in particular

whether the approximation breaks down in some sense). Figure 3.6 illustrates the results, which are arguably the best one could hope for: Our approximation has its mode in between the two modes of the true distribution, and correctly captures the width of the exact result.

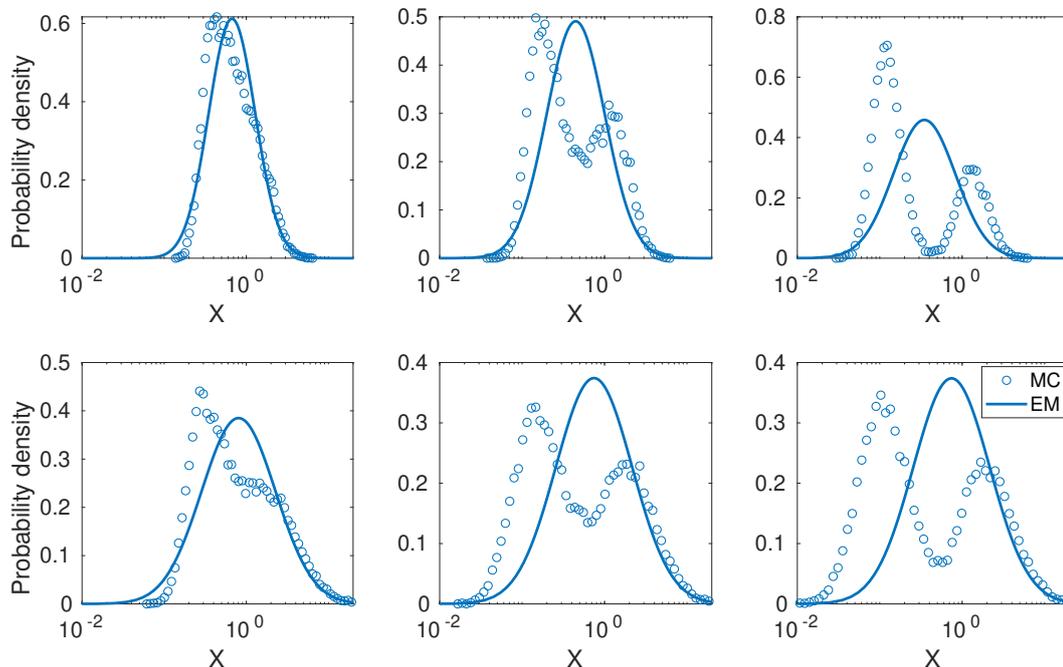


Figure 3.6: Comparison between log-normal entropic matching (EM, solid line) and Monte Carlo simulations (MC, circles, from 20,000 simulations) for the Schloegl system. Parameter values and initial conditions were independent log-normally distributed. For the rate constants, $c_i \sim \ln \mathcal{N}(\mu_i, \sigma^2)$ with $\mu_1 = \ln 2, \mu_2 = \ln 1.7, \mu_3 = \ln 0.1, \mu_4 = \ln 1$. For the initial conditions at time 0, $X(0) \sim \ln \mathcal{N}(\ln 2.5, \sigma^2)$. Top row corresponds to $\sigma^2 = 0.1$ and bottom row to $\sigma^2 = 0.4$. Columns correspond to time $t = 3$ (left column), $t = 6$ (middle column) and $t = 100$ (right column).

3.6 A diagrammatic technique for cumulant equations

The aim of this section is the derivation of the general formula for cumulant equations (and zero-cumulant closure) in the form of a diagrammatic technique. While software packages for the automatic generation of cumulant equations have been published [67], the general form of the equations appears to not have been investigated. The diagrammatic approach has several advantages. First of all, it provides a convenient mnemonic device when deriving the equations by hand for simple reaction networks. More importantly, it allows one to readily “see” the structure of the equations. This facilitates a convenient comparison with other types of approximations, especially if those other types can also be represented graphically. For example, here we will use our method to derive the very transparent relation between cumulant equations for the CME and for

the Kramers-Moyal expansion. Also, the investigation of the properties of the cumulant equations (and of zero-cumulant closure) can be simplified by our approach. We first describe our method in Section 3.6.1 and then provide the derivation of it in Section 3.6.2.

3.6.1 The diagrammatic rules

Consider the reaction network (2.3) with polynomial reaction kinetics, e.g. mass-action. The functions that specify the moment equations for such a system are polynomials, hence our task is to find a simple description of these polynomials. Instead of finding evolution equations for (centered or non-centered) moments, we can equally derive equations for cumulants. Moments can then be computed from the solution of the cumulant equations, if necessary. One advantage of considering cumulants is that they are more useful for understanding the underlying distributions. Additionally, when considering cumulant equations, the use of zero-cumulant closure becomes very simple. While zero-cumulant closure has been shown in the previous sections to not be very useful as a principled closure method, it nevertheless remains a fact that zero-cumulant closure is very popular, so that it is interesting to understand its properties.

The diagrammatic technique that we present will associate certain diagrams with each monomial of the cumulant equations. The full equations then emerge as a sum of diagrams. Because the rules that govern these diagrams are quite simple, we obtain a compact description of the cumulant equations.

Since we will derive the diagrammatic rules for both the CME and the Kramers-Moyal expansion (and in particular, for the CFPE), we will in this section use the notation $p(t, \mathbf{x})$ for the solution of the CME, and $\pi(t, \mathbf{z})$ for the solution of the Kramers-Moyal expansion, so that $\mathbf{x} \in \mathbb{N}_0^N$ and $\mathbf{z} \in \mathbb{R}_{\geq 0}^N$.

Moments and cumulants

We will require some basic facts about the relation of moments and cumulants. In this section, we will use multi-index notation, so that for $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{N}_0^N$ and $\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$, we write

$$\mathbf{x}^\alpha = \prod_{n=1}^N x_n^{\alpha_n}, \quad \alpha! = \prod_{n=1}^N \alpha_n!, \quad |\alpha| = \sum_{n=1}^N \alpha_n.$$

We define the moment of order $\alpha \in \mathbb{N}_0^N$ of $p(t, \mathbf{x})$ and $\pi(t, \mathbf{z})$ as

$$\sum_{\mathbf{x}} p(t, \mathbf{x}) \mathbf{x}^\alpha \quad \text{and} \quad \int d\mathbf{z} \pi(t, \mathbf{z}) \mathbf{z}^\alpha,$$

respectively, and the moment generating function as

$$G(t, \boldsymbol{\xi}) = \sum_{\mathbf{x}} p(t, \mathbf{x}) e^{\boldsymbol{\xi} \cdot \mathbf{x}} \quad \text{or} \quad G(t, \boldsymbol{\xi}) = \int d\mathbf{z} \pi(t, \mathbf{z}) e^{\boldsymbol{\xi} \cdot \mathbf{z}}.$$

The cumulant generating function F is defined by $F(t, \boldsymbol{\xi}) = \ln G(t, \boldsymbol{\xi})$, and the cumulant of order α as the coefficient of $\boldsymbol{\xi}^\alpha / \alpha!$ in the power series expansion of F around $\boldsymbol{\xi} = 0$.

It is convenient for our purposes to write moments and cumulants in a slightly different form. We will denote them, respectively, by $\langle i_1, \dots, i_M \rangle$ and $\langle\langle i_1, \dots, i_M \rangle\rangle$, where $i_1, \dots, i_M \in \{1, \dots, N\}$ are not necessarily distinct. This corresponds to a moment or cumulant of order $\alpha = \sum_{m=1}^M e_{i_m}$, where e_j is the j -th standard basis vector in \mathbb{R}^N . Note that the order in which these indices are written does not matter. For example, in terms of the power series expansion of the cumulant generating function F , the term $\langle\langle 2, 3, 3, 5, 5, 5 \rangle\rangle$ corresponds to the coefficient of $\frac{1}{2!3!} \xi_2 \xi_3^2 \xi_5^3$. The fact that moments and cumulants, and in particular their relation to each other, are often more transparent using this notation is well known [68].

The cumulant equations for the system are given by the sum of equations for each reaction separately. It is therefore sufficient to treat the case of a single reaction. Similarly, for a polynomial hazard function $h(\mathbf{x}) = \sum_{\alpha} b_{\alpha} \mathbf{x}^{\alpha}$, the cumulant equations are given by the sum of the equations for each monomial $b_{\alpha} \mathbf{x}^{\alpha}$. Therefore, we can consider the case of a monomial hazard function $\lambda(\mathbf{x}) = \mathbf{x}^{\alpha}$. We will write these monomials in the form $x_{j_1} \cdots x_{j_L}$, where $j_1, \dots, j_L \in \{1, \dots, N\}$ are not necessarily distinct. This is similar to our notation for cumulants and will make the diagrammatic technique especially simple. Note also that among “chemically meaningful” (i.e. at most bimolecular) reactions [37] for mass-action kinetics, all cases except reactions of the form $2X \rightarrow [\dots]$ actually have monomial reaction hazards with pairwise distinct factors.

Partitions and diagrams

Our method is based on partitions of sets. A partition is a decomposition of a set into disjoint, nonempty subsets (called blocks) such that the union of all blocks is equal to the set we started with. In our case, the elements of the sets for which we consider partitions will be called *points*. We have to distinguish two types of points, one corresponding to the factors in the hazard function (called by us *crosses*), the other corresponding to the indices of the cumulant for which we want to obtain the evolution equation (called by us *dots*).

For example, assume we want to obtain the evolution equation for $\langle\langle i_1, \dots, i_5 \rangle\rangle$ and a reaction hazard of the form $\lambda(\mathbf{x}) = x_j x_k x_l$. Then we will consider partitions of the point set shown in Figure 3.7a. Here the five dots correspond to the five indices i_1, \dots, i_5 , and the three crosses correspond to the three factors x_j, x_k and x_l . The relative positions in which the points are drawn are not important. Some example partitions are shown in Figure 3.7b – 3.7d. We represent blocks by placing the points belonging to them inside an oval, and use the convention that any point not in any oval belongs to a partition block consisting of only one element. Note that not all of the partitions shown in Figure 3.7b – 3.7d are allowed, as will be explained in Section 3.6.1. In the following, since we represent partitions by diagrams, we will use these two terms interchangeably.

The terms of the evolution equation for a cumulant are obtained by considering all possible partitions (subject to rules that forbid certain partitions), and then translating each partition into a term by certain translation rules. We now explain which diagrams are allowed, and subsequently explain how to translate diagrams into equation terms.

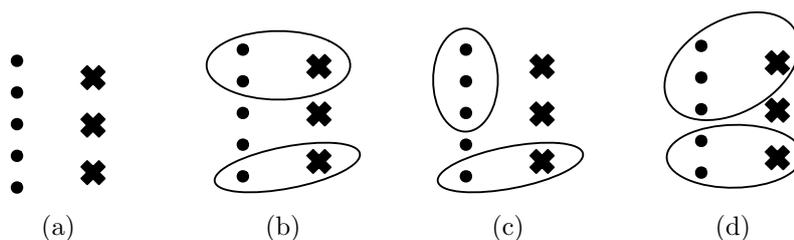


Figure 3.7: Example point set and partitions corresponding to a third-order reaction hazard and a fifth-order cumulant. Note that points that are not placed inside an oval are considered to be in a one-element block by themselves. For simplicity, we do not draw these blocks explicitly. Partitions (a) and (b) are allowed. Partition (c) is disallowed because of the cross rule, while partition (d) is disallowed because of the dot rule.

Allowed diagrams

As mentioned above, not all partitions are allowed. We now describe the rules which govern the cumulant equations. A special role is played by blocks with only one element, called singletons. The rules are:

Cross rule: Every non-singleton block has to contain at least one cross.

Dot rule: There has to be at least one singleton block that consists of a dot.

Examples of disallowed partitions are shown in Figure 3.7c and 3.7d. Apart from the partitions excluded by these rules, all partitions are allowed and contribute a term to the final evolution equation for the cumulant. The popular zero cumulant closure is then simply implemented using the following rule:

Zero-cumulant closure rule: For zero-cumulant closure of order C , a partition can have at most C points in each of its blocks.

A note regarding zero-order reactions: If $\lambda(\mathbf{x}) = 1$, there are no crosses and thus no cross partitions. The interpretation of our method in this case is that the only allowed partition blocks are the singleton dot blocks. Therefore, there exists only one diagram for this reaction for each cumulant.

Translation rules

We now explain the rules by which diagrams are translated into terms of the cumulant evolution equation. The term corresponding to a diagram is constructed by taking the product of the terms corresponding to each block of the diagram. The terms for each block are in turn constructed using two rules, one for singleton dot blocks and one for all other blocks:

Cross block: A block that contains at least one cross, and which is made up of points with the indices m_1, \dots, m_K , corresponds to the cumulant $\langle\langle m_1, \dots, m_K \rangle\rangle$.

Dot block: A singleton block made up of one dot with index i corresponds to the factor ν_i , where $\boldsymbol{\nu}$ is the stoichiometric change vector of the reaction.¹

¹Note that we distinguish between the i -th component ν_i of a stoichiometric vector $\boldsymbol{\nu}$ and the stoichiometric vector $\boldsymbol{\nu}_j$ of the j -th reaction by using bold typeface for the latter.

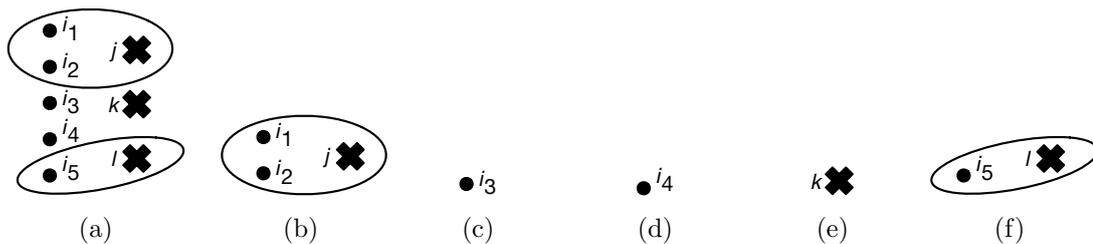
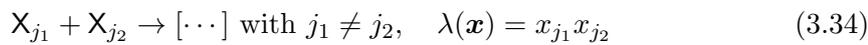


Figure 3.8: Example of translating a diagram into its corresponding term. The full diagram is shown in (a), and its blocks in (b) – (f). By the rule for cross blocks, the factor corresponding to (b) is $\langle\langle i_1, i_2, j \rangle\rangle$. For (c) and (d), we apply the rule for dot blocks, so the factors are ν_{i_3} and ν_{i_4} , respectively. For (e) and (f), we apply the rule for cross blocks, so the corresponding factors are $\langle\langle k \rangle\rangle$ and $\langle\langle i_5, l \rangle\rangle$. The full term corresponding to the diagram in (a) is thus the product $\nu_{i_3} \nu_{i_4} \langle\langle i_1, i_2, j \rangle\rangle \langle\langle k \rangle\rangle \langle\langle i_5, l \rangle\rangle$.

Finally, each diagram has to be multiplied by the constant factor of the monomial under consideration. An example of translating a diagram into its corresponding term is given in Figure 3.8. This completes the formulation of our diagrammatic technique.

Example: A bimolecular reaction

Let us write down the full set of diagrams for cumulants of first, second and third order for a bimolecular reaction with mass-action kinetics



and a stoichiometric change vector $\boldsymbol{\nu}$. The result is shown in Figure 3.9, where we also state the corresponding equations. For simplicity, we here wrote down not all diagrams, but rather diagram types, i.e. we omit the labels for the points. The actual set of diagrams can be obtained by appropriately labeling each diagram, as can be seen from the equations in Figure 3.9. If we were to employ third-order zero-cumulant moment closure, the first diagram in the third line would be omitted.

Kramers-Moyal expansion

A method to approximate the CME (which is defined on a discrete state space) by a partial differential equation defined on continuous state space $\mathbb{R}_{\geq 0}^N$ is the Kramers-Moyal expansion. This is simply the Taylor-series expansion in terms of the increment given by the stoichiometric change vectors $\boldsymbol{\nu}_j$ truncated to order Q and results in the equation

$$\partial_t \pi(t, \mathbf{z}) = \sum_{|\alpha|=1}^Q \sum_{j=1}^R \frac{(-1)^{|\alpha|} \boldsymbol{\nu}_j^\alpha}{\alpha!} \partial_{\mathbf{z}}^\alpha [h_j(\mathbf{z}) \pi(t, \mathbf{z})]. \quad (3.35)$$

The most important special case is the CFPE (2.7), which corresponds to $Q = 2$. As for the CME, solving the equation resulting from the Kramers-Moyal expansion is generally not possible, but just as for the CME, we can derive moment equations. These will not

$$\begin{aligned}
 \partial_t \langle\langle i_1 \rangle\rangle &= \bullet \begin{array}{c} \circlearrowleft \\ \times \\ \times \end{array} + \bullet \begin{array}{c} \times \\ \times \end{array} \\
 &= \nu_{i_1} \langle\langle j_1, j_2 \rangle\rangle + \nu_{i_1} \langle\langle j_1 \rangle\rangle \langle\langle j_2 \rangle\rangle, \\
 \partial_t \langle\langle i_1, i_2 \rangle\rangle &= \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \end{array} + \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \end{array} + \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \end{array} + \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \end{array} \\
 &= \nu_{i_2} \langle\langle i_1, j_1, j_2 \rangle\rangle + \nu_{i_1} \langle\langle i_2, j_1, j_2 \rangle\rangle \\
 &\quad + \nu_{i_2} \langle\langle i_1, j_1 \rangle\rangle \langle\langle j_2 \rangle\rangle + \nu_{i_2} \langle\langle i_1, j_2 \rangle\rangle \langle\langle j_1 \rangle\rangle \\
 &\quad + \nu_{i_1} \langle\langle i_2, j_1 \rangle\rangle \langle\langle j_2 \rangle\rangle + \nu_{i_1} \langle\langle i_2, j_2 \rangle\rangle \langle\langle j_1 \rangle\rangle \\
 &\quad + \nu_{i_1} \nu_{i_2} \langle\langle j_1, j_2 \rangle\rangle + \nu_{i_1} \nu_{i_2} \langle\langle j_1 \rangle\rangle \langle\langle j_2 \rangle\rangle, \\
 \partial_t \langle\langle i_1, i_2, i_3 \rangle\rangle &= \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \\ \bullet \end{array} + \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \end{array} + \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \\ \bullet \end{array} + \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \\ \bullet \end{array} \\
 &\quad + \begin{array}{c} \bullet \quad \times \\ \bullet \end{array} + \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \end{array} + \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \end{array} + \begin{array}{c} \bullet \quad \times \\ \bullet \quad \times \end{array}
 \end{aligned}$$

Figure 3.9: Cumulant diagram types and corresponding equation terms for a bimolecular reaction for cumulants of orders 1 to 3. The indices for the crosses are j_1 and j_2 . The indices for the dots are i_1 for the first-order cumulant and i_1 and i_2 for the second-order cumulant. Note that for the second-order cumulant, the first diagram type corresponds to 2 terms in the equation, while the second diagram type corresponds to 4 terms. The diagrams for the third order cumulant have not been translated because they lead to somewhat lengthy expressions.

be closed, and various moment-closure techniques can again be applied. For $Q > 2$, a theorem [69] states that the solution of the Kramers-Moyal expansion will not be positive everywhere and thus is not a valid probability distribution. This, however, does not mean that the Kramers-Moyal expansion is not useful as an approximation, in particular if one is interested in approximating moments or cumulants [70].

Assume now that instead of the CME, we consider the Kramers-Moyal expansion of order Q . It turns out that all rules formulated so far remain true, except that the set of allowed diagrams is further constrained by the following:

Kramers-Moyal rule: When using a Kramers-Moyal expansion of order Q , a diagram may contain at most Q dot blocks.

Returning to the example of a bimolecular reaction (3.34), when considering the Fokker-Planck equation (i.e. the case $Q = 2$), we see that the equations for $\partial_t \langle\langle i_1 \rangle\rangle$ and $\partial_t \langle\langle i_1, i_2 \rangle\rangle$

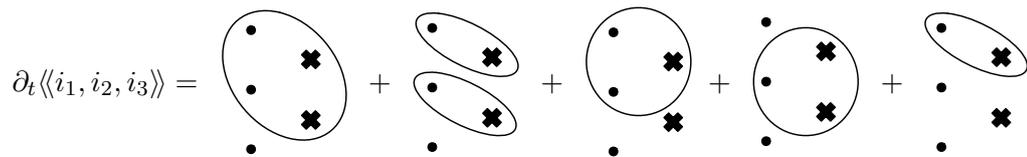


Figure 3.10: Diagram types for a third order cumulant and a second-order monomial hazard when using the Kramers-Moyal expansion of order $Q = 2$.

$$\partial_t \langle\langle 1 \rangle\rangle = \bullet \quad , \quad \partial_t \langle\langle 1, 1 \rangle\rangle = \begin{array}{c} \bullet \\ \bullet \end{array} \quad , \quad \partial_t \langle\langle 1, 1, 1 \rangle\rangle = \begin{array}{c} \bullet \\ \bullet \\ \bullet \end{array}$$

Figure 3.11: Diagrams for cumulants of order one to three for the Poisson process. These (and all higher order diagrams) contain only partition blocks of size 1, so they are closed. When using the Kramers-Moyal expansion of order Q , we are not allowed to have more than Q single-element dot blocks in a diagram, so the first Q cumulant equations will remain unchanged, whereas all higher cumulant equations will have no diagrams.

remain the same as given in Figure 3.9, while for $\partial_t \langle\langle i_1, i_2, i_3 \rangle\rangle$ we instead obtain the diagrams shown in Figure 3.10. We thus have a very transparent relation between the cumulant equations for the CME and for the Kramers-Moyal expansion. As an example, we can apply our technique to the simple Poisson process, for which the Kramers-Moyal expansion was investigated in [70]. There it was shown by calculation that cumulants up to the expansion order Q match the exact ones, and that higher cumulants vanish (if they vanish at $t = t_0$). From our diagrammatic technique, this is immediately obvious as illustrated in Figure 3.11.

Moment equations

For comparison, we briefly write down the rules for the usual moment equations. We obtain one additional rule:

Moment equation rule: When deriving moment equations instead of cumulant equations, all crosses have to be in the same partition block.

Translating diagrams into equations is done as before, except that partition blocks are now translated into moments instead of cumulants. Note that the rule for zero-cumulant closure is then no longer applicable. Computing moment equations for the Kramers-Moyal expansion instead of the CME again only requires the addition of the Kramers-Moyal rule. It is also possible to obtain the rules for centered moment equations together with centered-moment-neglect closure [40, 44], but we do not consider them here. From the above rule for moment equations, we see that considering moment equations instead of cumulant equations reduces the number of diagrams, and thus the number of terms, in the corresponding evolution equation.

Some practical aspects

When applying our technique to small systems by hand, many entries of stoichiometric vectors will be zero. This greatly reduces the number of diagrams that have to be summed. If, for some index i , one has $\nu_i = 0$, all diagrams in which i is in a dot block can be omitted. Additionally, some diagrams will occur several times for one and the same combination of cumulant and reaction. This can happen, for example, if we consider a higher-order cumulant with repeated entries (such as $\langle\langle 1, 1 \rangle\rangle$) or a reaction with repeated factors in the propensity (such as $\lambda(\mathbf{x}) = x_1^2$). In concrete cases, it is straightforward to figure out the necessary multiplicities. For theoretical purposes, on the other hand, the simple description given by us might be the more convenient one.

3.6.2 Derivation

In order to derive our diagrammatic technique, we will formulate a partial differential equation (PDE) for the cumulant generating function $F(t, \boldsymbol{\xi})$ of the marginal probability distributions $p(t, \mathbf{x})$ (for the solution of the CME) or $\pi(t, \mathbf{z})$ (for the solution of the Kramers-Moyal expansion, even though this might not be a probability density). Deriving moment equations by using the moment generating function is a standard approach, shown e.g. in [71], and deriving cumulant equations from the cumulant generating function is quite similar and has been described e.g. in [67], where a software package for generating the equations was presented. However, the systematic form that these equations take was not investigated.

Derivation for the Chemical master equation

To derive a partial differential equation for G (and subsequently for $F = \ln G$), we multiply the CME (2.5) by $e^{\boldsymbol{\xi} \cdot \mathbf{x}}$ and sum over \mathbf{x} . The result can be conveniently expressed as [72]

$$\partial_t G(t, \boldsymbol{\xi}) = \sum_{j=1}^R (e^{\boldsymbol{\xi} \cdot \boldsymbol{\nu}_j} - 1) h_j(\partial_{\boldsymbol{\xi}}) G(t, \boldsymbol{\xi})$$

where $h_j(\partial_{\boldsymbol{\xi}}) = h_j(\partial_{\xi_1}, \dots, \partial_{\xi_D})$ is defined by a formal substitution of the operators ∂_{ξ_d} into the polynomial h_j . From this, we obtain

$$\partial_t F(t, \boldsymbol{\xi}) = \sum_{j=1}^R (e^{\boldsymbol{\xi} \cdot \boldsymbol{\nu}_j} - 1) e^{-F(t, \boldsymbol{\xi})} h_j(\partial_{\boldsymbol{\xi}}) e^{F(t, \boldsymbol{\xi})}.$$

As noted above, it is sufficient to consider the case of a monomial $\lambda(\mathbf{x}) = x_{j_1} \cdots x_{j_L}$ with corresponding stoichiometric vector $\boldsymbol{\nu}$. Denote by $\{F\}_L$ the collection of F together with its various partial derivatives of order $\leq L$. Evaluating $e^{-F(t, \boldsymbol{\xi})} \lambda(\partial_{\boldsymbol{\xi}}) e^{F(t, \boldsymbol{\xi})}$ will result in an expression $\Lambda(\{F\}_L)$, where Λ is a polynomial the variables of which are the various partial derivatives of F . Determining this polynomial amounts to an application of Faà di Bruno's formula, but because we are dealing with the special case of e^F with a cumulant generating function F , we use the well known fact that Λ will be the polynomial that expresses the moment $\langle \lambda(\mathbf{x}) \rangle$ in terms of the cumulants. This in turn can be described

as a sum over all partitions of L points labeled by j_1, \dots, j_L , where each summand is the product over blocks of the partition:

$$\Lambda(\{F\}_L) = \sum_{\pi \in \Omega} \prod_{\sigma \in \pi} F_\sigma.$$

Here Ω is the set of all partitions of L points corresponding to the indices j_1, \dots, j_L . The product runs over all blocks $\sigma \in \pi$ of the partition. For such a partition block σ with indices j_{k_1}, \dots, j_{k_n} , we denote by F_σ the corresponding partial derivative $\partial_{j_{k_1}} \cdots \partial_{j_{k_n}} F$. See also Figure 3.12. Here and in the following, we use the notation $\partial_j = \partial_{\xi_j}$.

Briefly returning to the case of several reactions, the full PDE thus takes the form

$$\partial_t F(t, \boldsymbol{\xi}) = \sum_{j=1}^R (e^{\boldsymbol{\xi} \cdot \boldsymbol{\nu}_j} - 1) \Lambda_j(\{F\}_L), \quad (3.36)$$

where Λ_j is the sum of the polynomials Λ corresponding to each monomial λ of the hazard h_j . However, we continue to consider the case of a single monomial hazard function. Then the right-hand side of the PDE reads

$$\sum_{\pi \in \Omega} (e^{\boldsymbol{\xi} \cdot \boldsymbol{\nu}} - 1) \prod_{\sigma \in \pi} F_\sigma.$$

In order to derive the evolution equation for $\langle\langle i_1, \dots, i_M \rangle\rangle$, we merely have to apply $\partial_{i_1} \cdots \partial_{i_M}$ to this term and evaluate the resulting expressions at $\boldsymbol{\xi} = 0$, at which the derivatives of F will equal their corresponding cumulants. We now first consider the translation rules for diagrams, and subsequently check which diagrams are allowed.

