Zuse Institute Berlin

Takustrasse 7
D-14195 Berlin-Dahlem
Germany

VIKRAM SUNKARA[1,2]

[1]*Zuse Institute Berlin, Germany*

[2]*Department of Mathematics and Computer Science, Freie Universität Berlin, Germany*

ALGEBRAIC EXPRESSIONS OF CONDITIONAL EXPECTATIONS IN GENE REGULATORY NETWORKS

ZIB Report 18-52 (September 2018)

Zuse Institute Berlin
Takustrasse 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Algebraic Expressions of Conditional Expectations in Gene Regulatory Networks.

Vikram Sunkara[1,2]

[1]Computational Medicine, Konrad-Zuse-Zentrum für Informationstechnik, 14195 Berlin, Germany.
[2] Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany.
[1]Email: sunkara@mi.fu-berlin.de

September 12, 2018

Abstract

Gene Regulatory Networks are powerful models for describing the mechanisms and dynamics inside a cell. These networks are generally large in dimension and seldom yield analytical formulations. It was shown that studying the conditional expectations between dimensions (vertices or species) of a network could lead to drastic dimension reduction. These conditional expectations were classically given by solving equations of motions derived from the Chemical Master Equation. In this paper we deviate from this convention and take an *Algebraic* approach instead. That is, we explore the consequences of conditional expectations being described by a polynomial function. There are two main results in this work. Firstly: if the conditional expectation can be described by a polynomial function, then coefficients of this polynomial function can be reconstructed using the classical moments. And secondly: there are dimensions in Gene Regulatory Networks which inherently have conditional expectations with algebraic forms. We demonstrate through examples, that the theory derived in this work can be used to develop new and effective numerical schemes for forward simulation and parameter inference. The algebraic line of investigation of conditional expectations has considerable scope to be applied to many different aspects of Gene Regulatory Networks; this paper serves as a preliminary commentary in this direction.

Keywords: Markov Chains, Chemical Master Equation, Dimension Reduction.

1 Introduction

An average human being is made up of 37 trillion cells, each of these cells belongs to one of a hundred different phenotypes, and each of the phenotypes has differentiated from a single zygote. Each cell has an identical copy of DNA tracing back to that first zygote. Nevertheless, it performs its specialised function meticulously by using only specific segments of the DNA. It is natural to assume that a change in phenotype is permanent; that some physical change occurs which fixes the state of the cell to its purpose–a hammer doesn't turn into a screw driver. This intuition is true in human engineered systems, but far from reality in emergent complex systems like the cell. A cell is constantly active, at any moment several genes are being activated and deactivated resulting in what appears as a fixed phenotype on long time scales. These interactions/dynamics inside a cell can be described through the framework of networks (directed graphs), where the vertices of the network represent molecular structures such as genes, RNA, and proteins, and the edges represent reaction channels which describe the interactions between the respective vertices [42, 17, 22, 25]. These networks are referred to as *Gene Regulatory Networks* (GRNs). Every cell's cellular process can be described by a GRN; where a certain group of genes perform a function for the cell through transcription, translation, and regulation. One example is the Circadian rhythm in mammals, where the internal cellular clock is maintained by two genes activating and deactivating each other to produce a sinusoidal profile in $mCRY$ and $mPER$ protein concentrations through time [4, 41]. Another example is the lactose digestion in bacteria, where a single lactose molecule binds to the repressor protein near the lactase gene and frees the promotor region, which then starts the production of lactase needed for digesting the lactose [8]. Modelling a cellular process as a network helps in understanding its underlying mechanisms, which in turn can aid in predicting its behaviour.

Mathematically speaking, studying a GRN entails studying the various paths which can be traversed on that network. It was shown experimentally in the early 90s, that the internal processes of the cell are inherently stochastic [5, 35]. Hence, every path over the network is possible and each path has a probability of being realised.

Biological processes inherently have many interacting genes and molecules involved in complex interaction schemes, hence, investigating all possible paths over such networks is a real mathematical challenge. The class of GRNs which can be studied analytically is very small [19, 12]. In the majority of cases, numerical schemes have to be applied to study GRNs. These numerical schemes are covered under two overarching principles: *time scale separation* and *volume size expansion*. The idea behind time scale separation is that if there are sub-networks which are traversed faster than the rest of the network, these sub-networks can be reduced, which in turn would lead to a smaller network to study [30, 13, 11, 2, 6, 1, 26, 20, 29] . The second principle is that of volume size expansion, where in essence, the rate at which one traverses the network is adjusted such that the rate of traversing between any two connected vertices is of the same order. This in turn reduces the influence of stochasticity, which leads to an overall deterministic behaviour in the network [41, 32, 40, 7, 39, 29]. Both these principles are highly effective at reducing the computational complexity of many GRNs. However, they do not cover the whole range of dynamics, in particular, they have difficulties in accurately capturing GRNs which have cycles inside them [9, 15]. Cycles in GRNs imply that the system never reaches some particular steady state, rather it is in a permanent transient state through time. There are no fixed time scales or fixed volume sizes in such dynamic behaviour, these assumptions have to be updated in each phase of the cycle. Hence, a larger framework is needed, which encompasses the GRNs currently studied via the principles of time scale separation and volume size expansion, and further extends to include GRNs with cyclical dynamics.

In this paper, we formulate a new *algebraic conditional expectation* framework for studying GRNs. That is, we explore the consequences of conditional expectations being described by a polynomial function. We will begin by motivating the essentialness of conditional expectations for dimension reduction in the context of complex high dimensional GRNs. We then introduce the principle of algebraic conditional expectation forms, we will begin with the linear case to build intuition and then extend it to the general polynomial case. Then we will prove that the algebraic forms naturally arise in GRNs. We finish the paper by demonstrating how the properties of algebraic conditional expectations can be utilised to build new numerical schemes for studying GRNs.

2 Mathematical Background

In this section, we will introduce the key processes and notations which are to be used in this paper. We begin by defining a *Kurtz* process, a jump Markov process, which is commonly used for modelling GRNs. As the study is focused on investigating *conditional* expectations, we present the necessary assumptions to guarantee their existence and computability. After formulating the underlying assumptions, we will introduce the *Dimension Reduction* framework, and introduce a new algebraic perspective on the topic.