Translation rules: Performing the derivatives for each partition π over which the sum ranges, we have to assign the indices i_1, \dots, i_M to any of the terms $(F_\sigma)_{\sigma \in \pi}$ and $(e^{\boldsymbol{\xi} \cdot \boldsymbol{\nu}} - 1)$ in all possible ways, and sum over the various possibilities. We draw M dots and L crosses and first partition the crosses as defined by π . We then indicate the assignment of a partial derivative to one of the terms F_σ by placing the corresponding dot into the partition block σ . This will produce partitions on M dots and L crosses. Refer to Figure 3.12 for an example of the distribution of derivatives and the corresponding diagrams. All the indices that have not been assigned to any of the terms F_σ for some $\sigma \in \pi$ necessarily have to be assigned to $(e^{\boldsymbol{\xi} \cdot \boldsymbol{\nu}} - 1)$. In our diagrams, we leave these indices in singleton blocks. We see that on evaluating at $\boldsymbol{\xi} = 0$, these latter terms will produce the factors described in the rule for dot blocks. On the other hand, indices i_{l_1}, \dots, i_{l_m} assigned to one of the factors F_σ for a block σ with indices j_{k_1}, \dots, j_{k_n} will, on evaluating at $\boldsymbol{\xi} = 0$, give as a factor the corresponding cumulant $\langle\langle i_{l_1}, \dots, i_{l_m}, j_{k_1}, \dots, j_{k_n} \rangle\rangle$. This is the rule for cross blocks.

Allowed diagrams: We are summing over all possible partitions π of crosses. For every partition, the number and indices of dots assigned to partition blocks is arbitrary, and the dots can be distributed over cross partitions in an arbitrary way. However, dots have to be assigned either to some partition block of crosses, or to the term $(e^{\boldsymbol{\xi} \cdot \boldsymbol{\nu}} - 1)$. This produces the cross rule, i.e. it is not possible to have more than one dot in a block without any crosses. Additionally, assigning all indices to the factors $(F_\sigma)_{\sigma \in \pi}$ would leave the term $(e^{\boldsymbol{\xi} \cdot \boldsymbol{\nu}} - 1)$, which on evaluating at $\boldsymbol{\xi} = 0$ would vanish. Thus at least one

index has to be assigned to $(e^{\xi \cdot \nu} - 1)$, which we called the dot rule. This completes the derivation of the diagram rules when neglecting closure.

Closure: The zero-cumulant closure rule is not difficult to see. Cumulants of order larger than C correspond to partition blocks with more than C points. In order to perform zero-cumulant closure of order C , we merely have to remove diagrams for which any block has more than C points.

$$\begin{aligned}
 [\partial_{i_2}(e^{\xi \cdot \nu} - 1)][\partial_{i_1} F_{\{j_1, j_2\}}] &= \begin{array}{c} \bullet i_1 \quad j_1 \times \\ \bullet i_2 \quad j_2 \times \end{array} \\
 [\partial_{i_1} \partial_{i_2}(e^{\xi \cdot \nu} - 1)][F_{\{j_1, j_2\}}] &= \begin{array}{c} \bullet i_1 \quad j_1 \times \\ \bullet i_2 \quad j_2 \times \end{array} \\
 [\partial_{i_2}(e^{\xi \cdot \nu} - 1)][\partial_{i_1} F_{\{j_1\}}][F_{\{j_2\}}] &= \begin{array}{c} \bullet i_1 \quad j_1 \times \\ \bullet i_2 \quad j_2 \times \end{array} \\
 [\partial_{i_2}(e^{\xi \cdot \nu} - 1)][F_{\{j_1\}}][\partial_{i_1} F_{\{j_2\}}] &= \begin{array}{c} \bullet i_1 \quad j_1 \times \\ \bullet i_2 \quad j_2 \times \end{array} \\
 [\partial_{i_1} \partial_{i_2}(e^{\xi \cdot \nu} - 1)][F_{\{j_1\}}][F_{\{j_2\}}] &= \begin{array}{c} \bullet i_1 \quad j_1 \times \\ \bullet i_2 \quad j_2 \times \end{array}
 \end{aligned}$$

Figure 3.12: Example of distributions of derivatives over partition blocks and the term $(e^{\xi \cdot \nu} - 1)$, together with the corresponding diagrams. Here we consider a quadratic reaction hazard $\lambda(\mathbf{x}) = x_{j_1} x_{j_2}$ and diagrams for second-order cumulants. In this case, there are only two possible partitions of crosses, $\{j_1, j_2\}$ and $\{j_1\}, \{j_2\}$. These correspond to the formula $\langle j_1, j_2 \rangle = \langle\langle j_1, j_2 \rangle\rangle + \langle\langle j_1 \rangle\rangle \langle\langle j_2 \rangle\rangle$ expressing the moment in terms of cumulants. In the figure, the first two lines correspond to the partition $\{j_1, j_2\}$ while the last three lines correspond to the partition $\{j_1\}, \{j_2\}$. Note that not all possible combinations have been shown.

Derivation for the Kramers-Moyal expansion

We proceed analogously to the case of the CME. The moment generating function satisfies the PDE

$$\partial_t G(t, \boldsymbol{\xi}) = \sum_{j=1}^R \left(\sum_{|\alpha|=1}^Q \frac{\boldsymbol{\xi}^\alpha \nu_j^\alpha}{\alpha!} \right) h_j(\partial_{\boldsymbol{\xi}}) G(t, \boldsymbol{\xi}).$$

This can be seen by multiplying the Kramers-Moyal expansion (3.35) by $e^{\boldsymbol{\xi} \cdot \mathbf{z}}$, integrating and applying the usual rules for the Laplace transform of derivatives and multiplication by polynomials. Consequently the cumulant generating function satisfies the PDE

$$\begin{aligned} \partial_t F(t, \boldsymbol{\xi}) &= e^{-F(t, \boldsymbol{\xi})} \sum_{j=1}^R \left(\sum_{|\alpha|=1}^Q \frac{\boldsymbol{\xi}^\alpha \nu_j^\alpha}{\alpha!} \right) h_j(\partial_{\boldsymbol{\xi}}) e^{F(t, \boldsymbol{\xi})} \\ &= \sum_{j=1}^R \left(\sum_{|\alpha|=1}^Q \frac{\boldsymbol{\xi}^\alpha \nu_j^\alpha}{\alpha!} \right) \Lambda_j(\{F\}_N). \end{aligned}$$

We see that this equation differs from the corresponding equation (3.36) for the CME only in that the terms $(e^{\boldsymbol{\xi} \cdot \boldsymbol{\nu}_j} - 1)$ have been replaced by their Taylor expansions of order Q around $\boldsymbol{\xi} = 0$. Again, the evolution equation for the cumulant $\langle\langle i_1, \dots, i_M \rangle\rangle$ is obtained by taking derivatives and setting $\boldsymbol{\xi} = 0$. We can now repeat the argument for the CME, with one difference: When distributing some number m of the derivatives $\partial_{i_{i_1}} \cdots \partial_{i_{i_m}}$ to the term $\sum_{|\alpha|=1}^Q \frac{\boldsymbol{\xi}^\alpha \nu_j^\alpha}{\alpha!}$, we will obtain the term $\nu_{i_{i_1}} \cdots \nu_{i_{i_m}}$ (in analogy to the rule for the CME) only if $m \leq Q$. Otherwise, the corresponding term will be zero. Thus, we should not include diagrams with more than Q singleton dot blocks, which is what we called the Kramers-Moyal rule.

Chapter 4

The projection operator formalism for marginalization of heterogeneous reaction kinetics

A method for marginalization with a long history is the projection operator formalism of Nakajima [73], Zwanzig [74] and Mori [75]. In this chapter, a number of applications of this formalism to biomolecular reaction networks are discussed. We will consider a reaction network the species of which have been partitioned into two groups, the subnet $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N)$ and the environment $\hat{\mathbf{X}} = (\hat{\mathbf{X}}_1, \dots, \hat{\mathbf{X}}_{\hat{N}})$, as explained in Section 2.2:

$$\sum_{n=1}^N s_{nj} \mathbf{X}_n + \sum_{n=1}^{\hat{N}} \hat{s}_{nj} \hat{\mathbf{X}}_n \longrightarrow \sum_{n=1}^N r_{nj} \mathbf{X}_n + \sum_{n=1}^{\hat{N}} \hat{r}_{nj} \hat{\mathbf{X}}_n, \quad j = 1, \dots, R. \quad (4.1)$$

Our goal is to obtain (generally, approximate) descriptions of the marginal dynamics of the subnet species $\mathbf{X}_1, \dots, \mathbf{X}_N$. Throughout this chapter, a description in terms of forward equations is used, i.e., we are interested in (approximate) descriptions of the marginal distributions $p(t, \mathbf{x})$ of the subnet species. This is in contrast to the results in Chapter 5, where marginal equations for the *process* $\mathbf{X}(t)$ describing the subsystem dynamics will be obtained. This chapter includes material from [26].

4.1 The projection operator formalism across the model hierarchy

In this section, we briefly review the projection operator formalism in the context of reaction networks and show how applications of the formalism on different levels of the model hierarchy (Figure 2.1) are related to each other. We will introduce the idea behind the projection operator formalism on the special case of RREs with linear environment-environment and subnet-environment interactions. Thus, consider the RREs for the system (4.1), where $\boldsymbol{\mu}$ and $\hat{\boldsymbol{\mu}}$ denote, respectively, the concentrations of subnet and environment species:

$$\begin{aligned} \frac{d}{dt} \boldsymbol{\mu} &= \mathbf{f}(\boldsymbol{\mu}) + A^{\text{SE}} \hat{\boldsymbol{\mu}}, \\ \frac{d}{dt} \hat{\boldsymbol{\mu}} &= A^{\text{EE}} \hat{\boldsymbol{\mu}} + A^{\text{ES}} \boldsymbol{\mu} + \hat{\mathbf{b}}. \end{aligned} \quad (4.2)$$

The system state for the subnet components $\boldsymbol{\mu}$ evolves according to arbitrary dynamics \mathbf{f} , while the environment state evolves according to linear dynamics, specified by the

matrix A^{EE} and the inhomogeneity $\hat{\mathbf{b}}$. The subnet-environment interactions are also assumed to be linear and are given by the matrices A^{SE} and A^{ES} . Obtaining a marginal equation for the subnet variables $\boldsymbol{\mu}$ is now very simple. First, we solve the (linear, inhomogeneous) equation for $\hat{\boldsymbol{\mu}}$ in terms of $\boldsymbol{\mu}$:

$$\hat{\boldsymbol{\mu}}(t) = \int_0^t dt' e^{(t-t')A^{\text{EE}}} (A^{\text{ES}} \boldsymbol{\mu}(t') + \hat{\mathbf{b}}) + e^{tA^{\text{EE}}} \hat{\boldsymbol{\mu}}(0).$$

Plugging this equation into the equation for $\boldsymbol{\mu}$ in (4.2), we obtain a closed integro-differential equation for $\boldsymbol{\mu}$:

$$\dot{\boldsymbol{\mu}}(t) = \mathbf{f}(\boldsymbol{\mu}(t)) + \int_0^t dt' A^{\text{SE}} e^{(t-t')A^{\text{EE}}} (A^{\text{ES}} \boldsymbol{\mu}(t') + \hat{\mathbf{b}}) + A^{\text{SE}} e^{tA^{\text{EE}}} \hat{\boldsymbol{\mu}}(0). \quad (4.3)$$

The above derivation clearly depends on the linearity of environment-environment and subnet-environment interactions. A more general result is obtained by noting that forward equations, such as CME or CFPE, are always linear. This is the basis for the Nakajima-Zwanzig-Mori projection operator formalism, which we now review.

Consider an evolution equation

$$\frac{d}{dt} p_t = \mathcal{L} p_t \quad (4.4)$$

with a linear operator \mathcal{L} . This might be the CME, CFPE or the Liouville equation corresponding to the RRE. We are interested in finding a reduced equation for some lower-dimensional variable. This variable is defined by applying a projection operator \mathcal{P} to state p . For example, in the context of the CME where only a subset of the species is of interest, the projection operator might be defined by first applying a marginalization operator

$$[\mathcal{M}p](\mathbf{x}) = \sum_{\hat{\mathbf{x}}} p(\mathbf{x}, \hat{\mathbf{x}}). \quad (4.5)$$

The result is then lifted back to the original space by multiplying with a fixed probability distribution $q(\hat{\mathbf{x}})$ over the environment states [19, 76],

$$[\mathcal{U}p](\mathbf{x}, \hat{\mathbf{x}}) = q(\hat{\mathbf{x}}) p(\mathbf{x}).$$

The full projection operator is then given by $\mathcal{P} := \mathcal{U}\mathcal{M}$. Generally, the projection operator should satisfy $\mathcal{P}^2 = \mathcal{P}$, as it does here. Defining the ‘‘orthogonal’’ projection $\mathcal{Q} := 1 - \mathcal{P}$, we also have

$$\mathcal{Q}^2 = \mathcal{Q}, \quad \mathcal{P}\mathcal{Q} = \mathcal{Q}\mathcal{P} = 0.$$

Applying \mathcal{M} (respectively, \mathcal{Q}) to (4.4) and using $\mathcal{P} + \mathcal{Q} = 1$ and $\mathcal{P} = \mathcal{U}\mathcal{M}$, we obtain the two equations

$$\frac{d}{dt} \mathcal{M}p_t = \mathcal{M}\mathcal{L}\mathcal{U}\mathcal{M}p_t + \mathcal{M}\mathcal{L}\mathcal{Q}p_t, \quad (4.6)$$

$$\frac{d}{dt} \mathcal{Q}p_t = \mathcal{Q}\mathcal{L}\mathcal{Q}p_t + \mathcal{Q}\mathcal{L}\mathcal{U}\mathcal{M}p_t. \quad (4.7)$$

Formally solving (4.7) results in

$$\mathcal{Q}p_t = e^{t\mathcal{Q}\mathcal{L}} \mathcal{Q}p_0 + \int_0^t dt' e^{(t-t')\mathcal{Q}\mathcal{L}} \mathcal{Q}\mathcal{L}\mathcal{U}\mathcal{M}p_{t'},$$

and inserting the latter into (4.6) finally yields

$$\frac{d}{dt}\mathcal{M}p_t = \mathcal{M}\mathcal{L}\mathcal{U}\mathcal{M}p_t + \int_0^t dt' \mathcal{M}\mathcal{L}e^{(t-t')\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{U}\mathcal{M}p_{t'} + \mathcal{M}\mathcal{L}e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}p_0. \quad (4.8)$$

This is a closed equation for the marginalized distribution $\mathcal{M}p_t$, or in the general case for the lower-dimensional variable. The first term on the right-hand side here can be seen as a ‘‘Markovian’’ term, and the second, convolutional term as a ‘‘memory’’. The third term is usually referred to as a ‘‘noise’’ term, by virtue of the fact that it depends on the environment part $\mathcal{Q}p_0$ of the initial state p_0 . Equation (4.8) is the result of formal rewriting of the original evolution equation (4.4) and does not in itself result in any simplification. It can, however, be used as a basis for approximations, as will be done in Section 4.2.

Before considering approximations, it is worthwhile to understand some of the features of (4.8), and in particular to relate (4.8) for the CME to the corresponding result (4.3) for the RRE (4.2).

4.1.1 Mean equation for the projected CME

In this section, we demonstrate how to extract marginal mean equations from the marginal evolution equation (4.8) in the case when the environment-environment and subnet-environment interactions are linear. This is the situation for which the direct marginalization (4.3) of the RRE (4.2) was possible. We will relate the marginal mean equations derived in the following to the marginal RRE derived above. The derivation is presented for the case where the reaction network is modeled via the CME. A completely analogous derivation would be possible, for instance, for the CFPE. Our derivation will demonstrate how to obtain an exact marginal equation for the subnet species means when subnet-subnet interactions are non-linear.

Starting from (4.8), the expectation $\langle\phi\rangle_t$ of a function $\phi(\mathbf{x})$ with respect to the marginalized distribution $\mathcal{M}p_t$ evolves according to

$$\frac{d}{dt}\langle\phi\rangle_t = \langle\mathcal{U}^\dagger\mathcal{L}^\dagger\mathcal{M}^\dagger\phi\rangle_t + \int_0^t dt' \langle\mathcal{U}^\dagger\mathcal{L}^\dagger\mathcal{Q}^\dagger e^{(t-t')\mathcal{L}^\dagger\mathcal{Q}^\dagger}\mathcal{L}^\dagger\mathcal{M}^\dagger\phi\rangle_{t'} + \langle\mathcal{Q}^\dagger e^{t\mathcal{L}^\dagger\mathcal{Q}^\dagger}\mathcal{L}^\dagger\mathcal{M}^\dagger\phi\rangle_0. \quad (4.9)$$

Thus, we need expressions for the adjoints of the operators involved. Let $\psi(\mathbf{x}, \hat{\mathbf{x}})$ and $\phi(\mathbf{x})$ be two arbitrary functions. A brief computation shows that we have

$$\begin{aligned} [\mathcal{M}^\dagger\phi](\mathbf{x}, \hat{\mathbf{x}}) &= \phi(\mathbf{x}), \\ [\mathcal{U}^\dagger\psi](\mathbf{x}) &= \sum_{\hat{\mathbf{x}}} q(\hat{\mathbf{x}})\psi(\mathbf{x}, \hat{\mathbf{x}}) \end{aligned}$$

and thus

$$[\mathcal{P}^\dagger\psi](\mathbf{x}, \hat{\mathbf{x}}) = \sum_{\hat{\mathbf{x}}'} q(\hat{\mathbf{x}}')\psi(\mathbf{x}, \hat{\mathbf{x}}').$$

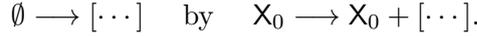
Choosing $\phi(\mathbf{x}) = \mathbf{x}$, we obtain from (4.9) equations for the mean abundances. To get an explicit expression for that equation, we first decompose the full time-evolution operator \mathcal{L} as $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1$, where \mathcal{L}_0 includes all linear reactions, and \mathcal{L}_1 contains the non-linear

subnet-subnet interactions. Strictly speaking, the operator \mathcal{L}_1 acts on distributions over the joint space of subnet and environment. However, in the following we will sometimes implicitly consider it to act on distributions over the subnet state only.

We first treat the case of a fully linear network, so that all reactions are of one of the two forms



where the first form can be seen as a special case of the second form by including an auxiliary species X_0 with abundance 1 and replacing the reaction



We will assume in the following that this rewriting has been performed.

Noting that ϕ is linear, we can obtain the mean equations in a more explicit form by first verifying that each operator involved in (4.9) maps the space of affine-linear functions to itself. Then (4.9), which is infinite-dimensional, reduces to a finite-dimensional equation for the coefficients of an affine-linear function. Thus, let

$$\psi(\mathbf{x}, \hat{\mathbf{x}}) = \mathbf{w}\mathbf{x} + \hat{\mathbf{w}}\hat{\mathbf{x}} + v, \quad \phi(\mathbf{w}) = \mathbf{w}\mathbf{x} + v$$

be general affine-linear functions, where \mathbf{w} and $\hat{\mathbf{w}}$ are row vectors and v is a scalar. As explained in Section 2.3.2, the first-order moment equations for a linear network are closed. The action of the operator \mathcal{L}_0^\dagger can then be specified via the matrix

$$A = \begin{bmatrix} A^{\text{SS}} & A^{\text{SE}} \\ A^{\text{ES}} & A^{\text{EE}} \end{bmatrix}$$

as

$$[\mathcal{L}_0^\dagger \psi](\mathbf{x}, \hat{\mathbf{x}}) = [\mathbf{w}, \hat{\mathbf{w}}] A \begin{bmatrix} \mathbf{x} \\ \hat{\mathbf{x}} \end{bmatrix},$$

where A^{SS} is the matrix specifying the linear part of the subnet-subnet interactions. Similarly,

$$\begin{aligned} [\mathcal{M}^\dagger \phi](\mathbf{x}, \hat{\mathbf{x}}) &= \mathbf{w}\mathbf{x} + v, \\ [\mathcal{U}^\dagger \psi](\mathbf{x}) &= \mathbf{w}\mathbf{x} + \hat{\mathbf{w}} \langle \hat{\mathbf{x}} \rangle_q + v, \\ [\mathcal{P}^\dagger \psi](\mathbf{x}, \hat{\mathbf{x}}) &= \mathbf{w}\mathbf{x} + \hat{\mathbf{w}} \langle \hat{\mathbf{x}} \rangle_q + v, \\ [\mathcal{Q}^\dagger \psi](\mathbf{x}, \hat{\mathbf{x}}) &= \hat{\mathbf{w}}(\hat{\mathbf{x}} - \langle \hat{\mathbf{x}} \rangle_q), \\ [\mathcal{L}_0^\dagger \mathcal{Q}^\dagger \psi](\mathbf{x}, \hat{\mathbf{x}}) &= (\hat{\mathbf{w}} A^{\text{ES}})\mathbf{x} + (\hat{\mathbf{w}} A^{\text{EE}})\hat{\mathbf{x}}. \end{aligned}$$

Thus, each operator maps the space of affine-linear functions to itself. Additionally, the operator \mathcal{L}_0^\dagger removes the inhomogeneous term v .

We can now evaluate each of the three terms in (4.9) in turn. For the Markovian part, we obtain

$$\langle \mathcal{U}^\dagger \mathcal{L}^\dagger \mathcal{M}^\dagger \phi \rangle_t = \langle \mathcal{L}_1^\dagger \phi \rangle_t + A^{\text{SS}} \langle \mathbf{x} \rangle_t + A^{\text{SE}} \langle \hat{\mathbf{x}} \rangle_q. \quad (4.10)$$

To evaluate the memory term, we first note that, because $\mathcal{M}^\dagger \phi$ does not depend on $\hat{\mathbf{x}}$, we have $\mathcal{Q}^\dagger \mathcal{L}^\dagger \mathcal{M}^\dagger \phi = \mathcal{Q}^\dagger \mathcal{L}_0^\dagger \mathcal{M}^\dagger \phi$. The latter does not depend on \mathbf{x} , so that we obtain $\mathcal{L}^\dagger \mathcal{Q}^\dagger \mathcal{L}^\dagger \mathcal{M}^\dagger \phi = \mathcal{L}_0^\dagger \mathcal{Q}^\dagger \mathcal{L}_0^\dagger \mathcal{M}^\dagger \phi$. Similarly, one can verify that $\mathcal{L}^\dagger \mathcal{Q}^\dagger \psi = \mathcal{L}_0^\dagger \mathcal{Q}^\dagger \psi$.

For the memory term of (4.9), we note that the coefficients of the linear function $e^{(t-t')\mathcal{L}^\dagger\mathcal{Q}^\dagger}\mathcal{L}^\dagger\mathcal{Q}^\dagger\mathcal{L}^\dagger\mathcal{M}^\dagger\phi = e^{(t-t')\mathcal{L}_0^\dagger\mathcal{Q}_0^\dagger}\mathcal{L}_0^\dagger\mathcal{Q}_0^\dagger\mathcal{L}_0^\dagger\mathcal{M}^\dagger\phi$ are the solution of the ODE system

$$\begin{aligned}\dot{\boldsymbol{w}} &= \hat{\boldsymbol{w}}A^{\text{ES}}, \\ \dot{\hat{\boldsymbol{w}}} &= \hat{\boldsymbol{w}}A^{\text{EE}}\end{aligned}$$

at time $t - t'$ with initial conditions at time 0 given by the coefficients of the linear function $\mathcal{L}_0^\dagger\mathcal{Q}_0^\dagger\mathcal{L}_0^\dagger\mathcal{M}^\dagger\phi$. Combining all of this, we obtain for the memory term

$$\begin{aligned}& \int_0^t dt' \left\langle \mathcal{U}^\dagger \mathcal{L}^\dagger \mathcal{Q}^\dagger e^{(t-t')\mathcal{L}^\dagger\mathcal{Q}^\dagger} \mathcal{L}^\dagger \mathcal{M}^\dagger \phi \right\rangle_{t'} \\ &= \int_0^t dt' A^{\text{SE}} e^{(t-t')A^{\text{EE}}} \left\{ A^{\text{ES}} \langle \boldsymbol{x} \rangle_{t'} + A^{\text{EE}} \langle \hat{\boldsymbol{x}} \rangle_q \right\} \\ &= \int_0^t dt' A^{\text{SE}} e^{(t-t')A^{\text{EE}}} A^{\text{ES}} \langle \boldsymbol{x} \rangle_{t'} + A^{\text{SE}} e^{tA^{\text{EE}}} \langle \hat{\boldsymbol{x}} \rangle_q - A^{\text{SE}} \langle \hat{\boldsymbol{x}} \rangle_q.\end{aligned}$$

For the noise term in (4.9), we use the relation $\mathcal{Q}^\dagger e^{t\mathcal{L}^\dagger\mathcal{Q}^\dagger} = \mathcal{Q}^\dagger e^{t\mathcal{Q}^\dagger\mathcal{L}^\dagger\mathcal{Q}^\dagger}$. Since $\mathcal{Q}^\dagger\mathcal{L}^\dagger\mathcal{Q}^\dagger\mathcal{L}_1^\dagger\mathcal{M}^\dagger\phi = 0$, we have

$$\mathcal{Q}^\dagger e^{t\mathcal{Q}^\dagger\mathcal{L}^\dagger\mathcal{Q}^\dagger} \mathcal{L}_1^\dagger \mathcal{M}^\dagger \phi = \mathcal{Q}^\dagger \mathcal{L}_1^\dagger \mathcal{M}^\dagger \phi = 0.$$

We then find

$$\begin{aligned}\left\langle \mathcal{Q}^\dagger e^{t\mathcal{Q}^\dagger\mathcal{L}^\dagger\mathcal{Q}^\dagger} \mathcal{L}^\dagger \mathcal{M}^\dagger \phi \right\rangle_0 &= \left\langle \mathcal{Q}^\dagger e^{t\mathcal{Q}^\dagger\mathcal{L}^\dagger\mathcal{Q}^\dagger} \mathcal{L}_0^\dagger \mathcal{M}^\dagger \phi \right\rangle_0 \\ &= \left\langle \mathcal{Q}^\dagger e^{t\mathcal{Q}^\dagger\mathcal{L}_0^\dagger\mathcal{Q}^\dagger} \mathcal{L}_0^\dagger \mathcal{M}^\dagger \phi \right\rangle_0 \\ &= A^{\text{SE}} e^{tA^{\text{EE}}} (\langle \hat{\boldsymbol{x}} \rangle_0 - \langle \hat{\boldsymbol{x}} \rangle_q).\end{aligned}$$

The full equation for the means thus reads

$$\frac{d}{dt} \langle \boldsymbol{x} \rangle_t = \langle \mathcal{L}_1^\dagger \boldsymbol{x} \rangle_t + A^{\text{SS}} \langle \boldsymbol{x} \rangle_t + \int_0^t dt' A^{\text{SE}} e^{(t-t')A^{\text{EE}}} A^{\text{ES}} \langle \boldsymbol{x} \rangle_{t'} + A^{\text{SE}} e^{tA^{\text{EE}}} \langle \hat{\boldsymbol{x}} \rangle_0. \quad (4.11)$$

We see that the non-linearity of the subnet enters only in the Markovian part.

At this point, we proceed to relate the projection operator formalism for the CME as given by (4.8) and the corresponding mean equation (4.11) to the formalism applied to the RRE model, as given by (4.3). The latter corresponds to the marginal mean equation as given by (4.11) if we choose the initial distribution at time 0 to have the correct mean, $\langle \hat{\boldsymbol{x}} \rangle_0 = \hat{\boldsymbol{\mu}}(0)$. This demonstrates that the marginal CME and the marginal equations obtained from (4.2) are consistent. Interestingly though, in our derivation above, the Markovian, memory and noise terms in (4.9) and (4.11) do not correspond to each other directly: The memory term in (4.9) includes $A^{\text{SE}} \langle \hat{\boldsymbol{x}} \rangle_q$ and $A^{\text{SE}} e^{tA^{\text{EE}}} \langle \hat{\boldsymbol{x}} \rangle_q$, which are canceled, respectively, by terms from the Markovian and noise parts of (4.9).

Note that the RRE and the resulting marginal equation (4.3) involve the replacement of the exact, non-closed moment dynamics $\langle \mathcal{L}_1^\dagger \mathbf{x} \rangle_t + A^{\text{SS}} \langle \mathbf{x} \rangle_t$ by the corresponding reaction rate equation dynamics as given by $\mathbf{f}(\boldsymbol{\mu})$. Thus, the derivation above shows how to obtain a marginal equation for the mean abundances when the subnet-subnet interactions are non-linear.