2.1 Kurtz Process

Let $\mathcal{S}_1, \dots, \mathcal{S}_{N_s}$ be population counts of N_s different *species* which can interact with each other. [1] A *reaction channel* is a transformation of a set of species into another. Mathematically, a reaction channel (R) is described by the following mapping:

$$R := \sum_{i=1}^{N_s} \chi_i^{in} \mathcal{S}_i \longrightarrow \sum_{i=1}^{N_s} \chi_i^{out} \mathcal{S}_i,$$

where $\chi_i^{in/out}$ are the number of species i which are entering and exiting the reaction channel respectively. The *stoichiometry* of the reaction channel (R) is the vector describing the net change in population after the reaction channel has fired. We denote this by

$$v_R := (\chi_i^{out} - \chi_i^{in})_{i=1}^{N_s}.$$

The *propensity/intensity* at which a reaction channel fires is given by the function:

$$f_R(\mathcal{S}_1, \dots, \mathcal{S}_{N_s}) := c_R \prod_{i=1}^{N_s} \binom{\mathcal{S}_i}{\chi_i^{in}}, \tag{2.1}$$

where c_R is a single event rate and the round brackets are the binomial coefficients. Functions in the form of f_R are referred to as *mass action* propensities [10]. That is, the rate of firing of a reaction channel is proportional to the product of the populations of the species involved in starting the reaction. This formalism emerged from thermodynamical description of the probability of molecules colliding and forming new molecules. We

[1] The terms species, dimensions, and vertices originate from different fields of study but refer to the same concept. Hence, we interchange between the terms to match the context.

are interested in systems with $N_s \in \mathbb{N}$ species undergoing $N_r \in \mathbb{N}$ reactions via mass action propensities. The stochastic process of such a system can be modelled by:

$$Z(t) = Z(0) + \sum_{j=1}^{N_r} \mathcal{P}\left(\int_0^t f_j(Z(s))ds\right) v_j, \text{ with } Z(t) \in \mathbb{N}_0^{N_s}, \tag{2.2}$$

where $\mathcal{P}$ is an inhomogeneous Poisson process [24]. As a shorthand, processes of the form (2.2) are referred to as a *Kurtz processes.*

The probability of observing $Z(t)$ in a state $\mathbf{z} \in \Omega \subset \mathbb{N}_0^{N_s}$ at time $t \in [0, \infty)$ is described by the Chemical Master Equation,

$$\frac{\partial\, p(Z = \mathbf{z}; t)}{\partial t} = \sum_{j=1}^{N_r} f_j(\mathbf{z} - v_j)\, p(Z = \mathbf{z} - v_j; t) - f_j(\mathbf{z})\, p(Z = \mathbf{z}; t). \tag{2.3}$$

For each reaction indexed by $j \in \{1, \ldots, N_r\}$, we denote the corresponding propensity functions by $f_j : \Omega \to \mathbb{R}^+$; and its stoichiometric vectors by $v_j \in \mathbb{Z}^{N_s}$. For simplicity of notation, we write the right-hand side of equation (2.3) as a shift minus an identity operator:

$$\frac{\partial\, p(\mathbf{z}; t)}{\partial t} = \sum_{j=1}^{N_r} (S_j - I)\, f_j(\mathbf{z})\, p(\mathbf{z}; t), \tag{2.4}$$

where S_j is the shift operator of the j^{th} reaction. We define p_t to be the vector $(p(Z = \mathbf{z}; t))_{\mathbf{z}\in\Omega}$, and dp_t/dt as the vector $(\partial\, p(Z = \mathbf{z}; t)/\partial t)_{\mathbf{z}\in\Omega}$. The solution of the CME (2.3) is found by solving the initial value problem:

$$\begin{cases} \frac{d\, p_t}{dt} = A\, p_t, & t > 0, \\ p_0 \in \ell_1(\Omega), & t = 0, \end{cases} \tag{2.5}$$

where A is an infinitesimal generator [42, 23, 37] with the properties:

$$A_{k,k} \leq 0, A_{k,l} \geq 0 \quad \text{for} \quad k \neq l \quad \text{and} \quad \sum_l A_{k,l} = 0,$$

for all $k, l \in \{1, \ldots, |\Omega|\}$. Solving the system of equations in (2.5) gives the full joint probability distribution of the system at a time point. Kurtz processes which yield analytical solutions to the equations in (2.5) are a small class [19]. In most common cases, numerical schemes have to be considered. We now introduce the necessary assumptions for guaranteeing the existence and computability of the solution of (2.5).

Assumption 2.1. *We assume the following:*

1. *the state space is finite,* $|\Omega| < \infty$.
2. *the operator* e^{At} *is honest* [3].
3. *the joint probability is non-negative over the state space,* $p(Z = \cdot\,; t) > 0$, *at all time points* t.

Assumption (1) and (2) guarantee that the probabilities of interest do exist and are computable. Assumption (3) is crucial for applying Bayes' Theorem–the critical tool for conditional probability–to help formulate and compute conditional moments.

2.2 Dimension Reduction

Any method which solves the CME (2.5) with less equations than $|\Omega|$ is technically a dimension reduction method. Within dimension reduction, we want to focus on a particular subclass of methods, that is, methods which reduce dimension by partitioning the species into two groups, a *stochastic* group and a *deterministic* group, and study their dynamics separately [21, 14, 27, 18, 16, 13, 20, 39, 41]. This class of methods is referred to as *hybrid methods.* In essence, the stochastic process $Z(t)$ is split into two sub-processes:

$$Z(t) = (X(t), Y(t)). \tag{2.6}$$

A hybrid method achieves dimension reduction by evolving $X(t)$ stochastically coupled with the statistics of $Y(t)$. With the splitting in (2.6) in mind, we now derive the general decomposition of a hybrid method for the CME. It is important to note that when we refer to a species as "deterministic," we do not model these species

deterministically, but we are rather referring to their statistics; which evolve deterministically (e.g. expectation, variance, or higher moments).

Let N_s^Y denote the number of species Y that are described deterministically, and N_s^X the number of species X described stochastically. Then the total number of species, $N_s = N_s^Y + N_s^X$, is equal to the sum of the two parts. Corresponding to (2.6), the state $\mathbf{z}$ is written as a tuple:

$$\mathbf{z} = (x, y), \text{ where } x \in \mathbb{N}_0^{N_s^X}, \ y \in \mathbb{N}_0^{N_s^Y}.$$

Similarly, the stoichiometric vector of the j^{th} reaction is written as a tuple $\upsilon_j = (\nu_j, \mu_j)$, where ν_j is the restriction of υ_j on the stochastically considered species and μ_j on the deterministic. The state space of the stochastically considered species is called the *hybrid state space*, and is denoted by $\Omega_X \subset \mathbb{N}_0^{N_s^X}$. Similarly, $\Omega_Y \subset \mathbb{N}_0^{N_s^Y}$, is the state space restricted to only deterministically considered species. The mass action propensities, (2.1), naturally split into the product structure:

$$f_j(\mathbf{z}) = g_j(x)\, h_j(y), \ \forall \mathbf{z} \in \Omega, \tag{2.7}$$

for non-negative functions g_j and h_j. The above splitting of species, stoichiometry and propensities is referred to as the *hybrid framework*. Given there exists a solution to the CME (2.3) at a given time $t > 0$ and the solution is non-zero on the state space–using Bayes' theorem–there exists a marginal probability distribution,

$$p(X = \cdot; t) : \Omega_X \to [0, 1],$$

and a family of conditional probability distributions,

$$\{p(Y = \cdot \mid X = x; t) : \Omega_Y \to [0, 1] \text{ for } x \in \Omega_X\},$$

such that,

$$p(\mathbf{z}; t) = p(X = x; t)\, p(Y = y \mid X = x; t). \tag{2.8}$$

Furthermore, by the law of conservation of probability, the marginal and conditional probability distributions satisfy:

$$\sum_x \frac{\partial\, p(X = x; t)}{\partial t} = 0 \ \text{ and } \ \forall x \in \Omega_X : \ \sum_y \frac{\partial\, p(Y = y \mid X = x; t)}{\partial t} = 0. \tag{2.9}$$

A hybrid method tries to reconstruct the marginal distribution $p(X = \cdot; t)$, using the statistics of the conditional distribution $p(Y = \cdot \,|\, X = \cdot; t)$. The equations of motion for $p(X = \cdot; t)$ are derived by substituting (2.8) and (2.9) into the CME and performing some simple algebra. We now derive the time derivatives of the marginal distribution and the conditional expectation.