The results in this section provide some insight into the behavior of the projection operator formalism, but do not themselves result in any simplification for the treatment of heterogeneous dynamics. In the following section, we apply the formalism to heterogeneous dynamics with random rate constants, where an approximate treatment leads to simplified expressions for the marginal dynamics.

4.2 Cumulant expansions for heterogeneous rate constants

We will consider the situation of either the partitioned reaction network (4.1), or the case of heterogeneity modeled by random rate constants \mathbf{c} . In both cases, we thus consider a process (\mathbf{X}, \mathbf{Z}) , where either $\mathbf{Z} = \hat{\mathbf{X}}$ or $\mathbf{Z} = \mathbf{c}$. The main assumption that we will make is that the joint process (\mathbf{X}, \mathbf{Z}) has a feed-forward structure. When considering the case $\mathbf{Z} = \hat{\mathbf{X}}$, this means that the marginal process $\hat{\mathbf{X}}$ is Markovian (and the reaction hazards of reactions that modify $\hat{\mathbf{X}}$ do not depend on \mathbf{X}), and that there are no reactions that simultaneously modify the state of both \mathbf{X} and $\hat{\mathbf{X}}$. The condition is automatically fulfilled in the case of heterogeneous rate constants \mathbf{c} . Additionally, for simplicity we will assume that the process \mathbf{Z} is stationary.

We denote by \underline{z} the entire trajectory, over $(-\infty, \infty)$, of the environment \mathbf{Z} (or, alternatively, the trajectory over an interval $[0, T]$ with T some finite time horizon of interest). Under these assumptions, it follows that the evolution of the (marginal) probability distribution of \mathbf{X} , given the trajectory \underline{z} of the process \mathbf{Z} , is described by a forward equation

$$\frac{d}{dt} p(t, \mathbf{x} | \underline{z}) = \mathcal{L} p(t, \mathbf{x} | \underline{z}) = \sum_{j=1}^{R_0} \hat{f}_j(\mathbf{z}(t)) \mathcal{L}_j p(t, \mathbf{x} | \underline{z}). \quad (4.12)$$

Here $\mathcal{L}_1, \dots, \mathcal{L}_{R_0}$ are the operators corresponding to the R_0 reactions that modify the state of \mathbf{X} , and the $\hat{f}_j(\mathbf{z}(t))$ give the dependence of the reaction hazards on the current state of the environment \mathbf{Z} . Thus, we also assume that the reaction hazards factorize as given by (2.11). This formalism is applicable to all levels of the model hierarchy in Figure 2.1. For example, for the CME, we have

$$\mathcal{L}_j p(\mathbf{x}) = f_j(\mathbf{x} - \boldsymbol{\nu}_j) p(\mathbf{x} - \boldsymbol{\nu}_j) - f_j(\mathbf{x}) p(\mathbf{x}), \quad j = 1, \dots, R,$$

while for the CFPE,

$$\mathcal{L}_j p(\mathbf{x}) = -(\boldsymbol{\nu}_j \cdot \nabla)(f_j(\mathbf{x}) p(\mathbf{x})) + \frac{1}{2} (\boldsymbol{\nu}_j \cdot \nabla)^2 (f_j(\mathbf{x}) p(\mathbf{x})), \quad (4.13)$$

where $\nabla = (\partial/\partial x_1, \dots, \partial/\partial x_N)$. Finally, when using a deterministic description, we can consider the associated Liouville equation, for which the corresponding operators are obtained by dropping the diffusion term in (4.13),

$$\mathcal{L}_j p(\mathbf{x}) = -(\boldsymbol{\nu}_j \cdot \nabla)(f_j(\mathbf{x}) p(\mathbf{x})).$$

In these expressions, $f_j(\mathbf{x})$ denotes the factor of the full reaction rate which gives the dependence on the subnet state \mathbf{x} .

4.2.1 Marginalization via Green's functions

One well-known approach [77, 78] for dealing with random parameters of problems governed by linear PDEs or ODEs, which we also employ in this chapter, relies on Green's functions. Generally speaking, the strategy for obtaining exact or approximate marginal descriptions proceeds along the following steps:

1. We formally solve the equation for a fixed realization of the heterogeneity by means of a Green's function $G(t, t')$,

$$p(t, \mathbf{x} \mid \underline{z}) = G(t, 0)p_0(\mathbf{x}).$$

Note that $G(t, t')$ depends on the realization \underline{z} of the heterogeneity, which we suppress in the notation.

2. We average over heterogeneity to arrive at

$$p(t, \mathbf{x}) = \mathbb{E}[G(t, 0)]p_0(\mathbf{x}).$$

Here we used the assumption that the initial condition $p_0(\mathbf{x})$ is independent of the heterogeneity.

3. We try to obtain a differential equation for the averaged Green's function,

$$\frac{d}{dt}\mathbb{E}[G(t, t')] = V[\mathbb{E}[G(t, t')]],$$

or a differential equation for the marginal solution $p(t, \mathbf{x})$.

In general, for time-dependent heterogeneity, the evolution operator $\mathcal{L}(t)$, for a fixed realization of the heterogeneity \underline{z} , is explicitly time-dependent. The most straightforward way to represent the corresponding Green's function is as an ordered exponential,

$$G(t, t') = \overleftarrow{\mathcal{T}} \exp \left\{ \int_{t'}^t d\tau \mathcal{L}(\tau) \right\} = 1 + \sum_{n=1}^{\infty} \int_{t'}^t dt_1 \int_{t'}^{t_1} dt_2 \cdots \int_{t'}^{t_{n-1}} dt_n \mathcal{L}(t_1) \cdots \mathcal{L}(t_n). \quad (4.14)$$

When combined with the projection operator formalism, this representation will allow us to obtain approximate marginal descriptions in terms of cumulant expansions.

For static heterogeneity, the Green's function is in principle given by an ordinary exponential,

$$G(t, t') = \exp \{ (t - t')\mathcal{L} \}.$$

However, it will often be useful to switch to the interaction picture relative to some chosen, simpler (time-independent) operator $\mathcal{L}^{(0)}$. Recall that this is done by considering the equation

$$\frac{d}{dt}p^{(1)}(t, \mathbf{x}) = \mathcal{L}^{(1)}(t)p^{(1)}(t, \mathbf{x})$$

with

$$p^{(1)}(t, \mathbf{x}) = e^{-t\mathcal{L}^{(0)}} p(t, \mathbf{x}) \quad \text{and} \quad \mathcal{L}^{(1)}(t) = e^{-t\mathcal{L}^{(0)}} (\mathcal{L}(t) - \mathcal{L}^{(0)}) e^{t\mathcal{L}^{(0)}}.$$

If we choose $\mathcal{L}^{(0)} = \mathbb{E}[\mathcal{L}(t)]$, which by assumption on the heterogeneity is time-independent, we have $\mathbb{E}[\mathcal{L}^{(1)}(t)] = 0$. This will be used in Section 4.2.2 to obtain cumulant expansions.

For some simple cases, it is possible to obtain the average over the Green's function in closed form. Here we first treat two such cases.

Example: The birth-death process

Perhaps the simplest possible non-trivial case of heterogeneous dynamics is the reaction rate equation for the birth-death process,

$$\dot{x} = a - c(t)x, \quad x(0) = x_0,$$

with a heterogeneous decay rate c , for which we here assume a CIR(μ, σ^2, γ)-distribution as defined in Section 2.2.1. It turns out that in that case, we can obtain the marginal equation exactly. Starting from the exact solution of the non-random equation with time-dependent decay rate $c(t)$,

$$x(t) = G(t, 0)x_0 + \int_0^t dt' G(t, t')a,$$

in terms of the Green's function for the equation $\dot{x} = -c(t)x$,

$$G(t, t') = \exp \left\{ - \int_{t'}^t d\tau c(\tau) \right\},$$

the average over realizations of the process can be computed in closed form [36]. We find

$$\begin{aligned} \mathbb{E}[G(t, t')] &= \left[\frac{\omega \cosh(\omega\tau) + \frac{\gamma}{2} \sinh(\omega\tau)}{\omega e^{\gamma\tau/2}} \right]^{-\mu^2/\sigma^2} \left[1 + \frac{\sigma^2}{\mu} \frac{\sinh(\omega\tau)}{\omega \cosh(\omega\tau) + \frac{\gamma}{2} \sinh(\omega\tau)} \right]^{-\mu^2/\sigma^2} \\ &= \left[\frac{\omega \cosh(\omega\tau) + (\frac{\gamma}{2} + \frac{\sigma^2}{\mu}) \sinh(\omega\tau)}{\omega e^{\gamma\tau/2}} \right]^{-\mu^2/\sigma^2} \end{aligned}$$

with $\tau = t - t'$ and $\omega = \sqrt{\gamma(\gamma/4 + \sigma^2/\mu)}$. From this we obtain

$$\frac{d}{dt} \mathbb{E}[G(t, t')] = -\mu\rho(t - t') \mathbb{E}[G(t, t')]$$

with

$$\rho(\tau) = \frac{\omega \cosh(\omega\tau) + \frac{\gamma}{2} \sinh(\omega\tau)}{\omega \cosh(\omega\tau) + (\frac{\gamma}{2} + \frac{\sigma^2}{\mu}) \sinh(\omega\tau)}.$$

We then get

$$\frac{d}{dt} \mathbb{E}[x] = -\mu\rho(t) \mathbb{E}[x] + \mathbb{E}[G(t, 0)]a + \mu\rho(t) \int_0^t dt' \mathbb{E}[G(t, t')]a, \quad (4.15)$$

which is an (explicitly time-dependent) equation for the heterogeneity-averaged process.

Example: Commuting operators

A situation in which closed-form expressions for the marginal evolution equation can be obtained is for heterogeneous rate constants when the operators $\mathcal{L}_1, \dots, \mathcal{L}_R$ commute. While this is a very special situation that will essentially never occur for a reaction network, it is nevertheless instructive to see the result. Denote by

$$M_j(\xi_j) = \mathbb{E}[e^{c_j \xi_j}], \quad j = 1, \dots, R,$$

the moment generating functions and by $F_j(\xi_j) = \ln M_j(\xi_j)$ the cumulant generating functions of the (static) heterogeneous parameters. Because the operators commute, the Green's function factorizes,

$$G(t, t') = \exp \left\{ (t - t') \sum_{j=1}^R c_j \mathcal{L}_j \right\} = \prod_{j=1}^R \exp \{ (t - t') c_j \mathcal{L}_j \}.$$

Noting that

$$\frac{d}{dt} M_j(t \mathcal{L}_j) = \mathcal{L}_j M_j'(t \mathcal{L}_j) = \mathcal{L}_j F_j'(t \mathcal{L}_j) M_j(t \mathcal{L}_j),$$

averaging over the heterogeneity and taking the time derivative we find

$$\frac{d}{dt} \mathbb{E}[G(t, t')] = \sum_{j=1}^R \mathcal{L}_j F_j'(t \mathcal{L}_j) \mathbb{E}[G(t, t')].$$

For the marginal probability distributions, we obtain correspondingly

$$\frac{\partial}{\partial t} p(t, \mathbf{x}) = \sum_{j=1}^R \mathcal{L}_j F_j'(t \mathcal{L}_j) p(t, \mathbf{x}).$$

We note that an analogous derivation is possible even when the heterogeneous rate constants are not independent.

As an example, consider the case of a single reaction with Gamma(α, β)-distributed heterogeneity, for which the derivative of the cumulant generating function is $F'(s) = \alpha/(\beta - s)$. We obtain the equation

$$\frac{\partial}{\partial t} p(t, \mathbf{x}) = \alpha(\beta - t \mathcal{L})^{-1} \mathcal{L} p(t, \mathbf{x})$$

which can be formally rewritten as

$$\beta \frac{\partial}{\partial t} p(t, \mathbf{x}) - t \frac{\partial}{\partial t} \mathcal{L} p(t, \mathbf{x}) = \alpha \mathcal{L} p(t, \mathbf{x}). \quad (4.16)$$

If \mathcal{L} corresponds to a linear reaction, say the decay process $X \xrightarrow{c} \emptyset$, the corresponding equation for the first moment

$$\beta \frac{d}{dt} \langle x \rangle - t \frac{d}{dt} \langle \mathcal{L}^\dagger x \rangle = \alpha \langle \mathcal{L}^\dagger x \rangle$$

is closed regardless of whether we consider the CME, the CFPE or the Liouville equation. We obtain

$$\frac{d}{dt} \langle x \rangle = -\frac{\alpha}{\beta + t} \langle x \rangle.$$

To demonstrate an application for several commuting operators, we briefly digress to a situation unrelated to reaction kinetics [79]. Consider a CTMC on the state space $\{0, \dots, N\} \times \{0, \dots, M\}$ where only nearest neighbor transitions to horizontally or vertically adjacent states are allowed, and where the evolution operators for horizontal moves \mathcal{L}_h and for vertical moves \mathcal{L}_v commute, $[\mathcal{L}_h, \mathcal{L}_v] = 0$, as described in [79]. Then the non-trivial marginal dynamics of a random linear combination $c_h \mathcal{L}_h + c_v \mathcal{L}_v$ can be computed using the above framework by solving a single time-dependent ODE. For example, using independent Gamma distributions $c_h \sim \mathcal{G}(\alpha_h, \beta_h)$, $c_v \sim \mathcal{G}(\alpha_v, \beta_v)$, we obtain

$$\partial_t p(t, \mathbf{x}) = \alpha_h \mathcal{L}_h (\beta_h - t \mathcal{L}_h)^{-1} p(t, \mathbf{x}) + \alpha_v \mathcal{L}_v (\beta_v - t \mathcal{L}_v)^{-1} p(t, \mathbf{x}).$$

4.2.2 Cumulant expansions

While the exact marginalizations presented in the previous section provide some insight into the behavior of the heterogeneous dynamics, they are applicable only in special cases. For general reaction networks, some form of approximate marginalization will be necessary. Here we will consider cumulant expansions, which have been proposed in a number of publications [77, 80, 78, 81, 82].

The particular type of cumulant expansion that will be considered here can be derived via the projection operator formalism, as was demonstrated in [78], and as we briefly re-derive in the following. We will apply the formalism not to the evolution equation (4.4), but to the interaction representation.

We define the projection operator \mathcal{P} via the expectation over the heterogeneity, so that

$$\mathcal{P} p(t, \mathbf{x} | \mathbf{z}) = \int d\mathbf{z} p(\mathbf{z}) p(t, \mathbf{x} | \mathbf{z}) = p(t, \mathbf{x}).$$

As explained in Section 4.1, we decompose the evolution equation into two terms,

$$\frac{d}{dt} p^{(1)}(t, \mathbf{x}) = \frac{d}{dt} \mathcal{P} p^{(1)}(t, \mathbf{x} | \mathbf{z}) = \mathcal{P} \mathcal{L}^{(1)}(t) (1 - \mathcal{P}) p^{(1)}(t, \mathbf{x} | \mathbf{z}) \quad (4.17)$$

where we used that in the interaction picture $\mathcal{P} \mathcal{L}^{(1)}(t) \mathcal{P} = 0$ due to the stationarity of the environment process \mathbf{Z} and the choice of operator $\mathcal{L}^{(0)} = \mathbb{E}[\mathcal{L}(t)]$ used to construct the interaction representation, so that one of the two terms that should appear on the right-hand side of (4.17) vanishes. Writing

$$q^{(1)}(t, \mathbf{x}, \mathbf{z}) := (1 - \mathcal{P}) p^{(1)}(t, \mathbf{x} | \mathbf{z}) = p^{(1)}(t, \mathbf{x} | \mathbf{z}) - p^{(1)}(t, \mathbf{x}),$$

we have

$$\frac{d}{dt} q^{(1)}(t, \mathbf{x}, \mathbf{z}) = (1 - \mathcal{P}) \mathcal{L}^{(1)}(t) q^{(1)}(t, \mathbf{x}, \mathbf{z}) + (1 - \mathcal{P}) \mathcal{L}^{(1)}(t) p^{(1)}(t, \mathbf{x}).$$

Similarly to the development in Section 4.1, this linear, time-dependent inhomogeneous equation for q can be solved in terms of the inhomogeneity $p^{(1)}(t, \mathbf{x})$. Unlike in Section 4.1, because the equation is explicitly time-dependent, we use the Green's function

$$\hat{U}(t, t') = \overleftarrow{\mathcal{T}} \exp \left\{ \int_{t'}^t d\tau (1 - \mathcal{P}) \mathcal{L}^{(1)}(\tau) \right\},$$

plugged back into (4.17), to obtain

$$\frac{d}{dt} p^{(1)}(t, \mathbf{x}) = \int_0^t dt' \mathcal{K}^{(1)}(t, t') p^{(1)}(t', \mathbf{x}), \quad (4.18)$$

where the kernel $\mathcal{K}^{(1)}(t, t')$ is given by

$$\mathcal{K}^{(1)}(t, t') = \mathcal{P} \mathcal{L}^{(1)}(t) \hat{U}(t, t') \mathcal{Q} \mathcal{L}^{(1)}(t').$$

This is a closed equation for $p^{(1)}(t, \mathbf{x})$, which however cannot be evaluated directly. However, it serves as a basis for an approximation, and the cumulant expansion is now obtained by expanding the time-ordered exponential as in (4.14), resulting in

$$\mathcal{K}^{(1)}(t, t') = \mathcal{K}_2^{(1)}(t, t') + \mathcal{K}_3^{(1)}(t, t') + \mathcal{K}_4^{(1)}(t, t') + \dots,$$

the first three terms being

$$\begin{aligned} \mathcal{K}_2^{(1)}(t, t') &= \mathcal{P} \mathcal{L}^{(1)}(t) \mathcal{Q} \mathcal{L}^{(1)}(t'), \\ \mathcal{K}_3^{(1)}(t, t') &= \int_{t'}^t dt_1 \mathcal{P} \mathcal{L}^{(1)}(t) \mathcal{Q} \mathcal{L}^{(1)}(t_1) \mathcal{Q} \mathcal{L}^{(1)}(t'), \\ \mathcal{K}_4^{(1)}(t, t') &= \int_{t'}^t dt_1 \int_{t'}^{t_1} dt_2 \mathcal{P} \mathcal{L}^{(1)}(t) \mathcal{Q} \mathcal{L}^{(1)}(t_1) \mathcal{Q} \mathcal{L}^{(1)}(t_2) \mathcal{Q} \mathcal{L}^{(1)}(t'). \end{aligned}$$

The indexes here haven been chosen to reflect the number of times that the operator \mathcal{L} appears, and thus the order of moments of the heterogeneity \mathbf{Z} that will appear in the expressions.

We can now evaluate these terms for the case considered by us. We note that $p^{(1)}(t', \mathbf{x})$, on which the kernel acts in (4.18), does not depend on the heterogeneity \mathbf{Z} . Also, we use the linear dependence of $\mathcal{L}^{(1)}$ on the heterogeneity as in (4.12). Writing for simplicity $\delta_j(t) := \hat{f}_j(\mathbf{z}(t)) - \mathbb{E}[\hat{f}_j(\mathbf{z}(t))]$ and $\bar{\mathcal{L}} = \mathcal{L}^{(0)} = \mathbb{E}[\mathcal{L}(t)]$, we obtain, for terms of order two to four, the expressions

$$\begin{aligned} \mathcal{K}_2(t, t') &= \sum_{j,k=1}^{R_0} C_{jk}(t, t') e^{-t\bar{\mathcal{L}}} \mathcal{L}_j e^{(t-t')\bar{\mathcal{L}}} \mathcal{L}_k e^{t'\bar{\mathcal{L}}}, \\ \mathcal{K}_3(t, t') &= \sum_{j,k,l=1}^{R_0} \int_{t'}^t dt_1 C_{jkl}(t, t_1, t') e^{-t\bar{\mathcal{L}}} \mathcal{L}_j e^{(t-t_1)\bar{\mathcal{L}}} \mathcal{L}_k e^{(t_1-t')\bar{\mathcal{L}}} \mathcal{L}_l e^{t'\bar{\mathcal{L}}}, \\ \mathcal{K}_4(t, t') &= \sum_{i,j,k,l=1}^{R_0} \int_{t'}^t dt_1 \int_{t'}^{t_1} dt_2 C_{ijkl}(t, t_1, t_2, t') e^{-t\bar{\mathcal{L}}} \mathcal{L}_i e^{(t-t_1)\bar{\mathcal{L}}} \mathcal{L}_j e^{(t_1-t_2)\bar{\mathcal{L}}} \mathcal{L}_k e^{(t_2-t')\bar{\mathcal{L}}} \mathcal{L}_l e^{t'\bar{\mathcal{L}}}, \end{aligned}$$

where

$$\begin{aligned} C_{jk}(t, t') &= \mathbb{E}[\delta_j(t)\delta_k(t')], \\ C_{jkl}(t, t_1, t') &= \mathbb{E}[\delta_j(t)\delta_k(t_1)\delta_l(t')], \\ C_{ijkl}(t, t_1, t_2, t') &= \mathbb{E}[\delta_i(t)\delta_j(t_1)\delta_k(t_2)\delta_l(t')] - \mathbb{E}[\delta_i(t)\delta_j(t_1)]\mathbb{E}[\delta_k(t_2)\delta_l(t')]. \end{aligned}$$

These terms are a type of non-commutative cumulants, known as partial cumulants [82]. Note that, while at order two and three, the expectation values do in fact correspond to standard (commutative) cumulants of the corresponding order, this is no longer the case for expectations in the expression for the kernel at order four.

Finally, transforming back from the interaction picture and plugging the expressions for the kernel into (4.18), the expansion in terms of partial cumulants to fourth order reads

$$\begin{aligned} \frac{d}{dt}p(t, \mathbf{x}) &= \bar{\mathcal{L}}p(t, \mathbf{x}) + \sum_{j,k=1}^{R_0} \int_0^t dt' C_{jk}(t, t') \mathcal{L}_j e^{(t-t')\bar{\mathcal{L}}} \mathcal{L}_k p(t', \mathbf{x}) \\ &+ \sum_{j,k,l=1}^{R_0} \int_0^t dt' \int_{t'}^t dt_1 C_{jkl}(t, t_1, t') \mathcal{L}_j e^{(t-t_1)\bar{\mathcal{L}}} \mathcal{L}_k e^{(t_1-t')\bar{\mathcal{L}}} \mathcal{L}_l p(t', \mathbf{x}) \\ &+ \sum_{i,j,k,l=1}^{R_0} \int_0^t dt' \int_{t'}^t dt_1 \int_{t'}^{t_1} dt_2 C_{ijkl}(t, t_1, t_2, t') \mathcal{L}_i e^{(t-t_1)\bar{\mathcal{L}}} \mathcal{L}_j e^{(t_1-t_2)\bar{\mathcal{L}}} \mathcal{L}_k e^{(t_2-t')\bar{\mathcal{L}}} \mathcal{L}_l p(t', \mathbf{x}). \end{aligned} \tag{4.19}$$

This is an integro-differential equation that is inconvenient to directly solve numerically even in the case when the operators \mathcal{L}_j can be represented as finite-dimensional matrices. A standard approach is to introduce auxiliary variables corresponding to the integral terms, and derive differential equations for them. This is done in Appendix A. In any case, for the CME, CFPE or Liouville equation, the corresponding high-dimensional system of differential-difference equations or PDEs cannot be solved either numerically or analytically. Just as for any of the models from the model hierarchy in Figure 2.1, it is possible to derive moment equations, which in general will not be closed. They can be closed using moment closure, and the results obtained in Chapter 3 apply. However, we then have to deal with approximation errors from two different sources: The approximate treatment of heterogeneity using the cumulant expansion, and the error introduced by moment closure.

Here, in order to be able to investigate the performance of the cumulant expansions themselves, we restrict ourselves to processes \mathbf{X} which are linear conditional on a realization of the process \mathbf{Z} . In that case, we can consider the (closed) mean equations for the process \mathbf{X} , and the operators \mathcal{L}_j can be replaced by (finite-dimensional) matrices. We now proceed to apply the cumulant expansion to example problems and to investigate their performance analytically and numerically. We restrict ourselves to expansions of second order.

4.2.3 Example: Birth-death process with CIR-distributed decay rate

We begin by considering a simple birth-death process,



where heterogeneity for the decay reaction c_2 is modeled by a CIR($\mu_2, \sigma_2^2, \gamma_2$) process. Here we introduced the auxiliary species A with constant abundance equal to 1 to obtain a linear system. The matrices corresponding to the birth and death reaction, respectively, taking the place of operators \mathcal{L}_1 and \mathcal{L}_2 , read

$$L_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad L_2 = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}.$$

We begin by investigating the second-order cumulant expansion analytically, which, after an obvious reduction to a one-dimensional problem, reads

$$\frac{d\mathbb{E}[x]}{dt} = c_1 - \mu_2\mathbb{E}[x] + \sigma_2^2 \int_0^t dt' e^{-(\gamma_2 + \mu_2)(t-t')} \mathbb{E}[x(t')].$$

As noted above, the cumulant expansion will depend on the heterogeneity c_2 only through its first two (multi-time) moments. Here, for the second-order expansion, only the stationary mean μ_2 and variance σ_2^2 , as well as the rate γ_2 of decay of correlations, enters the equation. Introducing the integral term as a new variable y , we obtain the linear two-dimensional system

$$\frac{d}{dt} \begin{bmatrix} \mathbb{E}[x] \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ 0 \end{bmatrix} + \begin{bmatrix} -\mu_2 & \sigma_2^2 \\ 1 & -(\mu_2 + \gamma_2) \end{bmatrix} \begin{bmatrix} \mathbb{E}[x] \\ y \end{bmatrix}.$$

The first question to ask is under which conditions the approximation is stable. The eigenvalues of the matrix of this system are

$$\frac{-\gamma_2 - 2\mu_2 \pm \sqrt{\gamma_2^2 + 4\sigma_2^2}}{2},$$

so that the system will be stable when $\sigma_2^2 < \mu_2(\mu_2 + \gamma_2)$. We see that the approximation will be stable for arbitrary variance of the noise as long as fluctuations are fast enough. For static heterogeneity (i.e. $\gamma_2 = 0$) however, we obtain the somewhat restrictive $\sigma_2^2 < \mu_2^2$. Interestingly, this condition coincides precisely with the condition for Gamma-distributed heterogeneity to result in a stationary solution: As long as $\sigma_2^2 < \mu_2^2$, the average $\mathbb{E}[c_1/c_2]$ over the fixed point c_1/c_2 of the RRE corresponding to (4.20) is finite (and the mode of the Gamma distribution is at a point > 0), while when $\sigma_2^2 > \mu_2^2$, we have $\mathbb{E}[c_1/c_2] = \infty$ (and the mode of the Gamma distribution is at 0). This is an interesting observation because the cumulant expansion itself does not rely on any assumption about the distribution of heterogeneity except for the moments. Nevertheless, as has been observed elsewhere [35, 12], Gamma-distributed (or CIR-process distributed) heterogeneity seems to have a particularly natural relation to reaction networks (which for CTMC models is explained by the Gamma distribution being conjugate to the path density). The stationary solution itself, within the approximation, is given by

$$\frac{c_1(\gamma_2 + \mu_2)}{\mu_2(\mu_2 + \gamma_2) - \sigma_2^2}.$$

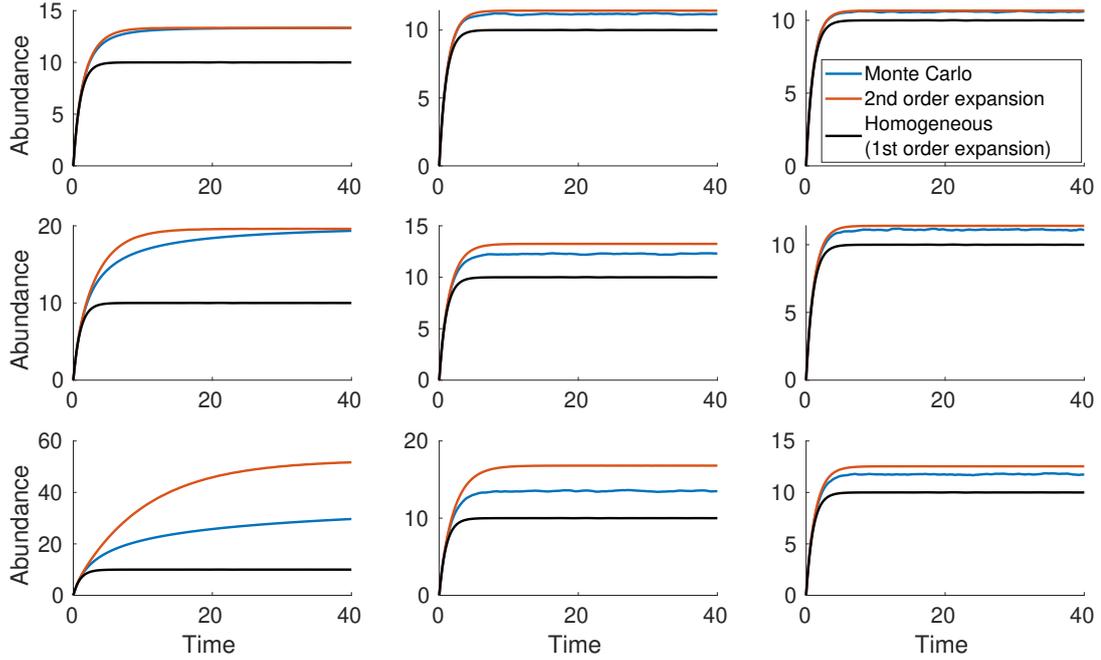


Figure 4.1: Marginal mean dynamics of the birth-death process, with CIR-distributed decay rate. Comparison of Monte Carlo simulations (10,000 simulations for each combination of parameters values) and second-order cumulant expansion. Corresponding homogeneous dynamics (i.e. the first-order expansion) shown for comparison. Parameter values $c_1 = 10$, $\mu_2 = 1$. Rows (top to bottom) correspond to σ_2 equal to 0.5, 0.7 and 0.9. Columns (left to right) corresponding to γ_2 equal to 0.0, 0.5 and 3.0. Initial value $x(0) = 0$.

For the static case $\gamma_2 = 0$ this coincides with the exact result for Gamma distributed heterogeneity.

In Figure 4.1, the performance of the second-order cumulant expansion is investigated numerically. We see that there exists a regime where the heterogeneity is non-negligible (there is a significant difference between homogeneous and heterogeneous mean) and where the cumulant expansion nevertheless produces reasonably accurate results. As expected, as the correlation of the CIR process decays faster (i.e., as γ_2 increases), the approximation becomes more accurate.

4.2.4 Example: Simple gene expression with CIR-distributed heterogeneity

Extending the previous example, we now consider the simple model of gene expression



where X denotes the mRNA and Y the protein. As in the previous example, we consider a heterogeneous decay rate c_2 modeled by a $\text{CIR}(\mu_2, \sigma_2^2, \gamma_2)$ process. Additionally, we make the same assumptions for the decay rate c_4 of the protein, modeling it via

a $\text{CIR}(\mu_4, \sigma_4^2, \gamma_4)$ process. To somewhat reduce the number of parameters for the presentation of the numerical results, we fix $(\mu_2, \sigma_2, \gamma_2) = (1.0, 0.7, 0.0)$, which is a set of parameters for which the birth-death process shows non-negligible heterogeneity and a good performance of the approximation. The performance of the approximation as the parameters (σ_4, γ_4) are varied is shown in Figure 4.2. Here, compared to the birth-death process, the second-order cumulant expansion is able to produce reasonably accurate approximations even when the heterogeneous mean differs significantly from the homogeneous mean.

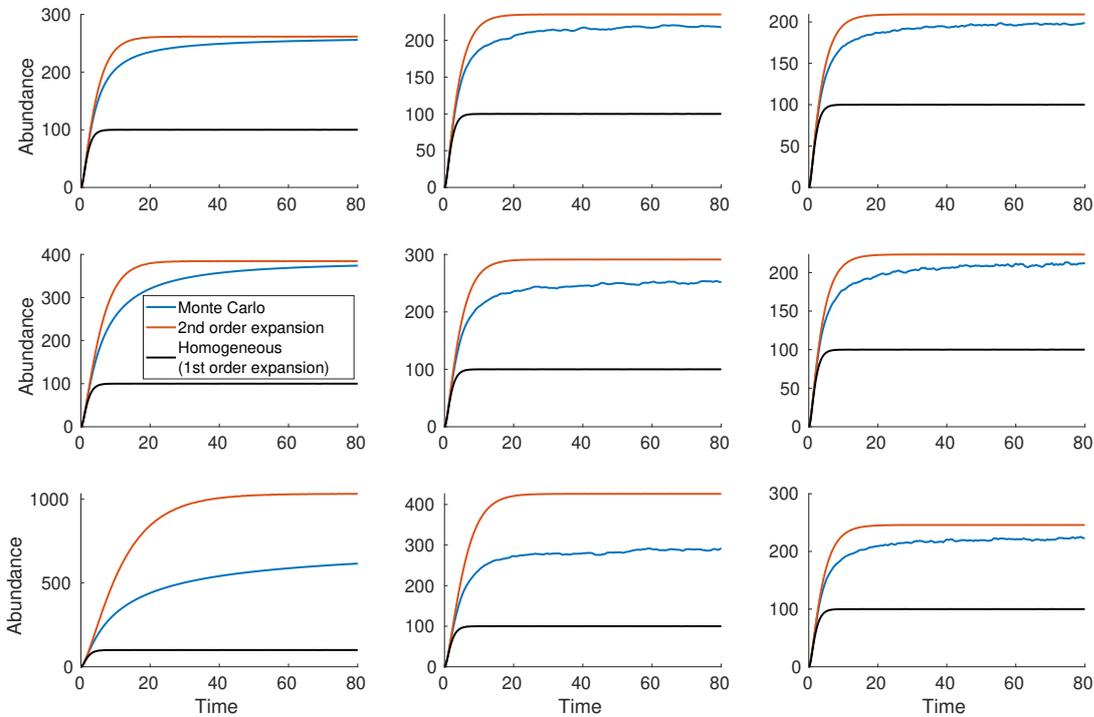


Figure 4.2: Marginal mean dynamics of species Y of the simple gene expression model (4.21), with CIR-distributed decay rates c_2 and c_4 . Comparison of Monte Carlo simulations (10,000 simulations for each combination of parameters values) and second-order cumulant expansion. Corresponding homogeneous dynamics (i.e. the first-order expansion) shown for comparison. Parameter values $c_1 = 10$, $\mu_2 = 1$, $c_3 = 10$, $\mu_4 = 1$ and $\sigma_2 = 0.7$, $\gamma_2 = 0$. Rows (top to bottom) correspond to σ_4 equal to 0.5, 0.7 and 0.9. Columns (left to right) corresponding to γ_4 equal to 0.0, 0.5 and 3.0. Initial values $x(0) = 0$, $y(0) = 0$.

4.2.5 Example: A case with a conservation relation

The last example we consider is a simple network with a conservation relation,



Here the sum of X_1 and X_3 is conserved. We take the rate constants c_1 and c_3 of the reactions involved in the conservation relation as heterogeneous, with $c_1 \sim \text{CIR}(\mu_1, \sigma_1^2, \gamma_1)$ and $c_3 \sim \text{Gamma}(\mu_3, \sigma_3^2)$, where the Gamma distribution is here parameterized in terms of mean and variance. The marginal dynamics of X_2 are shown in Figure 4.3. Again, we see that the second-order cumulant expansion provides a reasonable approximation to the marginal dynamics as long as the magnitude of the heterogeneity is not too large.

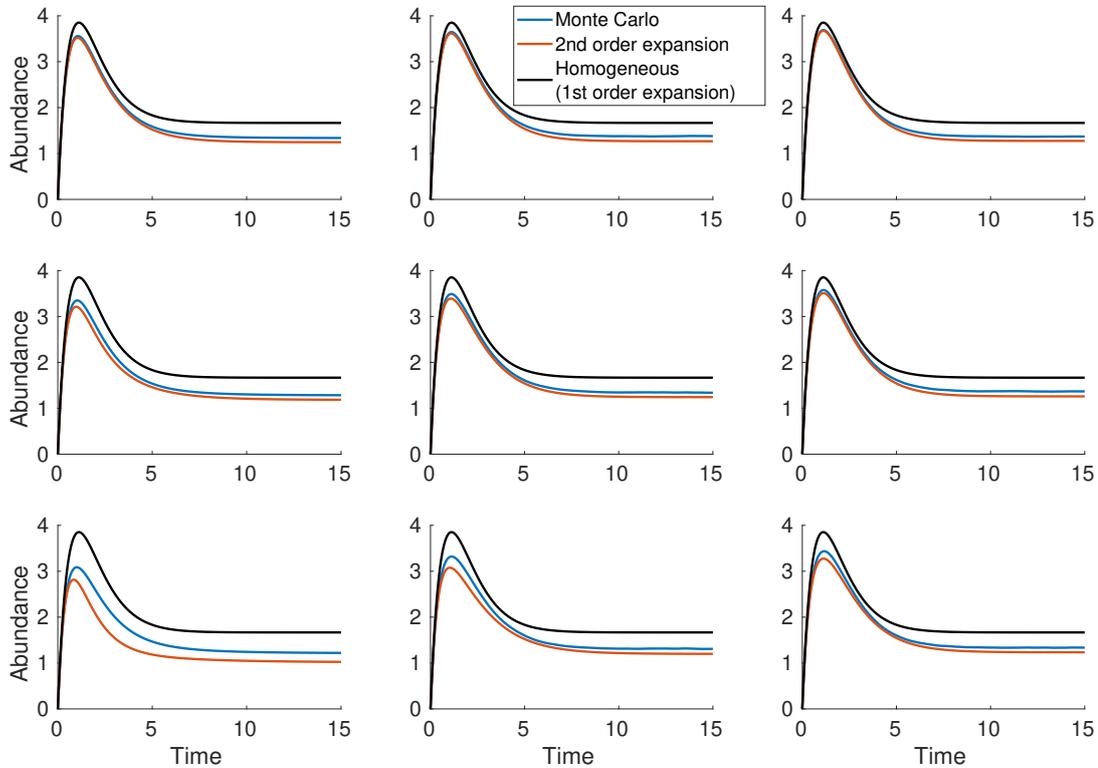


Figure 4.3: Marginal mean dynamics of species X_2 of the reaction network (4.22). Comparison of Monte Carlo simulations (20,000 simulations for each combination of parameters values) and second-order cumulant expansion. Corresponding homogeneous dynamics (i.e. the first-order expansion) shown for comparison. Parameter values $\mu_1 = 1$, $c_2 = 1$, $\mu_3 = 0.2$ and $\sigma_3 = 0.25$. Rows (top to bottom) correspond to σ_1 equal to 0.5, 0.7 and 0.9. Columns (left to right) corresponding to γ_1 equal to 0.0, 1.0 and 2.0. Initial values $x_1(0) = 10$, $x_2(0) = 0$, $x_3(0) = 0$.

In summary, our results indicate that the expansion in terms of partial cumulants can be a useful tool to efficiently obtain marginal dynamics for CIR-distributed heterogeneity, at least for linear systems.

Chapter 5

The marginal process framework for bi-directionally coupled reaction networks

In the previous chapter, a number of applications of the projection operator formalism to the treatment of heterogeneous or marginal reaction kinetics have been discussed. Those approaches operate at the level of forward equations and lead to evolution equations for single-time marginal probability distributions. In this chapter, a marginalization method is discussed which, for the subsystem of interest, operates on the process level.

We continue to treat the problem of marginalizing a reaction network partitioned into subnet and environment. We extend the marginalization approach of [83, 12] to a general reaction network with full coupling between variables of interest and nuisance variables. Marginalization requires the solution of the (in general, infinite-dimensional) filtering equation, which describes the evolution of the conditional probability distribution of the nuisance variables given the trajectory of the marginal process. We use entropic matching, as discussed in Section 3.1.2, to obtain a finite-dimensional approximation. The filtering equation and entropic matching can be interpreted as the result of projection operations consecutively applied to the full master equation of the joint process. In this way, we obtain a principled model reduction method. Our focus is on the marginal process framework as a theoretical tool for model reduction, rather than as a method for more efficient stochastic simulation.

For reaction networks with mass-action kinetics, a particularly simple reduced description is obtained when a product-Poisson ansatz distribution is used for the approximate solution of the filtering equation. We refer to the resulting reduced model as the *Poisson-marginal process* and investigate it in detail. Analogously, for exclusion processes, a product-Bernoulli ansatz distribution leads to what is the simplest possible reduced model within our framework. We investigate this reduced process for the example of the totally asymmetric simple exclusion process (TASEP) on the line with open boundaries.

Finally, for the special case of the marginal process with static, heterogeneous rate constants, we demonstrate how an auxiliary-variable master equation for the marginal process can be derived for CTMC and SDE models.

This chapter includes material from [27] and is organized as follows: After describing the problem setting in Section 5.1, we provide an outline of the proposed method in Section 5.2, using a simple model of constitutive gene expression as a running example. The general form of the marginal process framework is derived in Section 5.3. The finite-dimensional approximations necessary for a tractable description of the marginal process are discussed in Section 5.4, where we also apply our method to the TASEP as a first

example. The Poisson-marginal process for mass-action reaction networks is discussed in Section 5.5. In Section 5.6, the auxiliary-variable master equations for the marginal process with static heterogeneity are derived. Finally, in Section 5.7 the interpretation of the filtering equation as a projection of the full master equation is discussed.

5.1 Setting

It will be useful for conceptual clarity to generalize from our usual setting of a reaction network, and to consider a general Markov jump process (MJP) on an abstract state space. Specifically, we consider an MJP $(X, \hat{X}) = (X(t), \hat{X}(t))_{t \geq 0}$ on a product-form state-space $\mathbb{X} \times \hat{\mathbb{X}}$, where \mathbb{X} and $\hat{\mathbb{X}}$ are countable sets. The case of a reaction network

$$\sum_{n=1}^N s_{nj} X_n + \sum_{n=1}^{\hat{N}} \hat{s}_{nj} \hat{X}_n \longrightarrow \sum_{n=1}^N r_{nj} X_n + \sum_{n=1}^{\hat{N}} \hat{r}_{nj} \hat{X}_n, \quad j = 1, \dots, R \quad (5.1)$$

is then recovered with $\mathbb{X} = \mathbb{N}_0^N$ and $\hat{\mathbb{X}} = \mathbb{N}_0^{\hat{N}}$. For a state $(x, \hat{x}) \in \mathbb{X} \times \hat{\mathbb{X}}$, the rate of transitioning to another state $(y, \hat{y}) \in \mathbb{X} \times \hat{\mathbb{X}}$ over a short time-interval Δt is given by¹

$$\Pr(X(t + \Delta t) = y, \hat{X}(t + \Delta t) = \hat{y} \mid X(t) = x, \hat{X}(t) = \hat{x}) = L(y, \hat{y} \mid x, \hat{x}) \Delta t + o(\Delta t).$$

For the case of a reaction network, in the state $(\mathbf{x}, \hat{\mathbf{x}})$, for each $j = 1, \dots, R$ there exists a transition to the state $(\mathbf{x} + \boldsymbol{\nu}_j, \hat{\mathbf{x}} + \hat{\boldsymbol{\nu}}_j)$ with rate $h_j(\mathbf{x}, \hat{\mathbf{x}})$ and change vector $(\boldsymbol{\nu}_j, \hat{\boldsymbol{\nu}}_j)$ with $\boldsymbol{\nu}_j = (r_{1j} - s_{1j}, \dots, r_{Nj} - s_{Nj})$ and $\hat{\boldsymbol{\nu}}_j = (\hat{r}_{1j} - \hat{s}_{1j}, \dots, \hat{r}_{\hat{N}j} - \hat{s}_{\hat{N}j})$.

In the following, we will also require a description of the process $(\mathbf{X}, \hat{\mathbf{X}})$ in terms of the R reactions. We associate with each reaction channel j a counting process $Y_j(t)$ that counts the number of firings of reaction j over the time interval $[0, t]$. The process (Y_1, \dots, Y_R) can again be seen as a reaction network with values in \mathbb{N}_0^R , which can only change by increments of size 1 in any one of its components at a single time. The state of the original process $(\mathbf{X}, \hat{\mathbf{X}})$ is recovered from the state of these counting processes via

$$\mathbf{X}(t) = \sum_{j=1}^R Y_j(t) \boldsymbol{\nu}_j, \quad \hat{\mathbf{X}}(t) = \sum_{j=1}^R Y_j(t) \hat{\boldsymbol{\nu}}_j.$$

5.2 Outline of the method

Our goal is to derive a marginal process description for the process of interest X . While the joint process (X, \hat{X}) is Markovian, this is no longer the case for the marginal process X . The effect of the nuisance variables \hat{X} is implicitly contained in the memory of the process X . We now illustrate the method on a simple reaction network. We focus on the underlying ideas and postpone derivations to later sections.

A very simple model of constitutive gene expression is given by the reaction network



¹We use bold letters for states of general reaction networks, and ordinary non-bold letters for states of abstract CTMCs.

$\emptyset \rightarrow \text{mRNA}$	$\text{mRNA} \rightarrow \emptyset$	$\text{mRNA} \rightarrow \text{mRNA} + \text{P}$	$\text{P} \rightarrow \emptyset$
Ωc_1	$c_2 \hat{x}$	$c_3 \hat{x}$	$c_4 x$

Table 5.1: Transition rates for the reaction network (5.2) when the process is in state (x, \hat{x}) .

Assuming that we are interested primarily in the protein dynamics, we will consider the mRNA to be a nuisance species. Our goal is to obtain a marginal description of the protein dynamics. Thus, the mRNA plays the role of the nuisance variable \hat{X} and the protein the role of the variable of interest X . Assuming mass-action kinetics, the transition rates of the four reactions in the state (x, \hat{x}) are given in Table 5.1.

The steps to obtain a tractable approximate description of the marginal process are as follows: (i) Determine how the transition rates of the marginal process at time t depend on the process history $x_{[0,t]}$. (ii) Find a description for these marginal transition rates in terms of an evolution equation driven by the process X . The resulting equations are generally infinite-dimensional, but provide an exact description of the marginal process. (iii) Choose an approximation to obtain finite-dimensional equations. We will now carry out these steps for our simple example network.

Description of the marginal process Since the first two reactions in Table 5.1 do not change the state of X , the marginal process consists of two reactions, corresponding to the last two reactions in Table 5.1. Generally, since the marginal process is no longer Markovian, the transition rates for these reactions will depend on the entire history $x_{[0,t]}$ instead of just on the current state $x(t)$. However, the transition rate of the fourth reaction in Table 5.1 does not depend on the mRNA abundance. Consequently, its marginal transition rate remains unchanged and is given by $c_4 x(t)$. In particular, it depends only on the current state $x(t)$ of the marginal process. In contrast to this, the rate for the third reaction does depend on the mRNA abundance. As will be derived in Section 5.3.1, the corresponding marginal transition rate is given by $c_3 \mathbb{E}[\hat{X}(t) \mid x_{[0,t]}]$. This is an intuitive result, expressing the fact that in absence of information about the mRNA abundance, the marginal transition rate is given by the expectation of the transition rate conditional on all available information, i.e. conditional on the entire process history $x_{[0,t]}$.

Filtering equation We are now tasked with computing the expectation $\mathbb{E}[\hat{X}(t) \mid x_{[0,t]}]$. A convenient way to do this is to derive an evolution equation, driven by the marginal process $X(t)$, for the so-called filtering distribution $\pi_t(\hat{x}) := \Pr(\hat{X}(t) = \hat{x} \mid x_{[0,t]})$ with respect to which this expectation is computed. The resulting equation is called the filtering equation. As the process X is a jump process, the trajectory $x_{[0,t]}$ is piecewise constant. The filtering equation for $\pi_t(\hat{x})$ will thus consist of two parts: Continuous evolution (described by a differential equation) as long as X remains constant, and discontinuous jumps whenever X jumps. This is schematically illustrated in Figure 5.1. As will be derived in Section 5.3.3, the continuous evolution is given by

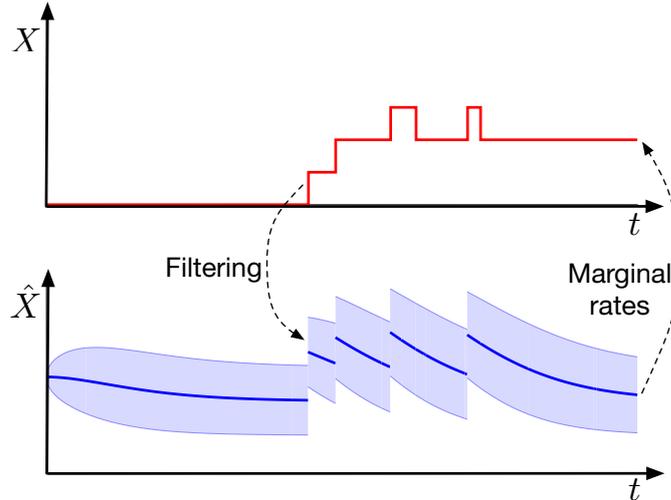


Figure 5.1: Schematic illustration of the concepts involved in the construction of the marginal process, based on the example network (5.2). Blue curve and shaded area show mean and plus/minus one standard deviation of the filtering distribution. The marginal process trajectory (red curve) drives the evolution of the filtering distribution, and the filtering distribution mean determines the transition rates of the marginal process X . Note that jumps of the filtering distribution occur only if the marginal process increases by a jump.

$$\frac{d}{dt}\pi_t(\hat{x}) = \Omega c_1[\pi_t(\hat{x} - 1) - \pi_t(\hat{x})] + c_2[(\hat{x} + 1)\pi_t(\hat{x} + 1) - \hat{x}\pi_t(\hat{x})] - c_3[\hat{x} - \langle \hat{x} \rangle_t]\pi_t(\hat{x}). \quad (5.3)$$

Here the expectation $\langle \hat{x} \rangle_t$ is computed with respect to the distribution $\pi_t(\hat{x})$ itself. The first two terms on the right-hand side of (5.3) simply correspond to the master equation for the mRNA alone, the dynamics of which do not depend on the protein abundance. The last term, however, is of a different form and describes how the information that is contained in the trajectory $x_{[0,t]}$ impacts our state of knowledge about the mRNA abundance. Note that the right-hand side of this equation does not depend on the state of the marginal process X . This is because the reaction network has a “feed-forward” structure. In general, the filtering equation will depend on the state of X .

As explained above, the filtering distribution will also jump instantaneously whenever the driving process X jumps (see Figure 5.1). At a jump of X at time t , the corresponding jump $\pi_{t+} - \pi_{t-}$ of the filtering distribution will depend on which reaction caused the change in X . Since the protein decay reaction does not depend on the mRNA abundance, no information about mRNA abundance is obtained when this reaction fires. Therefore, $\pi_{t+} - \pi_{t-} = 0$ in this case, in analogy to the continuous part (5.3) of the filtering equation in which the protein decay reaction likewise plays no role. When the protein abundance increases via the third reaction in Table 5.1, however, we instantaneously receive a finite amount of information about the mRNA state. To understand this, note for example that reaction three can fire only if there is at least one mRNA molecule present, i.e. $\hat{x} > 0$. Thus, the filtering distribution immediately after the jump, $\pi_{t+}(\hat{x})$, certainly

has to satisfy $\pi_{t+}(0) = 0$. As will be shown in Section 5.3.3, the jump in the filtering distribution when reaction three fires is given by

$$\pi_{t+}(\hat{x}) = \frac{\hat{x}}{\langle \hat{x} \rangle_{t-}} \pi_{t-}(\hat{x}). \quad (5.4)$$

In principle, (5.3) and (5.4) provide a full, exact description of the marginal process, allowing us to compute the marginal transition rates at any time t from the history $x_{[0,t]}$ of the marginal process X . For some simple processes, the corresponding equations can be solved in closed form, as will be demonstrated in Section 5.3.4. In general, however, these equations constitute an infinite-dimensional system that does not provide a sufficiently simple description of the marginal process dynamics. We thus have to look for finite-dimensional approximations.

Finite-dimensional approximation Since we are interested only in the expectation $\langle \hat{x} \rangle_t = \mathbb{E}[\hat{X}(t) \mid x_{[0,t]}]$ of the filtering distribution $\pi_t(\hat{x})$, it seems reasonable to consider the first-order moment equations for (5.3) and (5.4). However, the equations for the mean $\langle \hat{x} \rangle_t$ are not closed, because the second-order moment $\langle \hat{x}^2 \rangle_t$ enters: We obtain

$$\frac{d}{dt} \langle \hat{x} \rangle_t = \Omega c_1 - c_2 \langle \hat{x} \rangle_t - c_3 \left(\langle \hat{x}^2 \rangle_t - \langle \hat{x} \rangle_t^2 \right) \quad (5.5)$$

from (5.3) and

$$\langle \hat{x} \rangle_{t+} = \frac{\langle \hat{x}^2 \rangle_{t-}}{\langle \hat{x} \rangle_{t-}} \quad (5.6)$$

from (5.4). To find a tractable description of the marginal process, we employ moment closure to obtain a finite-dimensional system of equations. As will be explained in more detail in Section 5.5, we want to obtain the simplest possible description of the (approximate) marginal process, and so choose a first-order closure, incorporating the mean of the filtering distribution only. A natural choice for such a closure ansatz, and a principled method in the sense of Chapter 3, is the Poisson distribution. Writing $\theta(t)$ for the mean of the Poisson ansatz distribution, we obtain

$$\frac{d}{dt} \theta(t) = \Omega c_1 - c_2 \theta(t) - c_3 \theta(t) \quad (5.7)$$

from (5.5) and

$$\theta(t+) = \theta(t-) + 1 \quad (5.8)$$

from (5.6). These equations complete our description of the approximate marginal process, which we denote by X' (and refer to as the Poisson-marginal process) to distinguish it from the exact marginal process X . Using (5.7) and (5.8), we can compute the marginal transition rates at time t based on the full history $x'_{[0,t]}$ of the approximate marginal process X' . We use the fact that knowing the history $x'_{[0,t]}$ is equivalent to knowing the histories $(y_3)_{[0,t]}$ and $(y_4)_{[0,t]}$ of the two processes Y_3 and Y_4 (as defined in Section 5.1) that count firings of reactions three and four. Solving (5.7) and (5.8) in

terms of the process histories, we obtain for the marginal rate of the third reaction the expression

$$c_3 \left[e^{-(c_2+c_3)t} \theta(0) + \int_0^t e^{-(c_2+c_3)(t-\tau)} \{ \Omega c_1 d\tau + dy_3(\tau) \} \right].$$

The Stieltjes integral here reduces to a sum, because $(y_3)_{[0,t]}$ is piecewise constant.

In order to evaluate the quality of our chosen approximation, we can compute the mean and the variance of the approximate marginal process X' and of the exact marginal process X . Using results from Section 5.5.1, we find that the means of the exact and approximate marginal processes coincide at all times, assuming the initial conditions are chosen appropriately. At stationarity, the means are given by $\langle x \rangle_\infty = \langle x' \rangle_\infty = \Omega c_1 c_3 / c_2 c_4$. The variances of the processes, however, differ. We compute the relative error of the variance approximation at stationarity and find

$$\frac{\langle x'^2 \rangle_\infty - \langle x^2 \rangle_\infty}{\langle x^2 \rangle_\infty - \langle x \rangle_\infty^2} = \frac{c_3^2}{2c_2(c_2 + c_3 + c_4)}.$$

One particular regime where the error vanishes is time-scale separation, when $c_1, c_2 \rightarrow \infty$ with c_1/c_2 constant. It is thus natural to compare our approach with an approximation that directly invokes time-scale separation. As mentioned in the introduction, there exist a large number of approaches. For our simple network, however, there is one particularly natural option (e.g. [16]): We consider the process $\emptyset \xrightarrow{k_3} \text{Protein} \xrightarrow{c_4} \emptyset$ with the rate constant $k_3 = \Omega c_3 c_1 / c_2$. One easily checks that at stationarity, the time-scale separation ansatz reproduces the correct mean. We compare the approximation of the full distributions at stationarity numerically in Figure 5.2, where we make use of the stochastic simulation algorithm explained in Appendix B to sample from the approximate marginal process. We see that the Poisson-marginal process systematically improves on time-scale separation.