Firstly, for the time derivative of the marginal distribution of $X(t)$, we sum the CME (2.4) over all the states y. The derivative with respect to time of the product form in (2.8) is:

$$\partial_t\, p(Z = \mathbf{z}; t) = p(Y = y \mid X = x; t)\, \partial_t\, p(X = x; t) + p(X = x; t)\, \partial_t\, p(Y = y \mid X = x; t).$$

Then, summing the above expression over y and applying the second condition in (2.9), gives $\partial_t\, p(X = x; t) = \sum_y \partial_t\, p(Z = \mathbf{z}; t)$. By substituting (2.3) in the left-hand side term and then expanding, we will have derived the time derivative of the marginal distribution $p(X = \cdot; t)$:

$$\begin{aligned}
\frac{\partial\, p(X = x; t)}{\partial t} &= \sum_y \frac{\partial\, p(Z = \mathbf{z}; t)}{\partial t} \\
&= \sum_y \sum_{j=1}^{N_r} (S_j - I)\, g_j(x)\, h_j(y)\, p(X = x; t)\, p(Y = y \mid X = x; t), \\
&= \sum_{j=1}^{N_r} (S_j - I) \underbrace{\left[\sum_y h_j(y)\, p(Y = y \mid X = x; t) \right]}_{(\star)} g_j(x)\, p(X = x; t).
\end{aligned} \tag{2.10}$$

We have derived the formula of the time derivative of the marginal distribution in the form of the CME. Firstly, we should observe that the propensity of the deterministically considered species, the function preceding the probability term $(\star)$, has become time dependent. For $x \in \Omega_X$, if we denote $Y_x(t)$ as the process distributed according to the conditional distribution, $p(Y = \cdot \mid X = x; t)$, then the term $(\star)$ in (2.10)–by definition–is the expectation of the propensity of the conditional process $Y_x(t)$;

$$\sum_y h_j(y)\, p(Y = y \mid X = x; t) := \mathbb{E}\left[h_j(Y_x(t)) \right]. \tag{2.11}$$

Hence, (2.10) can be rewritten as

$$\frac{\partial\, p(X=x;t)}{\partial t} = \sum_{j=1}^{N_r}(S_j - I)\, \underbrace{\mathbb{E}\left[h_j(Y_x(t))\right]}_{(\star)}\, g_j(x)\, p(X=x;t). \tag{2.12}$$

In this form, it is clear that to study or evolve the sub-processes of a high-dimensional process, conditional moments are needed. If these were known a priori, then any sub-process could be studied independently to the full process. This highlights the importance of conditional moments in the process of dimension reduction. The key principle behind every hybrid method is to harness the conditional expectation structure. To understand how the conditional moments are computed, let us consider the following equation for the conditional expectation given by the *Method of Conditional Moments* (the derivation for the equations can be found in [18, 14]):

$$\begin{aligned}\frac{\partial(\mathbb{E}[Y_x(t)]\, p(X=x;t))}{\partial t} = \sum_{j=1}^{N_r} \Bigg(& \underbrace{\left[\sum_y h_j(y)\, p(Y=y \mid X=x-\nu_j, t)\right]}_{\mathbb{E}[h_j(Y_{x-\nu_j}(t))]} \mu_j\, g_j(x-\nu_j)\, p(X=x-\nu_j;t) \\ & + \underbrace{\left[\sum_y y\, h_j(y)\, p(Y=y \mid X=x-\nu_j;t)\right]}_{\mathbb{E}[Y_{x-\nu_j}(t)\, h_j(Y_{x-\nu_j}(t))]} g_j(x-\nu_j)\, p(X=x-\nu_j;t) \\ & - \underbrace{\left[\sum_y y\, h_j(y)\, p(Y=y \mid X=x;t)\right]}_{\mathbb{E}[Y_x(t)\, h_j(Y_x(t))]} g_j(x)\, p(X=x;t) \Bigg).\end{aligned} \tag{2.13}$$

The above equation solves for the conditional expectation. The first thing to notice is that the time derivative on the left-hand side contains the marginal probability, and a large time scale separation between the conditional expectation and the marginal distribution would be needed to uncouple these terms in the derivative. Secondly, if h_j is a polynomial of degree greater than zero, then there are higher order conditional moments in the square brackets on the right-hand side. Computing these would lead to solving an infinite system of non-linear differential equations, which is clearly not feasible. Hence, approximations like moment truncation and moment closures have to be applied to make (2.13) computable [21, 20, 14]. In summary, while conditional expectations are critical for dimension reduction, computing conditional expectations is computationally challenging and highly non-trivial.

Let us return to (2.12). Here is where we wish to deviate slightly from convention. Instead of writing down the derivatives of the conditional moments, as done in (2.13), and imposing closures and truncations, we want to pursue a more algebraic approach. We begin by asking the following questions:

- What dynamics would be observed if the conditional moments $(*)$ had a polynomial form?
- If the conditional expectation structure between two dimensions was a linear function, could the equations of motion for the slope and the intercept be derived?
- Do certain reactions or stoichiometries of GRNs guarantee that the conditional moments are embedded in some curve/manifold?

Giving a rigorous answer to the latter question is out of the scope of this work, but we begin by exploring the preliminary question: "If we were to consider some polynomial ansatz for the conditional moments, then which structures and properties would emerge?"

3 Algebraic Conditional Expectation

In this section we will study the relationship between conditional moments and classical moments [2]. Firstly, we will investigate the consequences of the conditional expectation having a linear algebraic form. After gaining intuition from the linear case, we will extend the results to a generalised polynomial form. That is, we will prove that the coefficients of a polynomial which describes the conditional moments can be computed by solving a

[2] In our context the results can be reformulated to be raw moments, factorial moments, or central moments .For this reason we say classical moments to encompass it all.

linear system of equations containing the joint and marginal moments of the random variables. Furthermore, we show that the existence of an algebraic conditional expectation inherently guarantees moment closure.