5.3 The marginal process and the filtering equation

In this section, we introduce the marginal process framework in full generality and derive the necessary equations for the case of a general MJP defined on a product-form state-space. We then specialize to the case of reaction networks, where it is useful to additionally introduce a slightly modified version of the marginal process.

5.3.1 The marginal process

As explained in Section 5.2, the marginal process X is in general no longer Markovian, so that the transition rates at time t will depend on the entire history $x_{[0,t]}$ of the process over the time interval $[0, t]$, instead of just on the current state $x(t)$. We now proceed to compute these marginal transition rates in a way analogous to [35].

For the marginal process, the probability for a transition into the state $y \in \mathbb{X}$ to happen in the time interval $[t, t + \Delta t]$, conditional on the process history $x_{[0,t]}$ (and

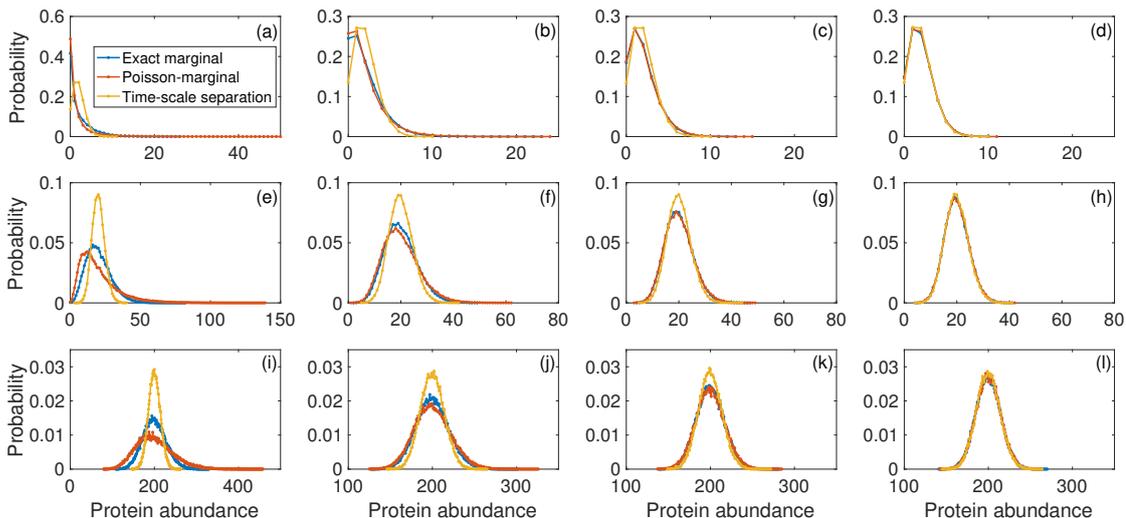


Figure 5.2: Monte Carlo evaluation of the approximation quality of the Poisson-marginal process and of time-scale separation. Distributions of protein abundance at stationarity from 50,000 Monte Carlo runs for each case. Parameters were $c_1 = \gamma$, $c_2 = \gamma/2$, $c_3 = 1$ and $c_4 = 0.1$. Rows correspond to system sizes $\Omega = 0.1$ for (a–d), $\Omega = 1$ for (e–h) and $\Omega = 10$ for (i–l). Columns correspond to mRNA process speeds of $\gamma = 0.5$ in (a,e,i), $\gamma = 2$ in (b,f,j), $\gamma = 5$ in (c,g,k) and $\gamma = 25$ in (d,h,l). Note that the Poisson-marginal process has a somewhat heavier right tail than the exact marginal process, especially at low system size and low value of γ .

assuming $y \neq x(t)$, is given by

$$\begin{aligned}
 & \Pr(X(t + \Delta t) = y \mid x_{[0,t]}) \\
 &= \sum_{\hat{x}, \hat{y}} \Pr(X(t + \Delta t) = y, \hat{X}(t + \Delta t) = \hat{y} \mid X(t) = x(t), \hat{X}(t) = \hat{x}) \\
 & \quad \times \Pr(X(t) = x(t), \hat{X}(t) = \hat{x} \mid x_{[0,t]}) \\
 &= \sum_{\hat{x}, \hat{y}} L(y, \hat{y} \mid x(t), \hat{x}) \Pr(X(t) = x(t), \hat{X}(t) = \hat{x} \mid x_{[0,t]}) \Delta t + o(\Delta t) \\
 &= \mathbb{E}[K(y \mid x(t), \hat{X}(t)) \mid x_{[0,t]}] \Delta t + o(\Delta t),
 \end{aligned}$$

where $K(y \mid x, \hat{x}) = \sum_{\hat{y}} L(y, \hat{y} \mid x, \hat{x})$ is the total rate for jumps from the state (x, \hat{x}) leading to any state in $\{y\} \times \hat{\mathbb{X}}$. Thus, the marginal transition rate is given by

$$\mathbb{E}[K(y \mid x(t), \hat{X}(t)) \mid x_{[0,t]}], \tag{5.9}$$

i.e. by the expectation of the total transition rate conditional on the entire history of the marginal process up to time t . The distribution $\Pr(\hat{X}(t) = \hat{x} \mid x_{[0,t]})$ with respect to which the expectation is computed is the filtering distribution for the stochastic process \hat{X} given the “observed” trajectory $x_{[0,t]}$ of the stochastic process X . The filtering distribution is the solution to the problem of estimating the state of the unobserved variable $\hat{X}(t)$ given the available information $x_{[0,t]}$ about the observed variable.

We see that, in order to obtain a useful description of the marginal process, we require a sufficiently simple description of the filtering distribution, or at least of the marginal transition rates $\mathbb{E}[K(y | x(t), \hat{X}(t)) | x_{[0,t]}]$ that are computed as expectations with respect to the filtering distribution. One way to obtain such a description is to formulate an evolution equation for the filtering distribution driven by the marginal process X . For the case of two fully coupled Markov jump processes, we are not aware of the required results existing in the literature, so we provide an elementary derivation. For an overview of stochastic filtering in general, see [84].

5.3.2 The filtering equation

The filtering distribution $\pi_t(\hat{x}) := \Pr(\hat{X}(t) = \hat{x} | x_{[0,t]})$ is, in principle, defined over the state-space $\hat{\mathbb{X}}$ of the nuisance variable. It is, however, convenient and natural to consider it as a distribution over the joint state-space $\mathbb{X} \times \hat{\mathbb{X}}$ via

$$\begin{aligned} \pi_t(x, \hat{x}) &:= \Pr(X(t) = x, \hat{X}(t) = \hat{x} | x_{[0,t]}) \\ &= \delta_{x(t), x} \Pr(\hat{X}(t) = \hat{x} | x_{[0,t]}), \end{aligned}$$

where $\delta_{x,y}$ is the Kronecker delta. This simply expresses the fact that conditional on $x_{[0,t]}$, the state of $X(t)$ is known to be $x(t)$ with probability one. Depending on the situation, either of these two views will be more convenient, so that in the following, we will repeatedly switch between considering the filtering distribution to be defined either on $\hat{\mathbb{X}}$ or on $\mathbb{X} \times \hat{\mathbb{X}}$.

For the derivations below, the following two operators will be useful: A summation operator \mathcal{S} and a projection operator \mathcal{P}_y (which depends on a state $y \in \mathbb{X}$), both of which act on functions $\psi : \mathbb{X} \times \hat{\mathbb{X}} \rightarrow \mathbb{R}$. They are defined by

$$\begin{aligned} [\mathcal{S}\psi] &= \sum_{x, \hat{x}} \psi(x, \hat{x}), \\ [\mathcal{P}_y\psi](x, \hat{x}) &= \delta_{y, x} \psi(y, \hat{x}). \end{aligned} \tag{5.10}$$

We can now derive the filtering equation. The filtering distribution π_t will evolve according to a differential equation in between jumps of the process X , and will jump whenever X jumps. The intuition here is that, over an infinitesimal time-interval dt , if the observed process X does not jump, we receive only an infinitesimal amount of information so that the change in the filtering distribution should also be infinitesimal. When, however, X does jump, we receive a finite amount of information and correspondingly, the filtering distribution has to jump, too. See also Figure 5.1.

Assuming that we have observed the process X over a time-interval $[0, t + \Delta t]$, these observations can be partitioned into the observations $x_{[0,t]}$ up to time t , and the observation $x(t + \Delta t)$. We assume Δt sufficiently small such that at most one jump occurred

during the time-interval $[t, t + \Delta t]$. Using Bayes' theorem, we have

$$\begin{aligned} & \Pr(\hat{X}(t + \Delta t) = \hat{x} \mid x(t + \Delta t), x_{[0,t]}) \\ &= \sum_{y, \hat{y}} \Pr(X(t + \Delta t) = x(t + \Delta t), \hat{X}(t + \Delta t) = \hat{x} \mid X(t) = y, \hat{X}(t) = \hat{y}) \\ & \quad \times \frac{\Pr(X(t) = y, \hat{X}(t) = \hat{y} \mid x_{[0,t]})}{\Pr(X(t + \Delta t) = x(t + \Delta t) \mid x_{[0,t]})} \\ &= \frac{[e^{\Delta t \mathcal{L}} \pi_t](x(t + \Delta t), \hat{x})}{\sum_{\hat{y}} [e^{\Delta t \mathcal{L}} \pi_t](x(t + \Delta t), \hat{y})} = \frac{\pi_t(x(t + \Delta t), \hat{x}) + \Delta t [\mathcal{L} \pi_t](x(t + \Delta t), \hat{x}) + o(\Delta t)}{\sum_{\hat{y}} \{\pi_t(x(t + \Delta t), \hat{y}) + \Delta t [\mathcal{L} \pi_t](x(t + \Delta t), \hat{y})\} + o(\Delta t)}. \end{aligned}$$

Multiplying this equation by $\delta_{x(t+\Delta t), x}$, using that

$$\pi_{t+\Delta t}(x, \hat{x}) = \delta_{x(t+\Delta t), x} \Pr(\hat{X}(t + \Delta t) = \hat{x} \mid x(t + \Delta t), x_{[0,t]}),$$

and noting the definition of \mathcal{P} and \mathcal{S} in (5.10), we find

$$\pi_{t+\Delta t} = \frac{\mathcal{P}_{x(t+\Delta t)} \pi_t + \Delta t [\mathcal{P}_{x(t+\Delta t)} \mathcal{L} \pi_t] + o(\Delta t)}{\delta_{x(t+\Delta t), x(t)} + \Delta t [\mathcal{S} \mathcal{P}_{x(t+\Delta t)} \mathcal{L} \pi_t] + o(\Delta t)}. \quad (5.11)$$

In the denominator, we also used that

$$\sum_{\hat{y}} \pi_t(x(t + \Delta t), \hat{y}) = \delta_{x(t+\Delta t), x(t)}.$$

We now have to distinguish the cases $x(t + \Delta t) = x(t)$ and $x(t + \Delta t) \neq x(t)$. When $x(t + \Delta t) = x(t)$, i.e. X remained constant over the time-interval $[t, t + \Delta t]$, we have $\mathcal{P}_{x(t+\Delta t)} \pi_t = \pi_t$. Subtracting π_t from (5.11), dividing by Δt and taking the limit $\Delta t \rightarrow 0$, we obtain

$$\frac{d}{dt} \pi_t(x, \hat{x}) = [\mathcal{P}_{x(t)} \mathcal{L} \pi_t](x, \hat{x}) - \pi_t(x, \hat{x}) [\mathcal{S} \mathcal{P}_{x(t)} \mathcal{L} \pi_t]. \quad (5.12)$$

This is the differential equation that the filtering distribution satisfies in between jumps of the process X . It turns out that (5.12) can also be obtained as an orthogonal projection of the full (joint) ME computed with respect to the Fisher-Rao information metric. This point of view is described in Section 5.7 and will allow us to better understand the finite-dimensional approximation of the filtering equation introduced in Section 5.4.2 below.

When $x(t + \Delta t) \neq x(t)$, i.e. when X jumps during the time-interval $[t, t + \Delta t]$, we have $\mathcal{P}_{x(t+\Delta t)} \pi_t = 0$. Taking the limit $\Delta t \rightarrow 0$ in (5.11), we obtain an expression for the filtering distribution immediately after the jump, π_{t+} , in terms of the filtering distribution immediately before the jump, π_{t-} , given by

$$\pi_{t+}(x, \hat{x}) = \frac{[\mathcal{P}_{x(t+)} \mathcal{L} \pi_{t-}](x, \hat{x})}{[\mathcal{S} \mathcal{P}_{x(t+)} \mathcal{L} \pi_{t-}]}, \quad (5.13)$$

where $x(t+)$ is the value of X after the jump.

We now write down expressions (5.12) and (5.13) explicitly in terms of the transition rates. The explicit expressions are simpler if we regard the filtering distribution as being defined only over \hat{X} , i.e. $\pi_t = \pi_t(\hat{x})$. We define

$$R(x, \hat{x}) = \sum_{y \neq x} \sum_{\hat{y}} L(y, \hat{y} \mid x, \hat{x}),$$

the total rate of those transitions out of state (x, \hat{x}) that change the \mathbb{X} -component. In between jumps of X , we then have

$$\begin{aligned} \frac{d}{dt} \pi_t(\hat{x}) = & \sum_{\hat{y}} \{L(x(t), \hat{x} | x(t), \hat{y}) \pi_t(\hat{y}) - L(x(t), \hat{y} | x(t), \hat{x}) \pi_t(\hat{x})\} \\ & - \{R(x(t), \hat{x}) - \langle R(x(t), \hat{x}) \rangle_t\} \pi_t(\hat{x}), \end{aligned} \quad (5.14)$$

where $\langle R(x(t), \hat{x}) \rangle_t = \sum_{\hat{x}} R(x(t), \hat{x}) \pi_t(\hat{x})$ denotes the expectation computed using the filtering distribution π_t . The first term on the right-hand side of (5.14) is an ME for the nuisance component \hat{X} involving only those transitions that do not change the \mathbb{X} -component of the state. Note that the corresponding transition rates can still depend on the current state of X . The second term in (5.14) accounts for the observations. Here the observations contain information by virtue of the fact that X does not jump as long as (5.14) is in effect. From this equation, we also see that the effect of “feedback” from the variable of interest to the nuisance variable is very simple: Because X is constant between its jumps, X is simply fixed to its current value in the transition rates entering (5.14).

When X does jump, so that $x(t+) \neq x(t-)$, the corresponding jump in the filtering distribution is given by

$$\pi_{t+}(\hat{x}) = \frac{\sum_{\hat{y}} L(x(t+), \hat{x} | x(t-), \hat{y}) \pi_t(\hat{y})}{\sum_{\hat{x}', \hat{y}} L(x(t+), \hat{x}' | x(t-), \hat{y}) \pi_t(\hat{y})}. \quad (5.15)$$

The combination of (5.14) and (5.15) with the marginal transition rates (5.9) provides a full description of the marginal process X . For simple processes, these expressions can be evaluated and solved in closed form, as will be demonstrated for a simple reaction network in Section 5.3.4. Before discussing the example, we specialize the discussion to reaction networks.

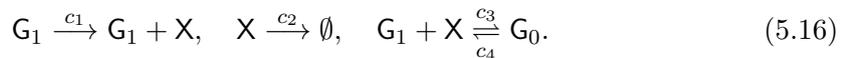
5.3.3 Application to reaction networks

As we can see from (5.14), the transitions of the MJP are naturally partitioned into two groups: Those that change the state of the \mathbb{X} -component, and those that do not. For a reaction network, we will denote by $J_{\mathbb{X}} \subseteq \{1, \dots, R\}$ the subset of indices of those reactions that can modify \mathbf{X} , and by $\overline{J_{\mathbb{X}}} = \{1, \dots, R\} \setminus J_{\mathbb{X}}$ the indices of all remaining reactions. This partitioning also results in a partitioning of the counting processes Y_1, \dots, Y_R (defined in Section 5.1) into two processes $\mathbf{Y} = (Y_j)_{j \in J_{\mathbb{X}}}$ and $\hat{\mathbf{Y}} = (Y_j)_{j \in \overline{J_{\mathbb{X}}}}$ with the former containing the reactions in $J_{\mathbb{X}}$ and the latter the remaining reactions in $\overline{J_{\mathbb{X}}}$. Note that the state of the subnet can then be recovered from \mathbf{Y} alone, while the state of the environment generally requires knowledge of both \mathbf{Y} and $\hat{\mathbf{Y}}$:

$$\mathbf{X}(t) = \sum_{j \in J_{\mathbb{X}}} Y_j(t) \nu_j, \quad \hat{\mathbf{X}}(t) = \sum_{j \in J_{\mathbb{X}}} Y_j(t) \hat{\nu}_j + \sum_{j \in \overline{J_{\mathbb{X}}}} Y_j(t) \hat{\nu}_j.$$

We now specialize the results obtained for the marginal process for general MJPs to the case of reaction networks. At this point, there arises an issue regarding the precise definition of the history of the marginal process on which we condition in the marginal

transition rates (5.9). Generally, it can happen that a reaction network contains two different reactions, with different change vectors $(\nu_i, \hat{\nu}_i)$ and $(\nu_j, \hat{\nu}_j)$, for which however the components corresponding to the subnet are identical, $\nu_i = \nu_j$. For example, this is the case for the simple gene expression model with negative feedback



Here G_0, G_1 are the two possible states of a gene, and X is the gene product that is produced when the gene is in state G_1 . The gene product can also reversibly bind to the gene and switch it to state G_0 , in which production of X is no longer possible. When the gene product X is considered to constitute the subnet, the reactions $G_1 \rightarrow G_1 + X$ and $G_0 \rightarrow G_1 + X$ both lead to an increase of X of size 1. Similarly, $X \rightarrow \emptyset$ and $G_1 + X \rightarrow G_0$ both lead to a decrease of X of size 1.

For such a reaction network, we obtain two different marginal processes depending on whether the history of the process is defined to be just the trajectory $\mathbf{x}_{[0,t]}$ (as was done in Section 5.3.1), or the trajectory $\mathbf{y}_{[0,t]}$ of the counting processes \mathbf{Y} of all reactions that change the subnet. In the former case we will speak of the marginal process \mathbf{X} and in the latter case of the marginal process \mathbf{Y} . Both marginal processes are meaningful, and only minor changes in the derivations presented in Section 5.3.1 and 5.3.2 are necessary. We will present expressions for both cases, because each version of the marginal process has advantages and disadvantages.

The marginal transition rate for the process \mathbf{Y} for reaction $j \in J_{\mathbb{X}}$ is given by

$$\mathbb{E}[h_j(\mathbf{x}(t), \hat{X}(t)) \mid \mathbf{y}_{[0,t]}] \quad (5.17)$$

where $\mathbf{x}(t) = \sum_{j \in J_{\mathbb{X}}} y_j(t) \nu_j$. This is different from (5.9), which for reaction networks reads

$$\sum_j \mathbb{E}[h_j(\mathbf{x}(t), \hat{X}(t)) \mid \mathbf{x}_{[0,t]}]$$

for a transition with change vector ν , and where the summation runs over all $j \in J_{\mathbb{X}}$ such that $\nu_j = \nu$.

The filtering equation, similarly, exists in two variants, depending on which form of the marginal process we consider. It turns out however that the continuous part (5.12) of the filtering equation is the same for both variants and explicitly reads

$$\begin{aligned} \frac{d}{dt} \pi_t(\hat{\mathbf{x}}) &= \sum_{j \in \overline{J_{\mathbb{X}}}} \{h_j(\mathbf{x}(t), \hat{\mathbf{x}} - \hat{\nu}_j) \pi_t(\hat{\mathbf{x}} - \hat{\nu}_j) - h_j(\mathbf{x}(t), \hat{\mathbf{x}}) \pi_t(\hat{\mathbf{x}})\} \\ &\quad - \sum_{j \in J_{\mathbb{X}}} \{h_j(\mathbf{x}(t), \hat{\mathbf{x}}) - \langle h_j(\mathbf{x}(t), \hat{\mathbf{x}}) \rangle_t\} \pi_t(\hat{\mathbf{x}}). \end{aligned} \quad (5.18)$$

For the marginal process \mathbf{X} as defined in Section 5.3.1, the jump in π_t when \mathbf{X} jumps is given by

$$\pi_{t+}(\hat{\mathbf{x}}) = \frac{\sum_j h_j(\mathbf{x}(t-), \hat{\mathbf{x}} - \hat{\nu}_j) \pi_{t-}(\hat{\mathbf{x}} - \hat{\nu}_j)}{\sum_j \langle h_j(\mathbf{x}(t-), \hat{\mathbf{x}}) \rangle_{t-}}, \quad (5.19)$$

where the sums in numerator and denominator each run over all reaction indices $j \in J_{\mathbb{X}}$ such that $\nu_j = \mathbf{x}(t+) - \mathbf{x}(t-)$. If instead we consider the marginal process \mathbf{Y} , a transition

$j \in J_{\mathbf{X}}$ leads to a jump in the filtering distribution given by

$$\pi_{t+}(\hat{\mathbf{x}}) = \frac{h_j(\mathbf{x}(t-), \hat{\mathbf{x}} - \hat{\nu}_j) \pi_{t-}(\hat{\mathbf{x}} - \hat{\nu}_j)}{\langle h_j(\mathbf{x}(t-), \hat{\mathbf{x}}) \rangle_{t-}}. \quad (5.20)$$

The absence of summations in (5.20) will be useful in Section 5.5.1. Here we proceed to discuss a simple example for which only the marginal process \mathbf{X} is useful.

5.3.4 Example: A case with finite-dimensional filtering equations

We consider the simple gene expression model (5.16), with the gene product \mathbf{X} chosen to constitute the subnet. For this model, every reaction changes the state of \mathbf{X} , so that the marginal process \mathbf{Y} would be equal to the full process and thus of no interest. Consequently, we instead consider the (one-dimensional) marginal process X . This process has two reactions, $\emptyset \rightarrow \mathbf{X}$ and $\mathbf{X} \rightarrow \emptyset$, with rates at time t given by

$$c_1\theta(t) + c_4(1 - \theta(t)) \quad \text{and} \quad (c_2 + c_3\theta(t))x(t),$$

respectively, where $\theta(t) = \langle g_1 \rangle_t$ is the filtering distribution mean of the gene state \mathbf{G}_1 , and where we assumed that only a single copy of the gene is present. The filtering distribution $\pi_t(g_0, g_1)$, initially defined on $\{0, 1\} \times \{0, 1\}$, is fully determined by a single number due to the conservation relation $G_0 + G_1 = 1$. Similarly, for the expectation values with respect to $\pi_t(g_0, g_1)$ we have $\theta(t) = \langle g_1 \rangle_t = \pi_t(0, 1) = 1 - \langle g_0 \rangle_t$. We can now write down the (one-dimensional) filtering equation using (5.18) and (5.19). In between jumps of \mathbf{X} , the result reads

$$\frac{d}{dt}\theta(t) = c(x(t))(1 - \theta(t))\theta(t), \quad (5.21)$$

where $c(x(t)) = c_4 - (c_3x(t) + c_1)$. This is solved, for an initial value of $\theta(t_0)$ at time t_0 , by

$$\theta(t) = \frac{\theta(t_0)e^{c(x(t_0))(t-t_0)}}{1 + \theta(t_0)(e^{c(x(t_0))(t-t_0)} - 1)},$$

where we used that $x(t)$ is constant and equal to $x(t_0)$ in between jumps. When the reaction $\emptyset \rightarrow \mathbf{X}$ fires, the filtering distribution mean θ jumps to 1. This is clear because both reactions of (5.16) that cause a change in X of size +1 lead to the gene being in state \mathbf{G}_1 . More interesting is the case when the reaction $\mathbf{X} \rightarrow \emptyset$ fires. Then the jump in the filtering distribution mean is given by

$$\theta(t+) = \frac{c_2\theta(t-)}{c_2 + c_3\theta(t-)}. \quad (5.22)$$

This completes the description of the marginal process. If we consider $\theta(t)$ as an auxiliary variable and use it to augment the process state, the resulting process (X, Θ) is a piecewise-deterministic Markov process with two reactions and deterministic evolution in between jumps given by (5.21). We show a sample from this augmented process in Figure 5.3.

While simple systems such as the one discussed in this section can be treated without approximation, more complicated systems will require an approximate solution of the filtering equations, as was already mentioned in Section 5.2. We address this issue next.

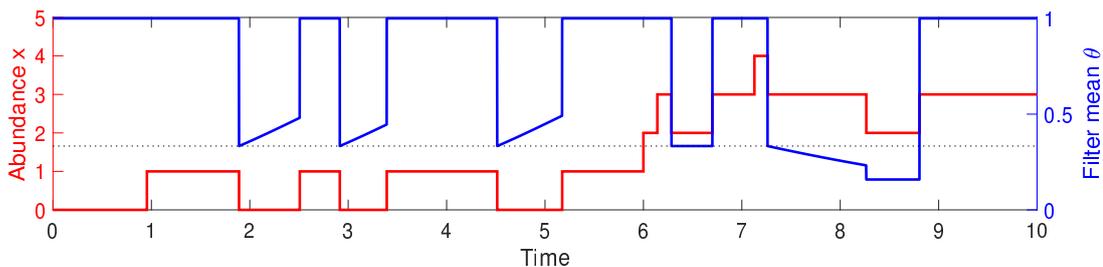


Figure 5.3: Example of a sampled trajectory of the process (X, Θ) . Red curve shows the abundance of gene product X over time. Blue curve shows the state of the filtering distribution mean Θ . Note that, as can be seen from (5.22), jumps in Θ occurring when $\Theta = 1$ always lead to the same value of $\Theta = c_2/(c_2 + c_3)$, indicated by the dotted line. Whether Θ increases or decreases after the jump depends on the value of X . Parameters were $c_1 = 1$, $c_2 = 0.25$, $c_3 = 0.5$ and $c_4 = 2$. Initial state was $X(0) = 0$ and $\Theta(0) = 1$.

5.4 Finite-dimensional approximations of the filtering equation

The filtering equation is in general infinite-dimensional and, just as the ME, far too complicated for either analytical or numerical solution. Thus, regardless of whether one is interested in the marginal process for analytical investigation or for stochastic simulation, an approximate treatment of the filtering equation is necessary.

5.4.1 Moment equations

Following the approach of Chapter 3, we can consider (variational) moment closure for the filtering equations. We first derive the filtering moment equations in their general form, starting from (5.12) and (5.13). The moments of the filtering distribution will evolve according to a differential equation in between jumps of the marginal process, and will jump whenever the marginal process jumps. Again considering π_t to be defined on $\mathbb{X} \times \hat{\mathbb{X}}$, we consider the moment equation for a function $\psi : \mathbb{X} \times \hat{\mathbb{X}} \rightarrow \mathbb{R}$, which can be obtained as demonstrated in (2.13) and is given by

$$\begin{aligned} \frac{d}{dt} \langle \psi \rangle_t &= \sum_{x, \hat{x}} \psi(x, \hat{x}) \{ [\mathcal{P}_{x(t)} \mathcal{L} \pi_t](x, \hat{x}) - \pi_t(x, \hat{x}) [\mathcal{S} \mathcal{P}_{x(t)} \mathcal{L} \pi_t] \} \\ &= \left\langle \mathcal{L}^\dagger \mathcal{P}_{x(t)} [\psi - \langle \psi \rangle_t] \right\rangle_t, \end{aligned} \quad (5.23)$$

in between jumps of X , where we used that $\mathcal{P}_{x(t)}^\dagger = \mathcal{P}_{x(t)}$. When X jumps, we have

$$\langle \psi \rangle_{t+} = \frac{\langle \mathcal{L}^\dagger \mathcal{P}_{x(t+)} \psi \rangle_{t-}}{\langle \mathcal{L}^\dagger \mathcal{P}_{x(t+)} 1 \rangle_{t-}}. \quad (5.24)$$

We skip explicit expressions in terms of rates for general MJPs and instead write down the simpler explicit expressions for reaction networks, which read (now for a function

$\phi : \hat{\mathbb{X}} \rightarrow \mathbb{R}$)

$$\begin{aligned} \frac{d}{dt} \langle \phi \rangle_t &= \sum_{j \in \overline{J_{\mathbb{X}}}} \{ \langle h_j(\mathbf{x}(t), \hat{\mathbf{x}}) \phi(\hat{\mathbf{x}} + \hat{\mathbf{v}}_j) \rangle_t - \langle h_j(\mathbf{x}(t), \hat{\mathbf{x}}) \phi(\hat{\mathbf{x}}) \rangle_t \} \\ &\quad - \sum_{j \in J_{\mathbb{X}}} \{ \langle h_j(\mathbf{x}(t), \hat{\mathbf{x}}) \phi(\hat{\mathbf{x}}) \rangle_t - \langle h_j(\mathbf{x}(t), \hat{\mathbf{x}}) \rangle_t \langle \phi(\hat{\mathbf{x}}) \rangle_t \} \end{aligned} \quad (5.25)$$

in between jumps. Focusing on the marginal process \mathbf{Y} , at a jump of \mathbf{Y} via reaction $j \in J_{\mathbb{X}}$ we have

$$\langle \phi \rangle_{t+} = \frac{\langle h_j(\mathbf{x}(t-), \hat{\mathbf{x}}) \phi(\hat{\mathbf{x}} + \hat{\mathbf{v}}_j) \rangle_{t-}}{\langle h_j(\mathbf{x}(t-), \hat{\mathbf{x}}) \rangle_{t-}}. \quad (5.26)$$

The moment equations are, as is generally the case, not closed: Choosing, for instance, $\phi(\hat{\mathbf{x}}) = \hat{x}_n$ for a reaction network to obtain first-order moments, the resulting equations will depend on moments of order higher than one. Just as for the CME, we require a way to close the system of equations. Note that for the filtering equations, the moment equations will not be closed even for a fully linear system.

Here, we focus on entropic matching, as presented in 3.1.2, to close the equations. The reason for considering entropic matching specifically will become clear in the following.

5.4.2 Entropic matching

The derivation of the entropic matching equations for the filtering equations is quite similar to the derivation for the CME, so that we just summarize it briefly. Choose a parametric family of probability distributions $p_{\boldsymbol{\theta}}(\hat{x})$ depending on parameters $\boldsymbol{\theta}$ ranging in some open subset of \mathbb{R}^K . Assume that, at time t , we have an approximation $p_{\boldsymbol{\theta}(t)}(\hat{x})$ of the filtering distribution $\pi_t(\hat{x})$ available. As for the filtering distribution itself, we identify the approximation $p_{\boldsymbol{\theta}}(\hat{x})$ on $\hat{\mathbb{X}}$ with $p_{\boldsymbol{\theta}}(x, \hat{x}) = \delta_{x(t), x} p_{\boldsymbol{\theta}}(\hat{x})$ on $\mathbb{X} \times \hat{\mathbb{X}}$.

We first consider the continuous part of the filtering equation. Then a short time-step Δt later, $p_{\boldsymbol{\theta}(t)}$ will have evolved to

$$p(x, \hat{x}) = p_{\boldsymbol{\theta}(t)}(x, \hat{x}) + \Delta t [\mathcal{P}_{x(t)} \mathcal{L} p_{\boldsymbol{\theta}(t)}](x, \hat{x}) - \Delta t p_{\boldsymbol{\theta}(t)}(x, \hat{x}) [\mathcal{S} \mathcal{P}_{x(t)} \mathcal{L} p_{\boldsymbol{\theta}(t)}].$$

We will obtain an approximation to $p(x, \hat{x})$ that lies in the parametric family $p_{\boldsymbol{\theta}}$ by choosing parameters $\boldsymbol{\theta}(t + \Delta t)$ to minimize the relative entropy $D[p \parallel p_{\boldsymbol{\theta}(t + \Delta t)}]$. We then take the limit $\Delta t \rightarrow 0$ to obtain an ordinary differential equation (ODE) for the parameters $\boldsymbol{\theta}$. Write, for brevity, $\boldsymbol{\theta} = \boldsymbol{\theta}(t)$ and $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}(t + \Delta t)$. Then we have, to first order in Δt ,

$$\begin{aligned} D[p \parallel p_{\tilde{\boldsymbol{\theta}}}] &= \left\langle \ln \frac{p_{\boldsymbol{\theta}} + \Delta t \{ \mathcal{P}_{x(t)} \mathcal{L} p_{\boldsymbol{\theta}} - p_{\boldsymbol{\theta}} [\mathcal{S} \mathcal{P}_{x(t)} \mathcal{L} p_{\boldsymbol{\theta}}] \}}{p_{\tilde{\boldsymbol{\theta}}}} \right\rangle_p \\ &= \left\langle \ln \frac{p_{\boldsymbol{\theta}}}{p_{\tilde{\boldsymbol{\theta}}}} \right\rangle_{\boldsymbol{\theta}} + \Delta t \left[\left\langle \left\{ \frac{\mathcal{P}_{x(t)} \mathcal{L} p_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}}} - [\mathcal{S} \mathcal{P}_{x(t)} \mathcal{L} p_{\boldsymbol{\theta}}] \right\} \ln \frac{p_{\boldsymbol{\theta}}}{p_{\tilde{\boldsymbol{\theta}}}} \right\rangle_{\boldsymbol{\theta}} + \text{const} \right] \end{aligned}$$

where ‘‘const’’ denotes terms independent of $\tilde{\boldsymbol{\theta}}$, and $\langle \cdot \rangle_{\boldsymbol{\theta}}$ denotes an expectation taken with respect to the distribution $p_{\boldsymbol{\theta}}$. The first term is simply equal to $D[p_{\boldsymbol{\theta}} \parallel p_{\tilde{\boldsymbol{\theta}}}]$, which to second order in $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is given by

$$D[p_{\boldsymbol{\theta}} \parallel p_{\tilde{\boldsymbol{\theta}}}] = \frac{1}{2} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^\dagger G(\boldsymbol{\theta}) (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

where $G(\boldsymbol{\theta})$ is the Fisher information matrix of the parametric family $p_{\boldsymbol{\theta}}$ at parameter value $\boldsymbol{\theta}$, the components of which are given by

$$G_{kl}(\boldsymbol{\theta}) = \left\langle \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_k} \frac{\partial \ln p_{\boldsymbol{\theta}}}{\partial \theta_l} \right\rangle_{\boldsymbol{\theta}}, \quad k, l = 1, \dots, K. \quad (5.27)$$

To minimize $D[p \parallel p_{\tilde{\boldsymbol{\theta}}}]$, we take the derivative with respect to $\tilde{\boldsymbol{\theta}}$ and obtain

$$\begin{aligned} 0 &= G(\boldsymbol{\theta})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \Delta t \left\langle \left\{ \frac{\mathcal{P}_{x(t)} \mathcal{L} p_{\boldsymbol{\theta}}}{p_{\boldsymbol{\theta}}} - [\mathcal{S} \mathcal{P}_{x(t)} \mathcal{L} p_{\boldsymbol{\theta}}] \right\} \nabla_{\tilde{\boldsymbol{\theta}}} \ln p_{\tilde{\boldsymbol{\theta}}} \right\rangle_{\boldsymbol{\theta}} \\ &= G(\boldsymbol{\theta})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) - \Delta t \left\langle \left\{ \mathcal{L}^\dagger \mathcal{P}_{x(t)}^\dagger \nabla_{\tilde{\boldsymbol{\theta}}} \ln p_{\tilde{\boldsymbol{\theta}}} \right\}_{\boldsymbol{\theta}} + [\mathcal{S} \mathcal{P}_{x(t)} \mathcal{L} p_{\boldsymbol{\theta}}] \langle \nabla_{\tilde{\boldsymbol{\theta}}} \ln p_{\tilde{\boldsymbol{\theta}}} \rangle_{\boldsymbol{\theta}} \right\}. \end{aligned}$$

Dividing by Δt , taking the limit $\Delta t \rightarrow 0$ and using that $\langle \nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}} \rangle_{\boldsymbol{\theta}} = 0$, we get

$$\frac{d}{dt} \boldsymbol{\theta} = G(\boldsymbol{\theta})^{-1} \left\langle \mathcal{L}^\dagger \mathcal{P}_{x(t)}^\dagger \nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}}. \quad (5.28)$$

This is a closed equation for the parameters $\boldsymbol{\theta}$. Using the resulting approximate solution $p_{\boldsymbol{\theta}}$ of the filtering equation, all necessary expectations, in particular the marginal transition rates, can be computed.

When the process X jumps, the filtering distribution jumps according to (5.13), so that the approximation $p_{\boldsymbol{\theta}(t-)}$ immediately before the jump is updated to

$$p = \frac{\mathcal{P}_{x(t+)} \mathcal{L} p_{\boldsymbol{\theta}(t-)}}{[\mathcal{S} \mathcal{P}_{x(t+)} \mathcal{L} p_{\boldsymbol{\theta}(t-)}]}.$$

Here too we can obtain an updated approximation within the parametric family by minimizing the relative entropy, i.e. choosing $\boldsymbol{\theta}(t+)$ to minimize $D[p \parallel p_{\boldsymbol{\theta}(t+)}]$. In general, this will be impractical. However, usually one will choose $p_{\boldsymbol{\theta}}$ to be an exponential family

$$p_{\boldsymbol{\theta}}(\hat{x}) = \frac{1}{Z(\boldsymbol{\theta})} \exp \left\{ \sum_{k=1}^K \theta_k \phi_k(\hat{x}) \right\} q(\hat{x}). \quad (5.29)$$

In this case, minimizing the relative entropy amounts to matching moments, i.e. choosing $\boldsymbol{\theta}(t+)$ so that $\langle \phi_k \rangle_{\boldsymbol{\theta}(t+)} = \langle \phi_k \rangle_p$ for $k = 1, \dots, K$, which is often practical.

The entropic matching equations (applied in the context of filtering for stochastic differential equations) were first proposed in [51] and derived using a projection argument employing the Fisher-Rao information metric. This geometrical approach to (5.28), which we describe in Section 5.7, is completely analogous to the projection leading to the filtering equation. In this way, entropic matching is seen to be a very natural way to produce a finite-dimensional approximation to the filtering equation, in addition to the justification provided above (which, in turn, has the advantage of allowing a unified treatment of the continuous and discrete parts of the filtering equation).

5.4.3 Example: The totally asymmetric exclusion process

In this section, we will apply the marginal process framework to the TASEP on the line with open boundaries. The TASEP [85] describes particles hopping on N sites

X_1, \dots, X_N , where each site can be occupied by at most one particle. We take $X_n = 1$ when site X_n is occupied, and $X_n = 0$ otherwise. If the first site X_1 is empty, a particle can enter at a rate α . If site X_{n+1} is empty and site X_n occupied, a particle can move from X_n to X_{n+1} with rate c . Finally, a particle at the last site X_N can leave the system with rate β .

We consider the situation where only the dynamics of the last site X_N is of interest to us, which might serve as a proxy for, say, the flux through the entire system. Thus, the only transitions which will be retained are the two transitions corresponding to a particle entering or leaving site X_N . The filtering moment equations for the mean occupancies read

$$\begin{aligned}\frac{d}{dt} \langle x_1 \rangle_t &= \alpha \langle 1 - x_1 \rangle_t - c \langle x_1(1 - x_2) \rangle_t, \\ \frac{d}{dt} \langle x_n \rangle_t &= c \langle x_{n-1}(1 - x_n) \rangle_t - c \langle x_n(1 - x_{n+1}) \rangle_t, \quad n = 2, \dots, N - 2, \\ \frac{d}{dt} \langle x_{N-1} \rangle_t &= c \langle x_{N-2}(1 - x_{N-1}) \rangle_t - c(1 - x_N(t)) \langle x_{N-1} \rangle_t \langle 1 - x_{N-1} \rangle_t.\end{aligned}$$

As expected and is well known, these contain second-order moments. Here we are interested in obtaining the simplest possible approximate marginal process. Thus, we will obtain closed equations in terms of the first-order moments $\langle x_1 \rangle_t, \dots, \langle x_{N-1} \rangle_t$ only. A very natural approach to obtain such a closure is to use entropic matching with a product-Bernoulli distribution ansatz:

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \prod_{n=1}^{N-1} \theta_n^{x_n} (1 - \theta_n)^{1-x_n}.$$

We refer to the resulting approximate marginal process as the *Bernoulli-marginal* TASEP. After application of product-Bernoulli entropic matching, the closed filtering moment equations, in between observations, are given by

$$\begin{aligned}\frac{d}{dt} \theta_1(t) &= \alpha(1 - \theta_1(t)) - c\theta_1(t)(1 - \theta_2(t)), \\ \frac{d}{dt} \theta_n(t) &= c\theta_{n-1}(t)(1 - \theta_n(t)) - c\theta_n(t)(1 - \theta_{n+1}(t)), \quad n = 2, \dots, N - 2, \\ \frac{d}{dt} \theta_{N-1}(t) &= c\theta_{N-2}(t)(1 - \theta_{N-1}(t)) - c(1 - x_N(t))\theta_{N-1}(t)(1 - \theta_{N-1}(t)).\end{aligned}$$

Unsurprisingly, the resulting equations are identical to a “naive” mean-field approximation. Note that when $x_N(t) = 1$, i.e. the last site is occupied, the observation term (the last term of the last line) vanishes because a particle cannot enter the last site.

When a particle leaves site X_N , no update to the filtering distribution moments is required. When a particle enters site X_N at time t , the update is simply given by

$$\begin{aligned}\theta_n(t+) &= \theta_n(t-), \quad n = 1, \dots, N - 2, \\ \theta_{N-1}(t+) &= 0.\end{aligned}$$

This is intuitively clear: Site X_{N-1} is necessarily empty immediately after a particle enters site X_N . The means of the remaining sites are left unchanged because of the product-form closure employed.

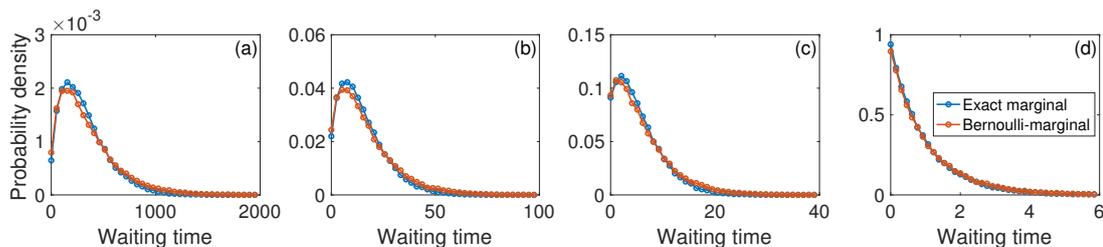


Figure 5.4: Numerical evaluation of the accuracy of the Bernoulli-marginal TASEP approximation on $N = 10$ sites at stationarity. Waiting-time distributions for a particle to enter site X_N after the previous particle left X_N . Parameters were $\alpha = \beta = 1$ and $c = 0.01$ in (a), $c = 0.1$ in (b), $c = 1$ in (c) and $c = 2$ in (d). Distributions from 100,000 samples.

We performed Monte Carlo simulations of both the Bernoulli-marginal TASEP and of the full TASEP to compare their behavior. In Figure 5.4, we plot the distribution of waiting times between a particle leaving site X_N and the next particle entering X_N when the process is at stationarity. The waiting-time distribution varies depending on the parameters of the process. The Bernoulli-marginal process reproduces the exact results with high accuracy, despite the fact that we have used a very simple closure for the filtering equation.

5.5 The product-Poisson marginal process

In this section, we will apply our results to general reaction networks. While the marginal process framework described in Section 5.3 and 5.4 is very flexible, it does not provide any indication of how to close the filtering moment equations. Each choice of closure leads to a different (approximate) marginal process. In Section 5.4.3, we chose what is presumably the simplest-possible non-trivial closure (depending on a single parameter for each variable of the filtering equation) for the TASEP.

In this section, in order to initiate the systematic study of the marginal process framework for reaction networks, we similarly investigate what is arguably the simplest non-trivial closure for reaction networks with mass-action kinetics. Throughout, we focus on the marginal process \mathbf{Y} as introduced in Section 5.3.3, which is more convenient here.

5.5.1 The product-Poisson closure

We consider a general reaction network (5.1) with mass-action rates. In Section 5.4.3, we employed product-Bernoulli entropic matching, which lead to naive mean-field equations. In the context of general reaction networks, we note that a naive first-order mean-field closure, in which the variance is set to zero, leads to the vanishing of the second term (corresponding to the observations) in (5.25). Instead, we will obtain a principled closure by employing entropic matching using a product-Poisson distribution

$$p_{\theta}(\hat{x}) = \prod_{n=1}^{\hat{N}} e^{-\theta_n} \frac{\theta_n^{\hat{x}_n}}{\hat{x}_n!}.$$

Applying product-Poisson entropic matching to a mass-action reaction network leads to equations for the Poisson means that coincide with the macroscopic reaction rate equations, as shown in Section 3.2.1. This result will then also hold for the first term in (5.25), which corresponds to the “prior” evolution of the environment species. In this sense, product-Poisson entropic matching behaves similar (though not identical) to a naive mean-field closure. However, unlike for a naive first-order closure, the term corresponding to the observations in (5.25) does not vanish. We obtain

$$\frac{d}{dt}\boldsymbol{\theta} = \sum_{j \in \overline{J_{\mathbf{x}}}} \Omega c_j f_j(\mathbf{x}(t)) \hat{g}_j(\boldsymbol{\theta}) \hat{\boldsymbol{\nu}}_j - \sum_{j \in J_{\mathbf{x}}} \Omega c_j f_j(\mathbf{x}(t)) \hat{g}_j(\boldsymbol{\theta}) \hat{\boldsymbol{s}}_j \quad (5.30)$$

for the continuous part of the filtering equation, and

$$\boldsymbol{\theta}(t+) = \boldsymbol{\theta}(t-) + \hat{\boldsymbol{r}}_j \quad (5.31)$$

when the marginal process jumps via reaction j , where $\hat{\boldsymbol{r}}_j = (\hat{r}_{1j}, \dots, \hat{r}_{N_j j})$ and $\hat{\boldsymbol{s}}_j = (\hat{s}_{1j}, \dots, \hat{s}_{N_j j})$. Here \hat{g}_j are the mass-action rates for the environment species in their macroscopic form, given by

$$\hat{g}_j(\boldsymbol{\theta}) = \prod_{n=1}^{\hat{N}} \left(\frac{\theta_n}{\Omega} \right)^{\hat{s}_{nj}}.$$

The simplicity of (5.31) is the reason for considering the marginal process \mathbf{Y} . If we instead consider the marginal process \mathbf{X} , the corresponding equation is more complicated, and the results obtained in the following would not hold.

In order to better understand the Poisson-marginal process, we can investigate its moment equations. For this purpose, we consider the augmented process $(\mathbf{X}', \boldsymbol{\Theta})$, where \mathbf{X}' is the approximate marginal process and $\boldsymbol{\Theta}$ is the stochastic process corresponding to the Poisson means $\boldsymbol{\theta}$. As in Section 5.3.4, $(\mathbf{X}', \boldsymbol{\Theta})$ is a piecewise deterministic Markov process. For such a process, from the known form of the backwards evolution operator [86], we obtain the moment equation for a function $\psi(\mathbf{x}, \boldsymbol{\theta})$ in the form

$$\begin{aligned} \frac{d}{dt} \langle \psi \rangle &= \sum_{j \in \overline{J_{\mathbf{x}}}} \left\langle \Omega c_j f_j(\mathbf{x}) \hat{g}_j(\boldsymbol{\theta}) \hat{\boldsymbol{\nu}}_j^\dagger \nabla_{\boldsymbol{\theta}} \psi \right\rangle - \sum_{j \in J_{\mathbf{x}}} \left\langle \Omega c_j f_j(\mathbf{x}) \hat{g}_j(\boldsymbol{\theta}) \hat{\boldsymbol{s}}_j^\dagger \nabla_{\boldsymbol{\theta}} \psi \right\rangle \\ &+ \sum_{j \in J_{\mathbf{x}}} \langle \Omega c_j f_j(\mathbf{x}) \hat{g}_j(\boldsymbol{\theta}) [\psi(\mathbf{x} + \boldsymbol{\nu}_j, \boldsymbol{\theta} + \hat{\boldsymbol{r}}_j) - \psi(\mathbf{x}, \boldsymbol{\theta})] \rangle. \end{aligned} \quad (5.32)$$

Here and in the following, we consider all vectors as column vectors. In particular (writing for brevity $\chi_j = \chi_j(\mathbf{x}, \boldsymbol{\theta}) = \Omega c_j f_j(\mathbf{x}) \hat{g}_j(\boldsymbol{\theta})$), the first-order moment equations are given by

$$\begin{aligned} \frac{d}{dt} \langle \mathbf{x} \rangle &= \sum_{j \in J_{\mathbf{x}}} \langle \chi_j \rangle \boldsymbol{\nu}_j = \sum_{j=1}^R \langle \chi_j \rangle \boldsymbol{\nu}_j, \\ \frac{d}{dt} \langle \boldsymbol{\theta} \rangle &= \sum_{j \in \overline{J_{\mathbf{x}}}} \langle \chi_j \rangle \hat{\boldsymbol{\nu}}_j - \sum_{j \in J_{\mathbf{x}}} \langle \chi_j \rangle \hat{\boldsymbol{s}}_j + \sum_{j \in J_{\mathbf{x}}} \langle \chi_j \rangle \hat{\boldsymbol{r}}_j \\ &= \sum_{j=1}^R \langle \chi_j \rangle \hat{\boldsymbol{\nu}}_j, \end{aligned}$$

where we used that $\hat{\nu}_j = \hat{r}_j - \hat{s}_j$ and, by definition of $\overline{J_{\mathbf{x}}}$, $\nu_j = 0$ for each $j \in \overline{J_{\mathbf{x}}}$. For a linear reaction network, these equations are identical to the first-order moment equations obtained for the full process $(\mathbf{X}, \hat{\mathbf{X}})$. Since these equations are closed, we see that the Poisson-marginal process, for a linear reaction network, reproduces the mean of the exact marginal process.

Similarly, we can investigate the relation between the covariance matrices of the Poisson-marginal and the full process by considering the second-order moment equations, which for the augmented Poisson-marginal process (\mathbf{X}', Θ) are given by

$$\begin{aligned}\frac{d}{dt} \langle \mathbf{x} \mathbf{x}^\dagger \rangle &= \sum_{j \in J_{\mathbf{x}}} \langle \chi_j [\mathbf{x} \nu_j^\dagger + \nu_j \mathbf{x}^\dagger + \nu_j \nu_j^\dagger] \rangle, \\ \frac{d}{dt} \langle \mathbf{x} \theta^\dagger \rangle &= \sum_{j=1}^R \langle \chi_j \mathbf{x} \hat{\nu}_j^\dagger \rangle + \sum_{j \in J_{\mathbf{x}}} \langle \chi_j [\nu_j \theta^\dagger + \nu_j \hat{r}_j^\dagger] \rangle, \\ \frac{d}{dt} \langle \theta \theta^\dagger \rangle &= \sum_{j=1}^R \langle \chi_j [\theta \hat{\nu}_j^\dagger + \hat{\nu}_j \theta^\dagger] \rangle + \sum_{j \in J_{\mathbf{x}}} \langle \chi_j \hat{r}_j \hat{r}_j^\dagger \rangle.\end{aligned}$$

Denote by

$$S = \begin{bmatrix} \langle \mathbf{x} \mathbf{x}^\dagger \rangle & \langle \mathbf{x} \hat{\mathbf{x}}^\dagger \rangle \\ \langle \hat{\mathbf{x}} \mathbf{x}^\dagger \rangle & \langle \hat{\mathbf{x}} \hat{\mathbf{x}}^\dagger \rangle \end{bmatrix} \quad \text{and} \quad S' = \begin{bmatrix} \langle \mathbf{x} \mathbf{x}^\dagger \rangle & \langle \mathbf{x} \theta^\dagger \rangle \\ \langle \theta \mathbf{x}^\dagger \rangle & \langle \theta \theta^\dagger \rangle \end{bmatrix}$$

the matrices of second-order moments for the full process $(\mathbf{X}, \hat{\mathbf{X}})$ and for the augmented Poisson-marginal process (\mathbf{X}', Θ) , respectively. For a linear reaction network, these then evolve according to

$$\begin{aligned}\frac{d}{dt} S &= AS + SA^\dagger + B(t), \\ \frac{d}{dt} S' &= AS' + S'A^\dagger + B'(t),\end{aligned}\tag{5.33}$$

with matrices $A, B(t)$ and $B'(t)$. The difference between the matrices $B(t)$ and $B'(t)$ is given by

$$B(t) - B'(t) = \sum_{j=1}^R \langle \chi_j \rangle_t \begin{bmatrix} 0 & -\nu_j \hat{s}_j^\dagger \\ -\hat{s}_j \nu_j^\dagger & \hat{\nu}_j \hat{\nu}_j^\dagger \end{bmatrix} - \sum_{j \in J_{\mathbf{x}}} \langle \chi_j \rangle_t \begin{bmatrix} 0 & 0 \\ 0 & \hat{r}_j \hat{r}_j^\dagger \end{bmatrix}.$$

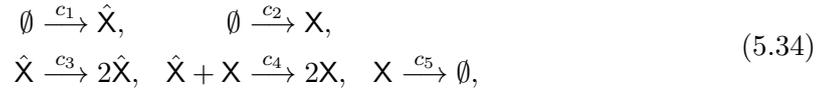
Using variation-of-constants to solve (5.33), we find that the difference between second-order moments of exact and approximate process is given by

$$S(t) - S'(t) = \int_0^t e^{(t-\tau)A} (B(\tau) - B'(\tau)) e^{(t-\tau)A^\dagger} d\tau,$$

where we assumed $S(0) = S'(0)$. In particular, if $B(t) - B'(t)$ is, say, positive semi-definite for all $t \geq 0$, this will also hold for $S(t) - S'(t)$.

Even when the reaction network is not linear, the macroscopic rates \hat{g}_j coincide with the transition rates \hat{f}_j to leading order in the system size Ω when expressed in terms of concentrations. One might then expect that in the large system size limit, the mean of

the Poisson-marginal process will coincide with the mean of the exact marginal process. We now investigate these findings numerically on the Lotka-Volterra system



a simple model of predator-prey interaction with oscillatory dynamics. Here we take the prey species \hat{X} to be part of the environment, while the predator species X constitutes the subnet. Numerical results for various system sizes are shown in Figure 5.5. We see the expected behavior: With increasing system size, the mean of the Poisson-marginal process approaches the exact mean. We also see that the Poisson-marginal process underestimates the variance of the exact marginal process.

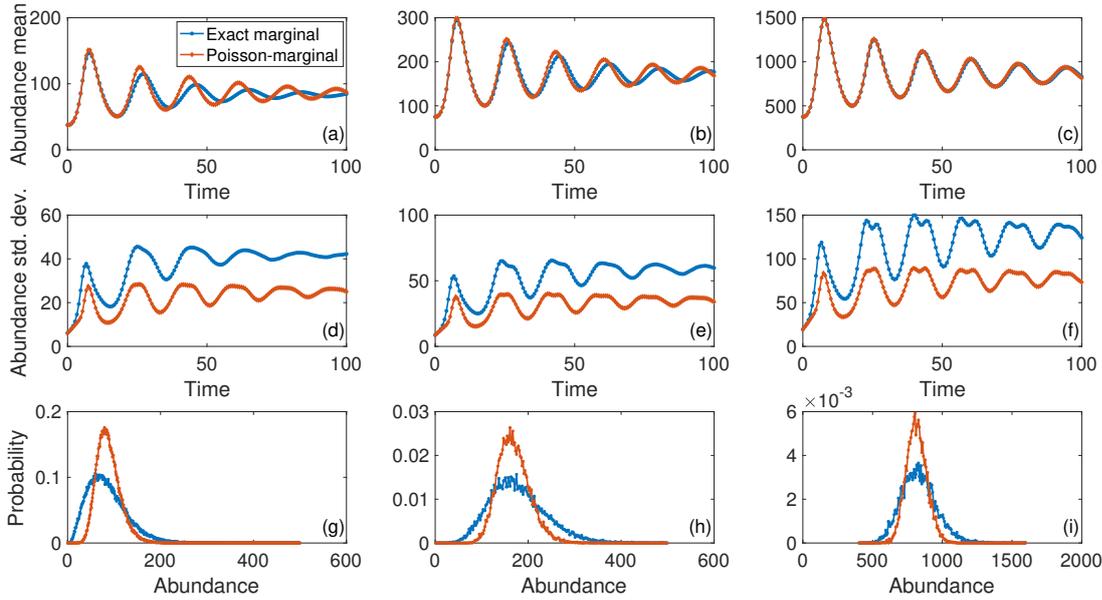


Figure 5.5: Monte Carlo evaluation of the accuracy of the Poisson-marginal process for the Lotka-Volterra system at system size $\Omega = 0.5$ in (a,d,g), $\Omega = 1$ in (b,e,h) and $\Omega = 5$ in (c,f,i). The mean of the Poisson-marginal process approaches the true mean as the system size increases. In all three cases, the Poisson-marginal process has a smaller standard deviation than the exact process. (g–i) show the full distributions at time $t = 100$. Parameters were $c_1 = 1$, $c_2 = 5$, $c_3 = 0.5$, $c_4 = 0.003$ and $c_5 = 0.3$. Initial conditions were of product-Poisson form with means $\langle X \rangle = \langle \hat{X} \rangle = 75\Omega$. The number of simulated trajectories was 100,000 for $\Omega = 0.5$, 20,000 for $\Omega = 1$ and 10,000 for $\Omega = 5$.