The notation from §2.2 is carried through and used in the rest of this section. In this section, the interest is strictly on processes at a fixed time point. Hence, the time variable is omitted from the notation. We quickly recall that X, Y are random variables over their respective state spaces: $\Omega_X \subset \mathbb{N}_0^{N_s^X}$, and $\Omega_Y \subset \mathbb{N}_0^{N_s^Y}$. We assume that the random variables, Y conditioned on x, for all $x \in \Omega_X \subset \mathbb{N}_0^{N_s^X}$ exist. These random variables are denoted as Y_x, for $x \in \Omega_X \subset \mathbb{N}_0^{N_s^Y}$. Henceforth, whenever we refer to "conditional expectation" or "conditional moments", the conditional variable is the dimension Y and the conditioning variable is the dimension X.

Remark 3.1. *To be rigorous, since the state space Ω_X is not continuous, the conditional expectation cannot be a continuous function, but rather there is a smooth manifold in which the conditional expectation, $\mathbb{E}[Y_x]$, is embedded. For brevity we will say that the conditional expectation has the form of the manifold in which it is embedded.*

3.1 Linear Conditional Expectation

Lemma 3.1. *Let $\alpha \in \mathbb{M}_{N_s^Y \times N_s^X}$ and $\beta \in \mathbb{M}_{N_s^Y \times 1}$ be fixed. If the expectation of Y conditioned on $x \in \Omega_X$ has a linear form, that is, for all $x \in \Omega_X$,*

$$\mathbb{E}[Y_x] = \alpha\, x + \beta, \tag{3.1}$$

then

1. $\mathbb{E}[Y] = \alpha\, \mathbb{E}[X] + \beta$,
2. $\operatorname{cov}(Y, X) = \alpha \operatorname{cov}(X, X)$,
3. $\mathbb{E}[\operatorname{cov}(Y_x, Y_x)] = \operatorname{cov}(Y, Y) - \alpha \operatorname{cov}(X, X)\, \alpha^T$.

Proof. Fix $\alpha \in \mathbb{M}_{N_s^Y \times N_s^X}$ and $\beta \in \mathbb{M}_{N_s^Y \times 1}$. We will prove the three statements separately.

Statement 1: We multiply (3.1) by the marginal probability of x and then sum over all x:

$$\begin{aligned} \mathbb{E}[Y_x] &= \alpha\, x + \beta, \\ \sum_{x \in \Omega_X} \mathbb{E}[Y_x]\; p(X = x) &= \sum_{x \in \Omega_X} (\alpha\, x + \beta)\, p(X = x), \\ \mathbb{E}[Y] &= \alpha\, \mathbb{E}[X] + \beta. \end{aligned}$$

Statement 2: We begin with the definition of the covariance of X and Y,

$$\operatorname{cov}(Y, X) := \sum_{x \in \Omega_X, y \in \Omega_Y} (y - \mathbb{E}[Y])\, (x - \mathbb{E}[X])^T\, p(X = x, Y = y).$$

Applying Bayes' Theorem to the joint distribution and then collating the y related terms gives us:

$$\begin{aligned} &= \sum_{x \in \Omega_X, y \in \Omega_Y} (y - \mathbb{E}[Y])\, (x - \mathbb{E}[X])^T\, p(X = x)\, p(Y = y \mid X = x), \\ &= \sum_{x \in \Omega_X} \left[\sum_{y \in \Omega_Y} (y - \mathbb{E}[Y])\, p(Y = y \mid X = x) \right] (x - \mathbb{E}[X])^T\, p(X = x), \end{aligned}$$

then expanding the square brackets gives

$$= \sum_{x \in \Omega_X} (\mathbb{E}[Y_x] - \mathbb{E}[Y])\, (x - \mathbb{E}[X])^T\, p(X = x).$$

Lastly, we substitute in the linear form (3.1) for the conditional expectation and then reduce:

$$\begin{aligned} &= \sum_{x \in \Omega_X} (\alpha\, x + \beta - \mathbb{E}[Y])\, (x - \mathbb{E}[X])^T\, p(X = x), \\ &= \sum_{x \in \Omega_X} \left[\alpha\, x\, x^T - \alpha\, x\, \mathbb{E}[X]^T + \beta\, x^T - \beta\, \mathbb{E}[X]^T - \mathbb{E}[Y]\, x^T + \mathbb{E}[Y]\, \mathbb{E}[X]^T \right] p(X = x), \\ &= \alpha\, \mathbb{E}[X X^T] - \alpha\, \mathbb{E}[X]\, \mathbb{E}[X]^T, \\ &= \alpha \operatorname{cov}(X, X). \end{aligned}$$

Statement 3: Proof in Appendix A. The idea of the proof is to substitute the linear conditional form (3.1) into Eve's Law and then reduce. □

Lemma 3.1-1 states that the conditional expectation form intersects the point $(\mathbb{E}[X], \mathbb{E}[Y])$. Furthermore, Lemma 3.1-2 states that the covariance of X and Y is a scaler of the variance of X. Writing these two equations together gives the following system of equations:

$$\begin{bmatrix} \operatorname{cov}(X,X) & 0 \\ \mathbb{E}[X] & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \operatorname{cov}(Y,X) \\ \mathbb{E}[Y] \end{bmatrix}.$$

We notice that solving the equation above gives the formula for the gradient and the intercept of the conditional expectation in terms of the moments:

$$\alpha = \frac{\operatorname{cov}(Y,X)}{\operatorname{cov}(X,X)} \text{ and } \beta = -\alpha\,\mathbb{E}[X] + \mathbb{E}[Y]. \tag{3.2}$$

We have deduced that if the conditional expectation has a linear form, then the gradient and the intercept can be calculated using classical moments. We will now extend this observation to the case of general polynomial forms.

3.2 Polynomial Conditional Expectation

For brevity, we fix $N_s^Y, N_s^X = 1$.

Theorem 3.1. *Let $m \in \mathbb{N}_0$ with $m \ll |\Omega_X|$. If the expectation of Y conditioned on $x \in \Omega_X$ has a degree m polynomial form, that is, for all $x \in \Omega_X$,*

$$\mathbb{E}[Y_x] = \kappa_m\, x^m + \kappa_{m-1}\, x^{m-1} + \ldots + \kappa_1\, x + \kappa_0, \tag{3.3}$$

then for $n \in \mathbb{N}_0$,

$$\mathbb{E}[Y\, X^n] = \sum_{i=0}^{m} \kappa_i\, \mathbb{E}[X^{i+n}]. \tag{3.4}$$

Proof. Fix $n \in \mathbb{N}_0$, $m \in \mathbb{N}_0$. We prove the statement by expanding the definition of $\mathbb{E}[Y\, X^n]$, then substituting in the polynomial form and reducing.

$$\mathbb{E}[Y\, X^n] := \sum_{x \in \Omega_X, y \in \Omega_Y} y\, x^n\; p(X = x, Y = y),$$

applying Bayes' theorem to the joint distribution gives

$$= \sum_{x \in \Omega_X, y \in \Omega_Y} y\, x^n\; p(Y = y \,|\, X = x)\; p(X = x),$$

then collating the y terms reduces the expression to

$$= \sum_{x \in \Omega_X} \left(\underbrace{\sum_{y \in \Omega_Y} y\; p(Y = y \,|\, X = x)}_{\mathbb{E}[Y_x]} \right) x^n\; p(X = x),$$

substituting (3.3) for the conditional expectation gives

$$= \sum_{x \in \Omega_X} \left(\sum_{i=0}^{m} \kappa_i\, x^i \right) x^n\; p(X = x).$$

Lastly, interchanging the summations gives:

$$= \sum_{i=0}^{m} \kappa_i \left(\sum_{x \in \Omega_X} x^i\, x^n\; p(X = x) \right),$$
$$= \sum_{i=0}^{m} \kappa_i\, \mathbb{E}[X^{i+n}].$$

□

Corollary 3.1. *Let*

- $\kappa := (\kappa_i)_{i=0}^m, \in \mathbb{R}_{m+1\times 1}$
- $\Xi := [\mathbb{E}[X^{i+j}]]_{i,j} \in \mathbb{M}_{m+1\times m+1}$
- $\mu := (\mathbb{E}[Y\, X^i])_{i=0}^m \in \mathbb{R}_{m+1\times 1}$

If Ξ is invertible, then

$$\kappa = \Xi^{-1}\,\mu. \tag{3.5}$$

Furthermore, if Ξ is invertible and $[\Xi^{-1}]_{1,m+1} \neq 0$, then

$$\mathbb{E}[Y\, X^m] = \left(\kappa_m - \sum_{i=0}^{m-1} [\Xi^{-1}]_{1,i+1}\, \mathbb{E}[Y\, X^i]\right) / [\Xi^{-1}]_{1,m+1}. \tag{3.6}$$

Proof. The linear system of equations in (3.5) arises by simply iterating the term $\mathbb{E}[Y\, X^n]$, as defined in (3.4), for $n = 0, \ldots, m$:

$$\underbrace{\begin{bmatrix} \mathbb{E}[X^m] & \mathbb{E}[X^{m-1}] & \ldots & \ldots & \mathbb{E}[X] & 1 \\ \mathbb{E}[X^{m+1}] & \mathbb{E}[X^m] & \ldots & \ldots & \mathbb{E}[X^2] & \mathbb{E}[X] \\ & & & & & \\ & & & & & \\ \mathbb{E}[X^{2m}] & \mathbb{E}[X^{2m-1}] & \ldots & \ldots & \mathbb{E}[X^{m+1}] & \mathbb{E}[X^m] \end{bmatrix}}_{\Xi} \bullet \underbrace{\begin{bmatrix} \kappa_m \\ \kappa_{m-1} \\ \\ \\ \kappa_0 \end{bmatrix}}_{\kappa} = \underbrace{\begin{bmatrix} \mathbb{E}[Y] \\ \mathbb{E}[Y\, X] \\ \\ \\ \mathbb{E}[Y\, X^m] \end{bmatrix}}_{\mu}$$

Since Ξ is invertible, it follows that coefficients in (3.3) can be computed by evaluating $\Xi^{-1}\,\mu$.

The second statement can be proved by taking the dot product of the first row of Ξ^{-1} and μ and rearranging to make $\mathbb{E}[Y\, X^m]$ the subject. We see that if $\kappa_m = 0$, then $\mathbb{E}[Y\, X^m]$ has a natural moment closure.

□

Definition 3.1. *Let $\eta^m_{Y|X}(x)$ denote the m^{th} degree polynomial approximation of $\mathbb{E}[Y_x] : \Omega_X \to \mathbb{R}_+$, where*

$$\eta^m_{Y|X}(x) := \sum_{i=0}^{m} \kappa_i\, x^i, \tag{3.7}$$

with the coefficients κ_i, for $i = 0, \ldots, m$, defined in (3.5).

Remark 3.2. *The idea of fitting polynomial structures to stochastic data was originally investigated by the data driven sciences; where functions were fitted to high dimensional point clouds to unravel the dynamics which generated that data set [31, 28]. In the data science framework, polynomial approximation–which we formulated in Def. 3.1–is equivalent to a polynomial regression. That is, if the moments were replaced with empirical moments of a dataset, the m^{th} degree polynomial approximation reduces to the m^{th} degree polynomial regression.*

Example 3.1. *So far, the hypotheses in the statements above have begun with "If the conditional expectation has the form..." A natural question that arises is if Kurtz processes can even satisfy this part of the hypotheses. To give some intuition into this question, let us consider the following three simple mRNA translation models:*

(***Model 1***)	$X \leftrightarrow \emptyset$	$Y \leftrightarrow \emptyset,$	
(***Model 2***)	$X \leftrightarrow \emptyset$	$X \to Y + X$	$Y \to \emptyset,$
(***Model 3***)	$X \leftrightarrow \emptyset$	$2X \to Y + 2X$	$Y \to \emptyset.$

All three models contain two species, X and Y. In the first model, both species are produced via a constant propensity function and decay via a linear propensity function, with their respective rate constants. The two species do not interact with each other. The second model is a simple mRNA translation model. In this model, X is produced and decayed as in Model 1. Species Y on the other hand, is produced by species X via an autocatalytic *reaction. That is, X produces Y and preserves its population in the process. The third model extends on the second model, with the autocatalytic reaction needing two Xs to perform the reaction. The model parameter can be found in Appendix B. We start all the models with an initial population of $(0, 0)$ and compute the stationary distribution using the Optimal Finite State Projection method. The left column plots in Fig. 1 show the stationary distribution of these three models, respectively. In the right column of Fig. 1, we plot their conditional expectations. We can see that these models exhibit algebraic conditional expectation forms.*

In summary, Theorem 3.1 states that if the conditional expectation has a polynomial form, then the coefficients of this polynomial form can be computed using the classical moments. In essence, we have reduced the information of the conditional moments to the classical moments. Furthermore, Corollary 3.1 shows that if the coefficient of the lead term of the polynomial is small, then natural moment closures arise. In §5, we will develop some simple numerical schemes–as a proof of concept–which exploit these results. Before starting with applications, we need to explore and study the switching behaviour, which is inherent in all GRNs, more rigorously.

4 Switching Behaviour

We define a *switch* to be a random variable which has only two states in its state space. Naturally, we interpret the two states as "on" and "off". The biological dynamics of a gene are accurately modelled as a switch, hence, all GRNs inherently contain switches. When genes are activated (transcription), they start producing some *particles* (RNAs/proteins), depending on which tier the GRN is modelling. It was shown that this switching behaviour of the genes is inherently stochastic and furthermore, induces multi-modal behaviour [12, 39]. In this section, we want to unravel the consequences of the switching behaviour using the conditional expectation forms derived earlier. We will construct the simplest possible GRN, a network containing one switch and one particle type, with non-zero covariance. We then derive the forms of the conditional expectation of particles from the perspective of the switch and vice versa. We also show that the derived results extend to all GRNs, as a general GRN can be decomposed into blocks of this simple *switch-particles* network.