5.5.2 Explicit representation of marginal rates

The representation of the marginal process that we have considered in the previous sections involves auxiliary variables, either the filtering distribution itself or the filtering distribution moments. It is interesting to represent the (approximate) marginal process

in a way that explicitly shows its memory. A subclass of systems for which this is readily done for the Poisson-marginal process are processes with transition rates linear in the environment variables. Note that this does not imply that the joint reaction network (5.1) is linear. For example, the Lotka-Volterra system (5.34) satisfies this condition. This will put the marginal process framework in a form more similar to other approaches for obtaining reduced models that have recently been investigated [13, 14, 15].

Assume that (5.30) has the form

$$\frac{d}{dt}\boldsymbol{\theta} = F(\mathbf{x}(t))\boldsymbol{\theta} \quad (5.35)$$

for the appropriate matrix F , which is explicitly time-dependent through $\mathbf{x}(t)$. The extension of the following results to the case where (5.35) contains an inhomogeneity is obvious. We write

$$V(t, \tau) = \overleftarrow{\mathcal{T}} \exp \left\{ \int_{\tau}^t F(\mathbf{x}(t')) dt' \right\}$$

for the time-ordered exponential (which reduces to a product of finitely many ordinary exponentials because $\mathbf{x}(t)$ is piecewise constant). Noting from (5.31) that the increments of $\boldsymbol{\theta}$ at jumps of the marginal process are independent of $\boldsymbol{\theta}$, we can represent the solution of the filtering equation as

$$\boldsymbol{\theta}(t) = V(t, 0)\boldsymbol{\theta}(0) + \sum_{i \in J_{\mathbb{X}}} \hat{r}_i \int_0^t V(t, \tau) dy_i(\tau).$$

Note that \mathbf{Y} is a piecewise-constant process with jumps of size 1, so that the Stieltjes integral reduces to a sum. Thus, a fully explicit representation for the marginal reaction rate of reaction channel $j \in J_{\mathbb{X}}$ at time t of the marginal process is

$$\langle h_j \rangle_{\boldsymbol{\theta}} = h_j \left(\mathbf{x}(t), V(t, 0)\boldsymbol{\theta}(0) + \sum_{i \in J_{\mathbb{X}}} \hat{r}_i \int_0^t V(t, \tau) dy_i(\tau) \right).$$

We can apply this result to the Lotka-Volterra system (5.34). The equation for the Poisson mean (5.30) reads (setting $\Omega = 1$ for simplicity)

$$\frac{d}{dt}\boldsymbol{\theta} = c_1 + c_3\boldsymbol{\theta} - c_4x(t)\boldsymbol{\theta}.$$

The only reaction with a rate depending on θ is $\hat{\mathbf{X}} + \mathbf{X} \xrightarrow{c_4} 2\mathbf{X}$. Setting $v(t, \tau) = \int_{\tau}^t (c_3 - c_4x(t')) dt'$ and noting that $\hat{r}_2 = \hat{r}_4 = \hat{r}_5 = 0$, we obtain for the marginal reaction rate of this reaction the explicit representation

$$\langle h_4 \rangle_{\boldsymbol{\theta}} = c_4x(t) \left[e^{v(t, 0)}\boldsymbol{\theta}(0) + \int_0^t e^{v(t, \tau)} c_1 d\tau \right].$$

5.5.3 Limitations of the product-Poisson closure

Using a product-Poisson ansatz to close the filtering equation will not always be appropriate. The most obvious situation where this approach might fail is in the presence of

conservation relations among the environment species. This will be particularly problematic when there is no intrinsic noise in the environment. A simple example for this behavior would be the gene expression network (5.2) with input rate c_1 and decay rate c_2 for the mRNA set to zero. Irrespective of the initial distribution of mRNA at time zero, the filtering distribution will converge to a unit mass at the true mRNA abundance as the time interval over that the subnet process is observed tends to infinity. Since for a Poisson distribution the variance is equal to the mean, the vanishing of the variance of the true filtering distribution over time cannot be captured by the Poisson closure.

5.6 Auxiliary-variable master equations for the marginal process with static heterogeneity

In the previous sections of this chapter, we have discussed the marginal process for a dynamic environment fully coupled to the subnet. The filtering distribution, in that case, could not be expressed in terms of a finite number of parameters.

In this section, we consider the marginal process for static heterogeneity in the form of random rate constants, for the case of both CTMC models and SDE models, where the filtering distribution has a parametric form. We demonstrate how auxiliary-variable master equations for the marginal process, using the sufficient statistics of the parametric filtering distribution, can be derived in that case.

5.6.1 Marginal CME for static heterogeneity

We consider a general reaction network (2.3) modeled as a CTMC \mathbf{X} with static, heterogeneous rate constants \mathbf{c} , so that the reaction rates are given by

$$h_j(\mathbf{x}) = c_j \lambda_j(\mathbf{x}), \quad j = 1, \dots, R.$$

Just as for the case of a dynamic environment, the transition rates of the marginal process at time t , given the trajectory $\mathbf{x}_{[0,t]}$, are given by

$$\mathbb{E}[h_j(\mathbf{X}(t)) \mid \mathbf{x}_{[0,t]}] = \mathbb{E}[c_j \mid \mathbf{x}_{[0,t]}] \lambda_j(\mathbf{x}(t)), \quad j = 1, \dots, R.$$

As shown in [35], the conditional expectation can in this case be expressed in terms of finite-dimensional sufficient statistics (\mathbf{Y}, \mathbf{W}) , $\mathbf{Y} = (Y_1, \dots, Y_R)$, $\mathbf{W} = (W_1, \dots, W_R)$, where $Y_j(t)$ is the number of times reaction j fired in the time-interval $[0, t]$, and

$$W_j(t) := \int_0^t dt' \lambda_j(\mathbf{X}(t')).$$

For the case of Gamma-distributed heterogeneity, $c_j \sim \text{Gamma}(\alpha_j, \beta_j)$, one gets explicitly

$$\mathbb{E}[c_j \mid \mathbf{x}_{[0,t]}] = \frac{\alpha_j + y_j(t)}{\beta_j + w_j(t)},$$

with y_j and w_j being the sufficient statistics computed from the observed trajectory $\mathbf{x}_{[0,t]}$. Since (\mathbf{Y}, \mathbf{W}) are sufficient to compute the marginal transition rates, and \mathbf{Y} and \mathbf{W} themselves appear to evolve according to Markovian dynamics, one expects that

the augmented process $(\mathbf{X}(t), \mathbf{Y}(t), \mathbf{W}(t))$ is Markov. To check this, and to derive the forward equation for this process, we follow the approach in [69], which also includes results for processes on a product-form state space with both discrete and continuous components. Let us denote the discrete components of the process by $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$, so that the state space of \mathbf{Z} is \mathbb{N}_0^{N+R} . Note that \mathbf{X} can actually be computed from \mathbf{Y} , so that we could reduce the state space of the process and only keep track of \mathbf{Y} . For simplicity of notation, we do not do this in the following. Generally, the forward equation for the probability distribution $p(t, \mathbf{z}, \mathbf{w})$ takes the form

$$\frac{\partial}{\partial t} p(t, \mathbf{z}, \mathbf{w}) = \sum_{\mathbf{z}' \in \mathbb{N}_0^{N+R}} a_{\mathbf{z}\mathbf{z}'}(t, \mathbf{w}) p(t, \mathbf{z}', \mathbf{w}) + \sum_{|\mathbf{n}| > 0} \frac{(-1)^{\mathbf{n}}}{\mathbf{n}!} \frac{\partial^{|\mathbf{n}|}}{\partial \mathbf{w}^{\mathbf{n}}} [A_{\mathbf{n}, \mathbf{z}}(t, \mathbf{w}) p(t, \mathbf{z}, \mathbf{w})],$$

where we use multi-index notation for $\mathbf{n} = (n_1, \dots, n_R)$. Here the coefficients a and A are defined by

$$a_{\mathbf{z}\mathbf{z}'}(t, \mathbf{w}) = \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} [\Pr(\mathbf{Z}(t + \Delta t) = \mathbf{z} \mid \mathbf{Z}(t) = \mathbf{z}', \mathbf{W}(t) = \mathbf{w}) - \delta_{\mathbf{z}\mathbf{z}'}]$$

and

$$A_{\mathbf{n}, \mathbf{z}}(t, \mathbf{w}) = \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \mathbb{E}[(\mathbf{W}(t + \Delta t) - \mathbf{w})^{\mathbf{n}} \mid \mathbf{W}(t) = \mathbf{w}, \mathbf{Z}(t) = \mathbf{z}, \mathbf{Z}(t + \Delta t) = \mathbf{z}].$$

In order to have $\mathbf{Z}(t) = \mathbf{Z}(t + \Delta t)$, we need either no jumps or at least two jumps in $[t, t + \Delta t]$. For small Δt , the probability of more than one jump occurring in $[t, t + \Delta t]$ is of order $o(\Delta t)$ and therefore negligible. When $\mathbf{Z}(t)$ is constant over $[t, t + \Delta t]$, the term $W_j(t + \Delta t) - W_j(t)$ is deterministic and equals $\lambda_j(\mathbf{Z}(t))\Delta t$. Consequently, $A_{\mathbf{n}, \mathbf{z}}$ vanishes unless $|\mathbf{n}| = 1$, and when $n_j = 1$ equals $\lambda_j(\mathbf{Z}(t))$. For the discrete part of the equation, we obtain the usual transition probabilities of the CME, now depending on the sufficient statistic \mathbf{W} for the memory.

The complete auxiliary-variable forward equation thus takes the form

$$\begin{aligned} \frac{\partial}{\partial t} p(t, \mathbf{z}, \mathbf{w}) = & \sum_{j=1}^R \{g_j(\mathbf{z} - (\boldsymbol{\nu}_j, \mathbf{e}_j), \mathbf{w}) p(t, \mathbf{z} - (\boldsymbol{\nu}_j, \mathbf{e}_j), \mathbf{w}) - g_j(\mathbf{z}, \mathbf{w}) p(t, \mathbf{z}, \mathbf{w})\} \\ & - \sum_{j=1}^R \lambda_j(\mathbf{z}) \frac{\partial}{\partial w_j} p(t, \mathbf{z}, \mathbf{w}), \end{aligned}$$

where \mathbf{e}_j denotes the j -th standard basis vector in \mathbb{R}^R and

$$g_j(\mathbf{z}, \mathbf{w}) = \frac{\alpha_j + y_j}{\beta_j + w_j} \lambda_j(\mathbf{x}), \quad j = 1, \dots, R.$$

This master equation might, for instance, be used as a starting point for approximate treatments of the marginal process.

5.6.2 Marginal SDE for static heterogeneity

Obtaining an analogous marginal description and auxiliary-variable master equation for an SDE is complicated by technical issues. Here, we heuristically derive a marginal description for an SDE with static heterogeneous rate constants, using the Onsager-Machlup action.

We consider an SDE with state-independent diffusion term

$$dx_n = \sum_{j=1}^R c_j f_{nj}(\mathbf{x}) dt + \sum_k \sigma_{nk} dW_k, \quad n = 1, \dots, N,$$

with heterogeneous rate constants $\mathbf{c} = (c_1, \dots, c_R)$. Denote by Q the matrix with entries $Q_{mn} = \sum_k \sigma_{nk} \sigma_{mk}$ and write $S = Q^{-1}$. We first have to determine the dynamics of the marginal process, which is technically more involved than for CTMCs. Heuristically, we can argue via the infinitesimal generator of the marginal process. Thus, for a test function $\phi(\mathbf{x})$, we can compute

$$\begin{aligned} & \frac{\mathbb{E}[\phi(\mathbf{x}(t+h)) \mid \mathbf{x}_{[0,t]}] - \phi(\mathbf{x}(t))}{h} \\ &= \frac{\int d\mathbf{x}(t+h) p(\mathbf{x}(t+h) \mid \mathbf{x}_{[0,t]}) \phi(\mathbf{x}(t+h)) - \phi(\mathbf{x}(t))}{h} \\ &= \frac{\int d\mathbf{x}(t+h) \int d\mathbf{c} p(\mathbf{x}(t+h) \mid \mathbf{c}, \mathbf{x}(t)) p(\mathbf{c} \mid \mathbf{x}_{[0,t]}) \phi(\mathbf{x}(t+h)) - \phi(\mathbf{x}(t))}{h} \\ &= \frac{\int d\mathbf{c} p(\mathbf{c} \mid \mathbf{x}_{[0,t]}) \left\{ \int d\mathbf{x}(t+h) p(\mathbf{x}(t+h) \mid \mathbf{c}, \mathbf{x}(t)) \phi(\mathbf{x}(t+h)) - \phi(\mathbf{x}(t)) \right\}}{h} \\ &\xrightarrow{h \rightarrow 0} \int d\mathbf{c} p(\mathbf{c} \mid \mathbf{x}_{[0,t]}) \mathcal{L}_{\mathbf{c}}^\dagger \phi[\mathbf{x}(t)], \end{aligned}$$

provided we can interchange limits in the last equality. Here $\mathcal{L}_{\mathbf{c}}^\dagger$ is the adjoint of the Fokker-Planck operator of the original non-marginal process with parameters \mathbf{c} . Since the distribution $p(\mathbf{x}(t+h) \mid \mathbf{x}_{[0,t]})$ agrees to first order in h with $p(\mathbf{x}(t+h) \mid \mathbf{x}(t), \mathbf{c} = \mathbb{E}[\mathbf{c} \mid \mathbf{x}_{[0,t]}])$, the process can be said to satisfy the ‘‘SDE’’ (or rather, an Itô process [87])

$$dx_n = \sum_{j=1}^R \mathbb{E}[c_j \mid \mathbf{x}_{[0,t]}] f_{nj}(\mathbf{x}) dt + \sum_k \sigma_{nk} dW_k, \quad n = 1, \dots, N, \quad (5.36)$$

where the increment dx_n is to be understood conditional on $\mathbf{x}_{[0,t]}$, which is suppressed in the notation.

Just as for the case of an MJP, we have to compute the filtering distribution $p(\mathbf{c} \mid \mathbf{x}_{[0,t]})$. Because the heterogeneity is assumed to be static, we can directly compute the filtering distribution using Bayes’ theorem, without deriving a differential equation for it. We need the path density $p(\mathbf{x}_{[0,t]} \mid \mathbf{c})$. This can be written as (see e.g. [88], and note the

remarks at the end of this section)

$$\begin{aligned}
 p(\mathbf{x}_{[0,t]} | \mathbf{c}) &\propto \exp \left\{ -\frac{1}{2} \sum_{n,m=1}^N S_{nm} \int_0^t dt' (\dot{x}_n - \sum_{i=1}^R c_i f_{ni}(\mathbf{x})) (\dot{x}_m - \sum_{j=1}^R c_j f_{mj}(\mathbf{x})) \right\} \\
 &= \exp \left\{ -\frac{1}{2} \sum_{n,m=1}^N S_{nm} \int_0^t dt' \dot{x}_n \dot{x}_m + \sum_{j=1}^R y_j c_j - \frac{1}{2} \sum_{i,j=1}^R z_{ij} c_i c_j \right\}
 \end{aligned} \tag{5.37}$$

with

$$\begin{aligned}
 y_j &= \sum_{n,m=1}^N S_{nm} \int_0^t dt' \dot{x}_n f_{mj}(\mathbf{x}) = \sum_{n,m=1}^N S_{nm} \int_0^t f_{mj}(\mathbf{x}) dx_n, & j = 1, \dots, R, \\
 z_{ij} &= \sum_{n,m=1}^N S_{nm} \int_0^t dt' f_{ni}(\mathbf{x}) f_{mj}(\mathbf{x}), & i, j = 1, \dots, R,
 \end{aligned}$$

or, in differential notation,

$$dy_j = \sum_{n,m=1}^N S_{nm} f_{mj}(\mathbf{x}) dx_n, \quad j = 1, \dots, R, \tag{5.38}$$

$$dz_{ij} = \sum_{n,m=1}^N S_{nm} f_{ni}(\mathbf{x}) f_{mj}(\mathbf{x}) dt, \quad i, j = 1, \dots, R. \tag{5.39}$$

Just as in the previous section, we can introduce the sufficient statistics $\mathbf{y} = (y_j)$ and $\mathbf{z} = (z_{ij})$. Then for the distribution of \mathbf{c} conditional on a path $\mathbf{x}_{[0,t]}$, we have $p(\mathbf{c} | \mathbf{x}_{[0,t]}) = p(\mathbf{c} | \mathbf{y}_t, \mathbf{z}_t)$. Denote by $\mathbf{v}(\mathbf{y}, \mathbf{z}) = (v_1(\mathbf{y}, \mathbf{z}), \dots, v_R(\mathbf{y}, \mathbf{z}))$ the corresponding conditional expectation, i.e. $\mathbf{v}(\mathbf{y}, \mathbf{z}) = \mathbb{E}[\mathbf{c} | \mathbf{y}, \mathbf{z}]$. For example, if we assume Gaussian heterogeneity $\mathbf{c} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, we obtain for the estimator

$$\mathbf{v}(\mathbf{y}, \mathbf{z}) = (\mathbf{z} + \Sigma^{-1})^{-1} (\mathbf{y} + \Sigma^{-1} \boldsymbol{\mu}).$$

Continuing with the marginal SDE, if we now write $\mathbb{E}[c_j | \mathbf{x}_{[0,t]}]$ in (5.36) in terms of the variables \mathbf{y}, \mathbf{z} , and also plug in (5.36) into (5.38), we obtain a system of SDEs

$$\begin{aligned}
 dx_n &= \sum_{j=1}^R v_j(\mathbf{y}, \mathbf{z}) f_{nj}(\mathbf{x}) dt + \sum_k \sigma_{nk} dW_k, \\
 dy_j &= \sum_{n,m} S_{nm} f_{mj}(\mathbf{x}) dx_n \\
 &= \sum_{n,m} S_{nm} f_{nj}(\mathbf{x}) \sum_i v_i(\mathbf{y}, \mathbf{z}) f_{mi}(\mathbf{x}) dt + \sum_{n,m} S_{nm} f_{mj}(\mathbf{x}) \sum_k \sigma_{nk} dW_k, \\
 dz_{ij} &= \sum_{n,m} S_{nm} f_{ni}(\mathbf{x}) f_{mj}(\mathbf{x}) dt,
 \end{aligned}$$

with a corresponding Fokker-Planck equation for the density $\pi = \pi(\mathbf{x}, \mathbf{y}, \mathbf{z})$,

$$\begin{aligned} \frac{\partial}{\partial t} \pi = & - \sum_n \frac{\partial}{\partial x_n} \sum_j v_j(\mathbf{y}, \mathbf{z}) f_{nj}(\mathbf{x}) \pi - \sum_j \frac{\partial}{\partial y_j} \sum_{n,m} S_{nm} f_{nj}(\mathbf{x}) \sum_i v_i(\mathbf{y}, \mathbf{z}) f_{mi}(\mathbf{x}) \pi \\ & - \sum_{i,j} \frac{\partial}{\partial z_{ij}} \sum_{n,m} S_{nm} f_{ni}(\mathbf{x}) f_{mj}(\mathbf{x}) \pi + \frac{1}{2} \sum_{n,m} \frac{\partial}{\partial x_n \partial x_m} Q_{nm} \pi + \frac{1}{2} \sum_n \sum_j \frac{\partial}{\partial x_n \partial y_j} f_{nj}(\mathbf{x}) \pi \\ & + \frac{1}{2} \sum_{i,j} \frac{\partial}{\partial y_i \partial y_j} \sum_{n,m} S_{nm} f_{ni}(\mathbf{x}) f_{mj}(\mathbf{x}) \pi. \end{aligned}$$

Note that infinite-dimensional approaches to path-dependent (backward) Kolmogorov equations also exist [89].

The derivations above were all based on the Onsager-Machlup expression for the path density (5.37). That expression is defined as a limit of the process time-discretized into M steps with step-size Δt , which (for simplicity, written for the 1-dimensional case, and setting $\Delta x_n = x_n - x_{n-1}$) is given by

$$\begin{aligned} p(x_{0:M} | \mathbf{c}) &= (2\pi\Delta t)^{-M/2} \exp \left\{ -\frac{1}{2\Delta t} \sum_{n=1}^M \left[x_n - \left(x_{n-1} + \Delta t \sum_{j=1}^R c_j f_j(x_{n-1}) \right) \right]^2 \right\} \\ &= (2\pi\Delta t)^{-M/2} \exp \left\{ -\frac{1}{2\Delta t} \sum_{n=1}^M (\Delta x_n)^2 \right\} \\ &\quad \times \exp \left\{ \sum_{j=1}^R c_j \sum_{n=1}^M f_j(x_{n-1}) \Delta x_n - \frac{1}{2} \sum_{i,j=1}^R c_i c_j \Delta t \sum_{n=1}^M f_i(x_{n-1}) f_j(x_{n-1}) \right\}, \end{aligned}$$

and it is not immediately clear whether the expression for the filtering distribution $p(\mathbf{c} | \mathbf{x}_{[0,t]})$ is meaningful. The reason is the presence of the divergent term

$$(2\pi\Delta t)^{-M/2} \exp \left\{ -\frac{1}{2\Delta t} \sum_{n=1}^M (\Delta x_n)^2 \right\}.$$

However, somewhat heuristically, we can argue as follows: The distribution $p(\mathbf{c} | x_{[0,t]})$ is well defined as

$$p(\mathbf{c} | x_{[0,t]}) = \lim_{M \rightarrow \infty} p(\mathbf{c} | x_{0:M}).$$

This is because, as is evident from the computation

$$\begin{aligned} p(\mathbf{c} | x_{0:M}) &= \frac{p(x_{0:M} | \mathbf{c}) p(\mathbf{c})}{\int d\mathbf{c} p(x_{0:M} | \mathbf{c}) p(\mathbf{c})} \\ &= \frac{\exp \left\{ \sum_j c_j \sum_{n=1}^M f_j(x_{n-1}) \Delta x_n - \frac{1}{2} \sum_{i,j} c_i c_j \Delta t \sum_{n=1}^M f_i(x_{n-1}) f_j(x_{n-1}) \right\} p(\mathbf{c})}{\int d\mathbf{c} \exp \left\{ \sum_j c_j \sum_{n=1}^M f_j(x_{n-1}) \Delta x_n - \frac{1}{2} \sum_{i,j} c_i c_j \Delta t \sum_{n=1}^M f_i(x_{n-1}) f_j(x_{n-1}) \right\} p(\mathbf{c})}, \end{aligned}$$

the divergent terms cancel when Bayes' theorem is applied, and now the limit $M \rightarrow \infty$ can be taken without difficulty.

5.7 Filtering equation and entropic matching as projection operations

In Chapter 4, we applied the projection operator formalism to heterogeneous reaction kinetics to obtain a marginal description of the process of interest. The process descriptions in that chapter were on the level of time-marginal distributions. The marginal, convolutional master equation (4.8) was expressed in terms of an integral over time-marginal distributions up to some time t . Those time-marginal distributions were propagated forward in time by the “orthogonal” dynamics, which, in the notation of Section 4.1, are given by the operator $\mathcal{QL} = (1 - \mathcal{P})\mathcal{L}$.

In this section, we discuss how the marginal process framework can be seen to be an analogous approach, which however operates on the process level. To do this, we discuss how the continuous part of the filtering equation (5.12) and the entropic matching equation (5.28) arise as an application of an orthogonal projection (using the Fisher-Rao information metric) applied to the vector field defined by the joint master equation.

Since the applicability of the results in this chapter does not actually depend on any of the results in this section, we restrict the discussion to a form that stresses the geometrical significance and neglects any technical difficulties. See [90] for a general treatment of information geometry.

For simplicity, assume that $\mathbb{X} \times \hat{\mathbb{X}}$ is finite, and define the set of probability distributions on $\mathbb{X} \times \hat{\mathbb{X}}$,

$$\mathbb{P} = \{p : \mathbb{X} \times \hat{\mathbb{X}} \rightarrow [0, 1] \mid \mathcal{S}p = 1\},$$

which inherits a manifold structure as a subset of finite-dimensional Euclidean space. The tangent space at a point $p \in \mathbb{P}$ is given by

$$T_p\mathbb{P} = \{p\} \times \{v : \mathbb{X} \times \hat{\mathbb{X}} \rightarrow \mathbb{R} \mid \mathcal{S}v = 0\}.$$

For an MJP (X, \hat{X}) on $\mathbb{X} \times \hat{\mathbb{X}}$, the master equation $dp_t(x, \hat{x})/dt = [\mathcal{L}p_t](x, \hat{x})$ defines a vector field on \mathbb{P} , the vector attached at a point p being $\mathcal{L}p$. We define a basepoint-dependent metric by

$$g_p(v, w) = \sum_{\substack{x, \hat{x} \\ p(x, \hat{x}) \neq 0}} \frac{v(x, \hat{x})w(x, \hat{x})}{p(x, \hat{x})}.$$

for $v, w \in T_p\mathbb{P}$. When restricted to $p \in \mathbb{P}$ with $p > 0$ everywhere, this is the information metric. Our extension to other p is somewhat ad hoc, but sufficient for our purposes.

It turns out that the continuous part of the filtering equation when X is in state $x(t)$ is obtained simply as an orthogonal projection of the vector field of the full master equation on the tangent space to the submanifold

$$\mathbb{P}_{x(t)} = \{p \in \mathbb{P} \mid p(x, \hat{x}) = \delta_{x(t), x} \hat{p}(\hat{x}) \text{ for some } \hat{p}(\hat{x})\}.$$

From now on, let $p \in \mathbb{P}_{x(t)}$ with $p(x, \hat{x}) = \delta_{x(t), x} \hat{p}(\hat{x})$ and assume $\hat{p} > 0$ everywhere. One easily checks that the linear operator

$$\mathcal{F}_p : T_p\mathbb{P} \rightarrow T_p\mathbb{P}, \quad \mathcal{F}_p v = (\mathcal{P}_{x(t)} - p\mathcal{S}\mathcal{P}_{x(t)})v,$$

satisfies $\mathcal{F}_p^2 = \mathcal{F}_p$ and $g_p(\mathcal{F}_p v, w) = g_p(v, \mathcal{F}_p w)$ for all $v, w \in T_p \mathbb{P}$, so that \mathcal{F}_p is an orthogonal projection. The projected vector field $p \mapsto (p, \mathcal{F}_p \mathcal{L}p)$ corresponds to the filtering equation.