Definition 4.1. *We define S to be a random variable over the state space $\Omega_S := \{0, 1\}$. Intuitively we refer to S as a "switch" and write 0 as **off** and 1 as **on**. We define Y to be a random variable over the state space $\Omega_Y \subset \mathbb{N}_0$. Since samples of Y are non-negative integers, we refer to Y as "particles". We couple these two random variables by imposing that $\mathrm{cov}(S, Y) \neq 0$. Furthermore, Assumptions 2.1 hold for the joint distribution of S and Y.*

4.1 A Switch's perspective

Lemma 4.1. *Let S be a switch and Y be particles as in Def. 4.1. Then the following statements hold:*

1. $\mathbb{E}[S] = p(S = \boldsymbol{on})$
2. $\mathrm{cov}(S, S) = \mathbb{E}[S] - \mathbb{E}[S]^2$
3. $\mathbb{E}[SY] = \mathbb{E}[Y_{\boldsymbol{on}}]\ p(S = \boldsymbol{on})$
4. $\mathrm{cov}(Y, S) = (\mathbb{E}[Y_{\boldsymbol{on}}] - \mathbb{E}[Y])\ p(S = \boldsymbol{on})$
5. $\mathbb{E}[S^2 Y] = \mathbb{E}[SY]$

Proof. We will prove the statements in order.

Statement 1:

$$\begin{aligned}\mathbb{E}[S] &:= 0\ p(S = \mathbf{off}) + 1\ p(S = \mathbf{on}),\\ &= p(S = \mathbf{on}).\end{aligned}$$

Statement 2:

$$\begin{aligned}\mathrm{cov}(S, S) &:= (0 - \mathbb{E}[S])^2\ p(S = \mathbf{off}) + (1 - \mathbb{E}[S])^2\ p(S = \mathbf{on}),\\ &= \mathbb{E}[S]^2\,(1 - p(S = \mathbf{on})) + (1 - 2\,\mathbb{E}[S] + \mathbb{E}[S]^2)\ p(S = \mathbf{on}),\\ &= \mathbb{E}[S]^2 - \mathbb{E}[S]^2\ p(S = \mathbf{on}) + p(S = \mathbf{on}) - 2\,\mathbb{E}[S]\ p(S = \mathbf{on}) + \mathbb{E}[S]^2\ p(S = \mathbf{on}),\\ &= \mathbb{E}[S]^2 - \mathbb{E}[S]^3 + \mathbb{E}[S] - 2\,\mathbb{E}[S]^2\ + \mathbb{E}[S]^3,\\ &= \mathbb{E}[S] - \mathbb{E}[S]^2.\end{aligned}$$

Statement 3:

$$\begin{aligned}\mathbb{E}[SY] &:= \sum_{y \in \Omega_Y} 0 \times y \times p(Y = y, S = \mathbf{off}) + 1 \times y \times p(Y = y, S = \mathbf{on}),\\ &= \sum_{y \in \Omega_Y} y\ p(Y = y, S = \mathbf{on}),\\ &= \sum_{y \in \Omega_Y} y\ p(Y = y \mid S = \mathbf{on})\ p(S = \mathbf{on}),\\ &= \mathbb{E}[Y_{\mathbf{on}}]\ p(S = \mathbf{on}).\end{aligned}$$

Statement 4:

$$\mathrm{cov}(Y,S) := \mathbb{E}[SY] - \mathbb{E}[S]\,\mathbb{E}[Y],$$

applying statement *3* and *1* reduces the right-hand side terms to:

$$\begin{aligned} &= \mathbb{E}[Y_{\mathbf{on}}]\; p(S=\mathbf{on}) - \mathbb{E}[Y]\; p(S=\mathbf{on}), \\ &= \left(\mathbb{E}[Y_{\mathbf{on}}] - \mathbb{E}[Y]\right)\; p(S=\mathbf{on}). \end{aligned}$$

Statement 5: Since $S = S^2$, the value of $\mathbb{E}[S^2Y]$ is the same as $\mathbb{E}[SY]$. □

Theorem 4.1. *Let S be a switch and Y be particles as in Def. 4.1, then*

$$\mathbb{E}[Y_{\mathbf{off}}] = -\frac{\mathrm{cov}(Y,S)}{\mathrm{cov}(S,S)}\mathbb{E}[S] + \mathbb{E}[Y],$$

$$\mathbb{E}[Y_{\mathbf{on}}] = \frac{\mathrm{cov}(Y,S)}{\mathrm{cov}(S,S)}\big(1 - \mathbb{E}[S]\big) + \mathbb{E}[Y].$$

Proof. We first prove the conditional expectation of the particles given the switch is **off**. We begin by rearranging the definition $\mathbb{E}[Y] := \mathbb{E}[Y_{\mathbf{on}}]\; p(S=\mathbf{on}) + \mathbb{E}[Y_{\mathbf{off}}]\; p(S=\mathbf{off})$, and then reduce using the statements of Lemma 4.1:

$$\mathbb{E}[Y_{\mathbf{off}}] := \frac{-\mathbb{E}[Y_{\mathbf{on}}]\; p(S=\mathbf{on}) + \mathbb{E}[Y]}{p(S=\mathbf{off})}.$$

Rearraging Lemma 4.1-4 for $\mathbb{E}[Y_{\mathbf{on}}]$ and substituting it in gives:

$$\begin{aligned} &= \frac{-\left(\mathrm{cov}(Y,S)/\; p(S=\mathbf{on}) + \mathbb{E}[Y]\right)\; p(S=\mathbf{on}) + \mathbb{E}[Y]}{(1-\mathbb{E}[S])}, \\ &= \frac{-\mathrm{cov}(Y,S) - \mathbb{E}[Y]\; p(S=\mathbf{on}) + \mathbb{E}[Y]}{(1-\mathbb{E}[S])}, \\ &= \frac{-\mathrm{cov}(Y,S)}{(1-\mathbb{E}[S])} + \frac{(1-\mathbb{E}[S])\,\mathbb{E}[Y]}{(1-\mathbb{E}[S])}, \end{aligned}$$

multiplying the top and bottom by $\mathbb{E}[S]$ and applying Lemma 4.1-2 gives us

$$= -\frac{\mathrm{cov}(Y,S)}{\mathrm{cov}(S,S)}\,\mathbb{E}[S] + \mathbb{E}[Y].$$

Now we prove the conditional expectation of the particles given the switch is **on**. We begin by rearranging Lemma 4.1-4 and reducing:

$$\mathbb{E}[Y_{\mathbf{on}}] = \frac{\mathrm{cov}(Y,S) + \mathbb{E}[S]\,\mathbb{E}[Y]}{\mathbb{E}[S]},$$

multiplying top and bottom by $1 - \mathbb{E}[S]$ gives us

$$= \frac{\mathrm{cov}(Y,S)\,(1-\mathbb{E}[S])}{\mathbb{E}[S]\,(1-\mathbb{E}[S])} + \mathbb{E}[Y],$$

then substituting Lemma 4.1-2 into the denominator gives us

$$= \frac{\mathrm{cov}(Y,S)}{\mathrm{cov}(S,S)}\,(1-\mathbb{E}[S]) + \mathbb{E}[Y].$$

If we compare the components of the conditional expectation of the switch to the coefficients of the linear conditional expectation form, (3.2), we see that the conditional expectation of particles with respect to a switch has a linear form. □

Theorem 4.1 states that the gradient of the line which intersects the points $(\mathbf{off}, \mathbb{E}[Y_{\mathbf{off}}])$ and $(\mathbf{on}, \mathbb{E}[Y_{\mathbf{on}}])$ is given by $\mathrm{cov}(Y,S)/\mathrm{cov}(S,S)$. By extending this further with the observations on linear conditional expectation forms in Lemma 3.1, we can show that the line which goes through $(\mathbf{off}, \mathbb{E}[Y_{\mathbf{off}}])$ and $(\mathbf{on}, \mathbb{E}[Y_{\mathbf{on}}])$ also goes through $(\mathbb{E}[S], \mathbb{E}[Y])$ (see Fig. 2A). Hence, if any three out of the four terms: $\mathbb{E}[S]$, $\mathbb{E}[Y]$, $\mathbb{E}[Y_{\mathbf{off}}]$, and $\mathbb{E}[Y_{\mathbf{on}}]$ are known, then the fourth term–which is unknown–can be reconstructed using Theorem 4.1.

Remark 4.1. *We have only given results for a single switch case. Nevertheless, these results can be extended to multiple switch cases since the multiple switch problem can be reformulated into coupled single switch problems by increasing the number of dimensions* (see Fig. 2B).

4.2 A particle's perspective

Theorem 4.2. *Let S be a switch and Y be particles as in Def. 4.1. Then for all $y \in \Omega_Y$,*

$$\mathbb{E}[S_y] = \frac{p(Y = y \mid S = \textit{on})\,\mathbb{E}[S]}{p(Y = y \mid S = \textit{on})\,\mathbb{E}[S] + p(Y = y \mid S = \textit{off})\,(1 - \mathbb{E}[S])}.$$

Furthermore, let $y* \in \Omega_Y$, then

$$p(Y = y* \mid S = \textit{on}) = p(Y = y* \mid S = \textit{off}) \text{ if and only if } \mathbb{E}[S_{y*}] = \mathbb{E}[S].$$

Proof. We can prove the first statement using Bayes' theorem.

$$\begin{aligned}\mathbb{E}[S_y] &:= 0 \times p(S = \textbf{off} \mid Y = y) + 1 \times p(S = \textbf{on} \mid Y = y) \\ &= \frac{p(S = \textbf{on}, Y = y)}{p(Y = y)},\end{aligned}$$

applying Bayes' Theorem to the numerator and denominator gives us

$$= \frac{p(Y = y \mid S = \textbf{on})\ p(S = \textbf{on})}{p(Y = y \mid S = \textbf{on})\ p(S = \textbf{on}) + p(Y = y \mid S = \textbf{off})\ p(S = \textbf{off})},$$

then simply substituting $p(S = \textbf{off}) = 1 - \mathbb{E}[S]$ gives us

$$= \frac{p(Y = y \mid S = \textbf{on})\,\mathbb{E}[S]}{p(Y = y \mid S = \textbf{on})\,\mathbb{E}[S] + p(Y = y \mid S = \textbf{off})\,(1 - \mathbb{E}[S])}.$$

The conditional expectation form of the switch conditioned on the particles seems to take a hyperbolic form. We now prove the second statement.

Case: ($\rightarrow$) Let $p(Y = y* \mid S = \textbf{on}) = p(Y = y* \mid S = \textbf{off})$. We begin with the result of the previous statement and then reduce:

$$\begin{aligned}\mathbb{E}[S_{y*}] &= \frac{p(Y = y* \mid S = \textbf{on})\,\mathbb{E}[S]}{p(Y = y* \mid S = \textbf{on})\,\mathbb{E}[S] + p(Y = y* \mid S = \textbf{off})\,(1 - \mathbb{E}[S])}, \\ &= \frac{p(Y = y* \mid S = \textbf{on})\,\mathbb{E}[S]}{p(Y = y* \mid S = \textbf{on})\,\mathbb{E}[S] + p(Y = y* \mid S = \textbf{on})\,(1 - \mathbb{E}[S])}, \\ &= \frac{p(Y = y* \mid S = \textbf{on})\,\mathbb{E}[S]}{p(Y = y* \mid S = \textbf{on})}, \\ &= \mathbb{E}[S].\end{aligned}$$

Case: ($\leftarrow$) Let $\mathbb{E}[S_{y*}] = \mathbb{E}[S]$.

$$\mathbb{E}[S_{y*}] = \frac{p(Y = y* \mid S = \textbf{on})\,\mathbb{E}[S]}{p(Y = y* \mid S = \textbf{on})\,\mathbb{E}[S] + p(Y = y* \mid S = \textbf{off})\,(1 - \mathbb{E}[S])},$$

dividing both sides by $\mathbb{E}[S]$ gives

$$1 = \frac{p(Y = y* \mid S = \textbf{on})}{p(Y = y* \mid S = \textbf{on})\,\mathbb{E}[S] + p(Y = y* \mid S = \textbf{off})\,(1 - \mathbb{E}[S])},$$

flipping the fraction and then multiplying top and bottom by the denominator gives

$$p(Y = y* \mid S = \textbf{on}) = p(Y = y* \mid S = \textbf{on})\,\mathbb{E}[S] + p(Y = y* \mid S = \textbf{off})\,(1 - \mathbb{E}[S]).$$

Finally, collating the like terms reduces the expression to

$$0 = (1 - \mathbb{E}[S])\,(p(Y = y* \mid S = \textbf{off}) - p(Y = y* \mid S = \textbf{on})).$$

Since $\mathbb{E}[S]$ cannot be one, we can conclude that $p(Y = y* \mid S = \textbf{off}) = p(Y = y* \mid S = \textbf{on})$. This completes the proof in both directions. □

Corollary 4.1. *If the conditional probabilities* $p(Y = y \mid S = \textbf{\textit{on}})$ *and* $p(Y = y \mid S = \textbf{\textit{off}})$ *are Poisson distributed, then*

$$\mathbb{E}[S_y] = \frac{(e^{-1}\lambda_1^y)^{\lambda_1}\,\mathbb{E}[S]}{(e^{-1}\,\lambda_1^y)^{\lambda_1}\,\mathbb{E}[S] + \left(e^{-\lambda_2/\lambda_1}\,\lambda_2^y\right)^{\lambda_1}(1-\mathbb{E}[S])},$$

and

$$y* = \left(\frac{\lambda_2}{\lambda_1} - 1\right)\frac{1}{\log(\lambda_1) - \log(\lambda_2)},$$

where $\lambda_1 := \mathbb{E}[Y_{\textbf{\textit{on}}}]$ *and* $\lambda_2 := \mathbb{E}[Y_{\textbf{\textit{off}}}]$.