The entropic matching equations can similarly be derived as an application of a further projection. They were in fact first derived in [51], using such a geometric approach. Considering a K -dimensional parametric family $p_{\boldsymbol{\theta}}$ of probability distributions on $\hat{\mathbb{X}}$, the map $\boldsymbol{\theta} \mapsto \delta_{x(t), x} p_{\boldsymbol{\theta}}(\hat{x})$ defines a submanifold of $\mathbb{P}_{x(t)}$ which we denote by $\mathbb{P}'_{x(t)}$. From now on, let $p \in \mathbb{P}'_{x(t)}$ with $p(x, \hat{x}) = \delta_{x(t), x} p_{\boldsymbol{\theta}}(\hat{x})$. The tangent space $T_p \mathbb{P}'_{x(t)}$ to this submanifold is spanned by the vectors

$$v_k = v_k(x, \hat{x}) = \delta_{x(t), x} \frac{\partial p_{\boldsymbol{\theta}}(\hat{x})}{\partial \theta_k}, \quad k = 1, \dots, K.$$

We then find for the Gram matrix

$$g_p(v_k, v_l) = G_{kl}(\boldsymbol{\theta}), \quad k, l = 1, \dots, K,$$

i.e. the information metric as defined by (5.27). The orthogonal projection $\mathcal{Q}_p : T_p \mathbb{P} \rightarrow T_p \mathbb{P}'_{x(t)}$ onto the tangent space $T_p \mathbb{P}'_{x(t)}$ is given by

$$\mathcal{Q}_p w = \sum_{k,l=1}^K g_p(w, v_k) [G(\boldsymbol{\theta})^{-1}]_{kl} v_l.$$

Because $T_p \mathbb{P}'_{x(t)}$ is a subspace of $T_p \mathbb{P}_{x(t)}$, we have $\mathcal{Q}_p \mathcal{F}_p = \mathcal{Q}_p$. The resulting projected ME, defined by the vector field

$$p \mapsto (p, \mathcal{Q}_p \mathcal{F}_p \mathcal{L}p) = (p, \mathcal{Q}_p \mathcal{L}p),$$

evolves on the manifold $\mathbb{P}'_{x(t)}$ when it is started there. When this is written in terms of the variables $\boldsymbol{\theta}$, we obtain the equations of entropic matching (5.28). Here we also see that entropic matching can be used to directly obtain a finite-dimensional approximation to the filtering equation from the master equation, without deriving the filtering equation in an intermediate step. The derivation presented in Section 5.4.2 could also be adapted in this way and would then be an application of variational inference [91].

We see that both the exact filtering equation and the approximated filtering equation after the application of entropic matching arise as orthogonal projections of the joint master equation.

Chapter 6

Summary and Outlook

The mathematical treatment of biomolecular reaction networks, taking into account both intrinsic and extrinsic noise, remains a challenging problem. In this thesis, two types of approximate treatments for biomolecular kinetics have been investigated: (i) approximations for the time-marginal probability density and (ii) approximate marginalizations of heterogeneous kinetics or of dynamic environments.

In Chapter 3, a variational approach to moment closure approximations was introduced and used to explain and partially correct problems usually attributed to moment closure, in particular the divergence of the approximation at low copy numbers. Mixtures of product-Poisson distributions were used to obtain a flexible class of probability distributions that can be used for moment closure for the CME. The variational approach results in full approximate solutions of the underlying evolution equation and was used to obtain approximate marginal descriptions of heterogeneous reaction kinetics with log-normally distributed rate constants. An extension to the approximation of multi-time probability distributions was also derived. It was shown that the multi-time approximations are consistent with approximations of the single-time marginal distributions. Therefore, variational moment closure can be used as a viable replacement for process-level approximations. Finally, a diagrammatic technique for the derivation of moment equations and cumulant equations was developed. The diagrammatic rules are very simple and allow one to readily understand the structure of cumulant equations. As an application, they were used to understand the relation between the cumulant equations for the CME and for the Kramers-Moyal expansion.

In Chapter 4, the application of the projection operator formalism for the treatment of heterogeneous reaction kinetics was discussed, working on the level of time-marginal probability distributions via Kolmogorov-forward equations. As a first step, treating the marginalization of the environment on time-marginal distributions as the projection operation, the relation between the formalism as applied to the CME and as applied on the level of the RRE was investigated. It was found that, while the resulting mean equations are consistent between CME and RRE, the Markovian, memory and noise terms of the projection operator formalism do not map to each other directly. In a second step, the focus was shifted to a projection operator marginalizing full trajectories of a stationary environment. Assuming CIR-distributed (or Gamma-distributed) heterogeneity, two exactly solvable cases were considered. In particular, a heterogeneous decay rate for the birth-death process and the case of commuting operators for an evolution equation were discussed. Since the majority of cases are not solvable in closed form, cumulant expansions in terms of partial cumulants were used to derive approximate marginal descriptions of a subnetwork embedded in a stationary environment when a feed-forward

structure is assumed. The approximation was investigated numerically and analytically for some simple linear example systems with CIR-distributed heterogeneity. The approximation was found to perform well as long as the variance of the heterogeneity was not too large.

In Chapter 5, the exact process-level marginal description of a fully coupled MJP was derived. Using entropic matching, a principled approximate description of the marginal process was obtained. The resulting approximation was interpreted in terms of an orthogonal projection of the full joint master equation. Using product-Poisson entropic matching, a particularly simple approximate description of the marginal process for mass-action kinetics was obtained. The properties of that process were investigated both analytically and numerically. It was found that for linear networks, the marginal process mean is reproduced exactly by the Poisson-marginal process. Bounds for the difference between the covariance matrices of exact and approximate marginal process were also derived. Restricting considerations to the case of static heterogeneous rate constants, auxiliary-variable master equations for the marginal process were derived. These results were also extended to systems modeled by SDEs, where results were based on the Onsager-Machlup path density.

Outlook

Using the results of this thesis, a number of ideas for future investigations suggest themselves.

The variational approach to moment closure of Chapter 3 appears to be very suitable for the development of inference algorithms. One reason for this is the fact that it yields a full approximating distribution, and not merely a number of moments. The other reason is the possibility to use the approach to obtain multi-time joint distributions. Variational moment closure is also a natural framework for combining different existing approximation and model reduction methods. This is important because for real-world networks with many species and reactions, a variety of properties such as heterogeneity, time-scale separation or abundance separation will make the simultaneous use of several types of approximation methods both necessary and desirable. The diagrammatic technique developed might allow one to compare moment closure approximations to the system size expansion, for which a diagrammatic technique using Feynman diagrams exists [92].

The use of the projection operator formalism, particularly via cumulant expansions, was here investigated for (conditionally) linear systems. Future work might focus on the investigation of the approximations on non-linear systems, where a further approximation (say, via moment closure) is necessary to obtain tractable equations. Another interesting problem is to find further exactly marginalizable cases of heterogeneous reaction kinetics.

The marginal process framework was considered for MJPs only. Analogous results for SDEs might allow a comparison to other methods used for obtaining marginal process equations [13, 14, 15]. Since the environment of a system will often not be known exactly, the memory effects obtained from the marginal process framework could be used to investigate the properties of environments in general. For instance, one might consider large, random environments in the limit of infinite environment size [14].

Appendix A

Differential equations for the solution of the cumulant expansions

In this appendix, we state the auxiliary-variable differential equations that can be used to numerically solve the integro-differential equations (4.19) in Section 4.2.2. We assume that the correlation matrix $C(t, t') = C(t-t') = [C_{jk}(t-t')]_{jk}$ satisfies a linear differential equation

$$\frac{\partial}{\partial t} C_{jk}(t-t') = - \sum_i A_{ji} C_{ik}(t-t'),$$

as it will for many environments of practical interest, such as for the CIR process.

Now consider, for instance, the cumulant expansion to second order, given by the first two terms on the right-hand side of (4.19). We introduce the quantities

$$q_{jk}(t, \mathbf{x}) := \int_0^t dt' C_{jk}(t-t') e^{(t-t')\bar{\mathcal{L}}} \mathcal{L}_k p(t', \mathbf{x}), \quad j, k = 1, \dots, R_0.$$

These satisfy the differential equations

$$\frac{d}{dt} q_{jk}(t, \mathbf{x}) = C_{jk}(0) \mathcal{L}_k p(t, \mathbf{x}) + \bar{\mathcal{L}} q_{jk}(t, \mathbf{x}) - \sum_i A_{ji} q_{ik}(t, \mathbf{x}), \quad j, k = 1, \dots, R_0.$$

Together with the equation for $p(t, \mathbf{x})$ written using the new variables,

$$\frac{d}{dt} p(t, \mathbf{x}) = \bar{\mathcal{L}} p(t, \mathbf{x}) + \sum_{j,k=1}^{R_0} \mathcal{L}_j q_{jk}(t, \mathbf{x}),$$

we have obtained a closed system.

Generally, this will be a complicated system of either high-dimensional ODEs or PDEs, and not amenable to direct solution. We can, however, extract moment equations in the usual way. For a function $\phi(\mathbf{x})$, these are

$$\begin{aligned} \frac{d}{dt} \langle \phi \rangle &= \langle \bar{\mathcal{L}}^\dagger \phi \rangle + \sum_{j,k=1}^R \langle \mathcal{L}_j^\dagger \phi \rangle_{jk} \\ \frac{d}{dt} \langle \phi \rangle_{jk} &= C_{jk}(0) \langle \mathcal{L}_k^\dagger \phi \rangle + \langle \bar{\mathcal{L}}^\dagger \phi \rangle_{jk} - \sum_i A_{ji} \langle \phi \rangle_{ik}, \quad j = 1, \dots, R, \end{aligned}$$

where $\langle \cdot \rangle_{jk}$ denotes an ‘‘expectation’’ with respect to q_{jk} . Note that the moment equations, just as for the standard CME or CFPE, will in general not be closed, so that in addition to the approximation introduced by the cumulant expansion, some form of moment closure will be required.

Appendix B

The marginal simulation algorithm

Here we describe one possible way to simulate the (approximate) marginal process for reaction networks. Let

$$\frac{d}{dt}\boldsymbol{\theta} = v(\boldsymbol{\theta}, x) \quad (\text{B.1})$$

be the differential equation governing the parameters $\boldsymbol{\theta}$ of the (approximate or exact) solution of the filtering equation in between jumps, which in general will depend on the marginal process state x . For example, for the Poisson-marginal process, v is given by the right-hand side of (5.30). Similarly, let

$$\boldsymbol{\theta}_+ = v_j(\boldsymbol{\theta}_-, x_-) \quad (\text{B.2})$$

be the equation specifying the update to the parameters $\boldsymbol{\theta}$ when the subnet jumps via reaction j . For example, for the Poisson-marginal process, v_j is given by the right-hand side of (5.31). An algorithm [93] based on the modified next reaction method [5] can be formulated as follows: The expected reaction rates $\langle h_j \rangle_{\boldsymbol{\theta}}$, $j \in J_{\mathbb{X}}$ of those reactions that modify the state of \mathbf{Y} are functions of $\boldsymbol{\theta}$. We augment the ODE system (B.1) to include new variables

$$\frac{d}{dt}\tau_j = -\langle h_j \rangle_{\boldsymbol{\theta}}, \quad j \in J_{\mathbb{X}}. \quad (\text{B.3})$$

Algorithm 2 Marginal stochastic simulation algorithm
(modified next reaction method)

```
Set  $t \leftarrow 0, \mathbf{x} \leftarrow \mathbf{x}_0, \boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_0$ . ▷ Initialization
for  $j \in J_{\mathbb{X}}$  do
  Sample  $u \sim \text{Uniform}(0, 1)$ .
  Set  $\tau_j \leftarrow -\ln u$ .
end for

while  $t < T$  do ▷ Main loop
  Solve (B.1), (B.3) until the first variable  $\tau_{j^*}$  reaches 0 for
  some index  $j^* \in J_{\mathbb{X}}$ .
  Update  $\boldsymbol{\theta} \leftarrow v_{j^*}(\boldsymbol{\theta}, x)$ .
  Update  $\mathbf{x} \leftarrow \mathbf{x} + \boldsymbol{\nu}_{j^*}$ .
  Sample  $u \sim \text{Uniform}(0, 1)$ .
  Set  $\tau_{j^*} \leftarrow -\ln u$ .
end while
```

The system can then be simulated using Algorithm 2, which samples a trajectory of the (approximate) marginal process over the time-interval $[0, T]$ starting from an initial subnet state \boldsymbol{x}_0 and initial parameters $\boldsymbol{\theta}_0$ for the filtering distribution at time 0. The algorithm has to find the time-point at which a function of the ODE system state crosses a specified threshold (one of the variables τ_j reaches 0). This is a functionality provided by many ODE solvers, so that the algorithm is straightforward to implement.

Appendix C

List of Acronyms

CFPE	Chemical Fokker-Planck equation
CIR	Cox-Ingersoll-Ross
CLE	Chemical Langevin equation
CME	Chemical Master equation
CTMC	Continuous-time Markov chain
EM	Entropic matching
FPE	Fokker-Planck equation
KME	Kramers-Moyal expansion
MC	Monte Carlo
ME	Master equation
MJP	Markov jump process
ODE	Ordinary differential equation
PDE	Partial differential equation
RRE	Reaction rate equation(s)
SDE	Stochastic differential equation
SSA	Stochastic simulation algorithm
TASEP	Totally asymmetric simple exclusion process
ZC	Zero cumulant
ZI	Zero information

Bibliography

- [1] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.
- [2] Arjun Raj and Alexander van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226, 2008.
- [3] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [4] Jakob Ruess, Heinz Koepl, and Christoph Zechner. Sensitivity estimation for stochastic models of biochemical reaction networks in the presence of extrinsic variability. *The Journal of Chemical Physics*, 146(12):124122, 2017.
- [5] David F Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *The Journal of Chemical Physics*, 127(21):214107, 2007.
- [6] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.
- [7] David F Anderson and Desmond J Higham. Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics. *Multiscale Modeling & Simulation*, 10(1):146–179, 2012.
- [8] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, 124(4):044104, 2006.
- [9] NG van Kampen. A power series expansion of the master equation. *Canadian Journal of Physics*, 39(4):551–567, 1961.
- [10] P Whittle. On the use of the normal approximation in the treatment of stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 268–281, 1957.
- [11] Stefan Engblom. Computing the moments of high dimensional solutions of the master equation. *Applied Mathematics and Computation*, 180(2):498–515, 2006.
- [12] Christoph Zechner and Heinz Koepl. Uncoupled analysis of stochastic reaction networks in fluctuating environments. *PLoS Computational Biology*, 10(12):e1003942, 2014.

- [13] Katy J Rubin, Katherine Lawler, Peter Sollich, and Tony Ng. Memory effects in biochemical networks as the natural counterpart of extrinsic noise. *Journal of Theoretical Biology*, 357:245–267, 2014.
- [14] Barbara Bravi, Peter Sollich, and Manfred Opper. Extended Plefka expansion for stochastic dynamics. *Journal of Physics A: Mathematical and Theoretical*, 49(19):194003, 2016.
- [15] B Bravi and P Sollich. Statistical physics approaches to subnetwork dynamics in biochemical systems. *Physical Biology*, 14(4):045010, 2017.
- [16] Yang Cao, Daniel T Gillespie, and Linda R Petzold. The slow-scale stochastic simulation algorithm. *The Journal of Chemical Physics*, 122(1):014116, 2005.
- [17] Andreas Hellander and Per Lötstedt. Hybrid method for the chemical master equation. *Journal of Computational Physics*, 227(1):100–122, 2007.
- [18] Tobias Jahnke. On reduced models for the chemical master equation. *Multiscale Modeling & Simulation*, 9(4):1646–1676, 2011.
- [19] Philipp Thomas, Ramon Grima, and Arthur V Straube. Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators. *Physical Review E*, 86(4):041110, 2012.
- [20] George WA Constable, Alan J McKane, and Tim Rogers. Stochastic dynamics on slow manifolds. *Journal of Physics A: Mathematical and Theoretical*, 46(29):295002, 2013.
- [21] Hye-Won Kang and Thomas G Kurtz. Separation of time-scales and model reduction for stochastic reaction networks. *The Annals of Applied Probability*, 23(2):529–583, 2013.
- [22] Jae Kyoung Kim, Grzegorz A Rempala, and Hye-Won Kang. Reduction for stochastic biochemical reaction networks with multiscale conservations. *Multiscale Modeling & Simulation*, 15(4):1376–1403, 2017.
- [23] L Bronstein, C Zechner, and H Koepl. Bayesian inference of reaction kinetics from single-cell recordings across a heterogeneous cell population. *Methods*, 85:22–35, 2015.
- [24] Leo Bronstein and Heinz Koepl. A variational approach to moment-closure approximations for the kinetics of biomolecular reaction networks. *The Journal of Chemical Physics*, 148(1):014105, 2018.
- [25] Leo Bronstein and Heinz Koepl. A diagram technique for cumulant equations in biomolecular reaction networks with mass-action kinetics. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 5857–5862. IEEE, 2016.
- [26] Johannes Falk, Leo Bronstein, Maleen Hanst, Barbara Drossel, and Heinz Koepl. Context in synthetic biology: Memory effects of environments with mono-molecular reactions. *The Journal of Chemical Physics*, 150(2):024106, 2019.

-
- [27] Leo Bronstein and Heinz Koepl. Marginal process framework: A model reduction tool for Markov jump processes. *Physical Review E*, 97(6):062147, 2018.
- [28] Christopher Schneider, Leo Bronstein, Jascha Diemer, Heinz Koepl, and Beatrix Suess. ROC'n'Ribo: Characterizing a riboswitching expression system by modeling single-cell data. *ACS Synthetic Biology*, 6(7):1211–1224, 2017.
- [29] Leo Bronstein and Heinz Koepl. Scalable inference using PMCMC and parallel tempering for high-throughput measurements of biomolecular reaction networks. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 770–775. IEEE, 2016.
- [30] Ron Milo and Rob Phillips. *Cell biology by the numbers*. Garland Science, 2015.
- [31] Darren J Wilkinson. *Stochastic Modelling for Systems Biology*. CRC Press, 2011.
- [32] Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58:35–55, 2007.
- [33] Daniel T Gillespie. The chemical Langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.
- [34] Wilson W Wong, Tony Y Tsai, and James C Liao. Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Molecular Systems Biology*, 3(1):130, 2007.
- [35] Christoph Zechner, Michael Unger, Serge Pelet, Matthias Peter, and Heinz Koepl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods*, 11(2):197, 2014.
- [36] John C Cox and E Ingersoll. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.
- [37] NG van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 2007.
- [38] Christian Kuehn. Moment closure — a brief review. In *Control of self-organizing nonlinear systems*, pages 253–271. Springer, 2016.
- [39] David Schnoerr, Guido Sanguinetti, and Ramon Grima. Validity conditions for moment closure approximations in stochastic chemical kinetics. *The Journal of Chemical Physics*, 141(8):08B616_1, 2014.
- [40] David Schnoerr, Guido Sanguinetti, and Ramon Grima. Comparison of different moment-closure approximations for stochastic chemical kinetics. *The Journal of Chemical Physics*, 143(18):11B610_1, 2015.
- [41] Arnab Ganguly, Derya Altintan, and Heinz Koepl. Jump-diffusion approximation of stochastic reaction dynamics: error bounds and algorithms. *SIAM Multiscale Modeling & Simulation*, 13(4):1390–1419, 2015.

- [42] Claudia Cianci, David Schnoerr, Andreas Pehler, and Ramon Grima. An alternative route to the system-size expansion. *Journal of Physics A: Mathematical and Theoretical*, 50(39):395003, 2017.
- [43] Abhyudai Singh and Joao Pedro Hespanha. A derivative matching approach to moment closure for the stochastic logistic model. *Bulletin of Mathematical Biology*, 69(6):1909–1925, 2007.
- [44] Eszter Lakatos, Angelique Ale, Paul DW Kirk, and Michael PH Stumpf. Multivariate moment closure techniques for stochastic kinetic models. *The Journal of Chemical Physics*, 143(9):094107, 2015.
- [45] Khem Raj Ghusinga, Cesar A Vargas-Garcia, Andrew Lamperski, and Abhyudai Singh. Exact lower and upper bounds on stationary moments in stochastic biochemical systems. *Physical Biology*, 14(4):04LT01, 2017.
- [46] Juan Kuntz, Philipp Thomas, Guy-Bart Stan, and Mauricio Barahona. Rigorous bounds on the stationary distributions of the chemical master equation via mathematical programming. *arXiv preprint arXiv:1702.05468*, 2017.
- [47] Ramon Grima. A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *The Journal of Chemical Physics*, 136(15):154105, 2012.
- [48] Tiago Ramalho, Marco Selig, Ulrich Gerland, and Torsten A Enßlin. Simulation of stochastic network dynamics via entropic matching. *Physical Review E*, 87(2):022719, 2013.
- [49] Gregory L Eyink. Action principle in nonequilibrium statistical dynamics. *Physical Review E*, 54(4):3419, 1996.
- [50] Reimar H Leike and Torsten A Enßlin. Optimal belief approximation. *Entropy*, 19(8):402, 2017.
- [51] Damiano Brigo, Bernard Hanzon, and François Le Gland. Approximate nonlinear filtering by projection on exponential manifolds of densities. *Bernoulli*, 5(3):495–534, 1999.
- [52] Damiano Brigo and Giovanni Pistone. Dimensionality reduction for measure valued evolution equations in statistical manifolds. In *Computational Information Geometry*, pages 217–265. Springer, 2017.
- [53] Cedric Archambeau, Dan Cornford, Manfred Opper, and John Shawe-Taylor. Gaussian process approximations of stochastic differential equations. In *Gaussian Processes in Practice*, pages 1–16, 2007.
- [54] Manfred Opper and Guido Sanguinetti. Variational inference for Markov jump processes. In *Advances in Neural Information Processing Systems*, pages 1105–1112, 2008.

-
- [55] Tobias Sutter, Arnab Ganguly, and Heinz Koepl. A variational approach to path estimation and parameter inference of hidden diffusion processes. *Journal of Machine Learning Research*, 17(190):1–37, 2016.
- [56] J Trębicki and K Sobczyk. Maximum entropy principle and non-stationary distributions of stochastic systems. *Probabilistic Engineering Mechanics*, 11(3):169–178, 1996.
- [57] Patrick Smadbeck and Yiannis N Kaznessis. A closure scheme for chemical master equations. *Proceedings of the National Academy of Sciences*, 110(35):14261–14265, 2013.
- [58] Alexander Andreychenko, Lina Mikeev, and Verena Wolf. Model reconstruction for moment-based stochastic chemical kinetics. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 25(2):12, 2015.
- [59] David F Anderson, Gheorghe Craciun, and Thomas G Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bulletin of Mathematical Biology*, 72(8):1947–1970, 2010.
- [60] David F Anderson and Simon L Cotter. Product-form stationary distributions for deficiency zero networks with non-mass action kinetics. *Bulletin of Mathematical Biology*, 78(12):2390–2407, 2016.
- [61] David F Anderson, David Schnoerr, and Chaojie Yuan. Time-dependent product-form poisson distributions for reaction networks with higher order complexes. *arXiv preprint arXiv:1904.11583*, 2019.
- [62] Peter Deuffhard, Wilhelm Huisinga, Tobias Jahnke, and Michael Wulkow. Adaptive discrete Galerkin methods applied to the chemical master equation. *SIAM Journal on Scientific Computing*, 30(6):2990–3011, 2008.
- [63] C W Gardiner and S Chaturvedi. The Poisson representation. i. a new technique for chemical master equations. *Journal of Statistical Physics*, 17(6):429–468, 1977.
- [64] Jun Ohkubo. Approximation scheme for master equations: Variational approach to multivariate case. *The Journal of Chemical Physics*, 129(4):044108, 2008.
- [65] Maria Infusino, Tobias Kuna, Joel L Lebowitz, and Eugene R Speer. The truncated moment problem on \mathbb{N}_0 . *Journal of Mathematical Analysis and Applications*, 452(1):443–468, 2017.
- [66] Friedrich Schlögl. Chemical reaction models for non-equilibrium phase transitions. *Zeitschrift für Physik*, 253(2):147–161, 1972.
- [67] Timothy I Matis and Ivan G Guardiola. Achieving moment closure through cumulant neglect. *The Mathematica Journal*, 12, 2010.
- [68] Peter McCullagh. Cumulants and partition lattices. In *Selected works of Terry Speed*, pages 277–293. Springer, 2012.

- [69] R Pawula. Generalizations and extensions of the Fokker-Planck-Kolmogorov equations. *IEEE Transactions on Information Theory*, 13(1):33–41, 1967.
- [70] H Risken and HD Vollmer. On the application of truncated generalized Fokker-Planck equations. *Zeitschrift für Physik B Condensed Matter*, 35(3):313–315, 1979.
- [71] Colin S Gillespie. Moment-closure approximations for mass-action models. *Systems Biology, IET*, 3(1):52–58, 2009.
- [72] Norman TJ Bailey. *The elements of stochastic processes with applications to the natural sciences*. John Wiley & Sons, 1990.
- [73] Sadao Nakajima. On quantum theory of transport phenomena: steady diffusion. *Progress of Theoretical Physics*, 20(6):948–959, 1958.
- [74] Robert Zwanzig. Ensemble method in the theory of irreversibility. *The Journal of Chemical Physics*, 33(5):1338–1341, 1960.
- [75] Hazime Mori. Transport, collective motion, and brownian motion. *Progress of Theoretical Physics*, 33(3):423–455, 1965.
- [76] Daniele Venturi and George Em Karniadakis. Convolutionless Nakajima-Zwanzig equations for stochastic analysis in nonlinear dynamical systems. *Proceedings of the Royal Society A*, 470(2166):20130754, 2014.
- [77] NG Van Kampen. A cumulant expansion for stochastic linear differential equations. i. *Physica*, 74(2):215–238, 1974.
- [78] RH Terwiel. Projection operator method applied to stochastic linear differential equations. *Physica*, 74(2):248–265, 1974.
- [79] Steven N Evans, Bernd Sturmfels, and Caroline Uhler. Commuting birth-and-death processes. *The Annals of Applied Probability*, pages 238–266, 2010.
- [80] NG Van Kampen. A cumulant expansion for stochastic linear differential equations. ii. *Physica*, 74(2):239–247, 1974.
- [81] Gerhard C Hegerfeldt and Henrik Schulze. Noncommutative cumulants for stochastic differential equations and for generalized Dyson series. *Journal of Statistical Physics*, 51(3-4):691–710, 1988.
- [82] Peter Neu and Roland Speicher. A self-consistent master equation and a new kind of cumulants. *Zeitschrift für Physik B Condensed Matter*, 92(3):399–407, 1993.
- [83] Kay Giesecke, Hossein Kakavand, and Mohammad Mousavi. Exact simulation of point processes with stochastic intensities. *Operations Research*, 59(5):1233–1245, 2011.
- [84] Alan Bain and Dan Crisan. *Fundamentals of stochastic filtering*. Springer, 2009.
- [85] Bernard Derrida. An exactly soluble non-equilibrium system: The asymmetric simple exclusion process. *Physics Reports*, 301(1-3):65–83, 1998.

-
- [86] Alina Crudu, Arnaud Debussche, Aurélie Muller, and Ovidiu Radulescu. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *The Annals of Applied Probability*, 22(5):1822–1859, 2012.
- [87] Bernt Øksendal. *Stochastic differential equations*. Springer, 2003.
- [88] Robert Graham. Path integral formulation of general diffusion processes. *Zeitschrift für Physik B Condensed Matter*, 26(3):281–290, 1977.
- [89] Franco Flandoli, Giovanni Zanco, et al. An infinite-dimensional approach to path-dependent Kolmogorov equations. *The Annals of Probability*, 44(4):2643–2693, 2016.
- [90] Nihat Ay, Jürgen Jost, Hông Vân Lê, and Lorenz Schwachhöfer. *Information geometry*. Springer, 2017.
- [91] Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008.
- [92] Philipp Thomas, Christian Fleck, Ramon Grima, and Nikola Popović. System size expansion using Feynman rules and diagrams. *Journal of Physics A: Mathematical and Theoretical*, 47(45):455007, 2014.
- [93] Aurélien Alfonsi, Eric Cances, Gabriel Turinici, Barbara Di Ventura, and Wilhelm Huisinga. Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. In *ESAIM: Proc.*, volume 14, pages 1–13. EDP Sciences, 2005.

Erklärungen laut Promotionsordnung

§ 8 Abs. 1 lit. c PromO

Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§ 8 Abs. 1 lit. d PromO

Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§ 9 Abs. 1 PromO

Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§ 9 Abs. 2 PromO

Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Datum und Unterschrift