Idea of Proof. The statement is proved by substituting the corresponding formulas of the Poisson distribution into Theorem 4.2 and then reducing. □

From Theorem 4.2, it is clear that the conditional expectation of the switch with respect to particles takes a hyperbolic form. In the case where the conditional probabilities are Poisson distributed, the conditional expectation takes the shape of a *sigmoid function* (see Fig. 3). Furthermore, like in the case of the conditional expectation of the particles with respect to the switch, if three out of four terms: $p(Y = y \mid S = \textbf{on})$, $p(Y = y \mid S = \textbf{off})$, $\mathbb{E}[S_y]$, and $\mathbb{E}[S]$ are known, then the fourth term can be reconstructed using Theorem 4.2.

In this section, we have proven some fundamental results regarding the conditional expectation forms from the perspective of the switch and that of the particles. We have shown that there is an inherent algebraic structure in a switch-particle coupled system. To gain further intuition into the theorems given in this section, we now introduce some simple numerical schemes which use these theorems. We will apply those schemes to some toy models to see how the conditional expectation forms behave.

5 Application

In the previous two sections we pursued a theoretical exploration in which we derived the relationship between classical moments and conditional expectations. From our theoretical results, a natural question arises: "Can the insights from §3 and §4 be used to help improve existing numerical methods used for simulating GRNs?" Answering such a question rigorously is beyond the scope of this work. However, as a preliminary step, we propose a new numerical solver which we call *ACE-Ansatz*; and a probability distribution reconstruction method based purely on moments, which we refer to as *ACE-Reconstruction*. We will introduce these methods and apply them to some simple examples. It is important to state that the aim is not to perform a comparative study, but rather to study examples which use the theorems derived in this paper, and through these examples gain further intuition into the structures between dimensions in GRNs.

5.1 ACE-Ansatz

We propose a new hybrid scheme which exploits the property that polynomial conditional expectation forms can be derived from classical moments. In essence, our new method is analogous to the *Method of Conditional Moments* (MCM), with an alternative method for deriving the conditional expectations. That is, in the MCM, the equations of motion of the conditional expectations are derived from the CME (§2.2), whereas in our new method, we will propose a polynomial ansatz for the derivation of the conditional expectations. Then, utilising Theorem 3.1, we will simply solve for the corresponding classical moments. We refer to this method as the *Algebraic Conditional Expectation Ansatz* method (ACE-Ansatz), which we will now compute on the well studied *Simple Gene Switch* model [12].

5.1.1 A Simple Gene Switch Model

The model describes a system which consists of a gene interacting with a well mixed pool of $mRNA$ and proteins. At any point in time only the following three variables of the system can be observed: the state of the gene, the population counts of $mRNA$, and the population counts of proteins. We denote this as the processes $X(t) := (G(t), M(t), A(t))$, where

- $G(t)$ has a binary state space $\{\textbf{on}, \textbf{off}\}$, describing the state of the gene at time t,
- $M(t)$ has a positive integer state space, describing the counts of $mRNA$ at time t,
- $A(t)$ has a positive integer state space, describing the counts of proteins at time t.

#	Reaction	Coefficient	Stoichiometry	Description
1	$\mathbf{off} \xrightarrow{\tau_{\mathbf{on}}} \mathbf{on}$	$\tau_{\mathbf{on}} = 0.1$	$(1, 0, 0)$	Basal activation
2	$\mathbf{on} \xrightarrow{\tau_{\mathbf{off}}} \mathbf{off}$	$\tau_{\mathbf{off}} = 0.05$	$(-1, 0, 0)$	Basal inactivation
3	$\mathbf{on} \xrightarrow{k_1} \mathbf{on} + mRNA$	$k_1 = 10.0$	$(0, 1, 0)$	Transcription
4	$mRNA \xrightarrow{\gamma_1} \emptyset$	$\gamma_1 = 1.0$	$(0, -1, 0)$	mRNA degradation
5	$mRNA \xrightarrow{k_2} mRNA + protein$	$k_2 = 4.0$	$(0, 0, 1)$	Translation
6	$protein \xrightarrow{\gamma_2} \emptyset$	$\gamma_2 = 0.5$	$(0, 0, -1)$	Protein degradation
7	$\mathbf{off} + Protein \xrightarrow{\hat{\tau}_{\mathbf{on}}} \mathbf{on}$	$\hat{\tau}_{\mathbf{on}} = 0.015$	$(1, 0, -1)$	Promoter activation

Table 1: Simple Gene Switch model reactions, propensities and stoichiometries. The system is initialised at $G(0) = \mathbf{off}$, $M(0) = 8$, and $A(0) = 80$.

The system can undergo seven reactions which alter its state (see Table 1). Verbosely, reactions one and two describe the background basal switching of the gene in the system. Reaction three describes the transcription process, where the gene in the "**on**" state starts producing mRNA. Reaction five describes the translation process, where the $mRNA$ is translated to produce a protein. Reaction four and six describe the degradation of the $mRNA$ and proteins respectively. Reaction seven describes the activation of the promotor region of the gene by the protein.

5.1.2 Linear ACE-Ansatz Approximation

We will use Theorem 3.1 to perform dimension reduction on the simple gene switch model. In previous literature, where the simple gene switch model was introduced for dimension reduction, the authors demonstrated that the marginal distribution of the gene and the proteins could be well approximated by numerous dimension reduction schemes [29, 39, 12, 14]. We will keep the same setting as in the previous literature to help contrast the application of our theorems. We will first derive the derivative of the marginal distribution of genes and proteins. Then, as in (2.10), we will highlight the conditional expectation terms needed to solve for the marginal distribution. We then apply a linear form ansatz to the conditional expectations. We know from §3.1, that the coefficients of the linear form are given by the first and second moments. Hence, we will derive the equations for the moments up to degree two and then close the higher order terms.

We begin by deriving the derivative of the marginal distribution of G and A with respect to time. The steps from the CME to the marginal distribution are given in Appendix C. For a fixed $a \in \Omega_A$, the following two equations describe the marginal distributions for the states **off** and **on**, respectively:

$$\begin{aligned}\frac{d\,p(G=\mathbf{off}, A=a;t)}{dt} &= \tau_{\mathbf{off}}\; p(G=\mathbf{on}, A=a;t) \\ &\quad + k_2\, \underbrace{\mathbb{E}[M_{\mathbf{off},a-1}(t)]}_{(*)}\; p(G=\mathbf{off}, A=a-1;t) \\ &\quad + \gamma_2\,(a+1)\;\; p(G=\mathbf{off}, A=a+1;t) \\ &\quad - \left(\tau_{\mathbf{on}} + k_2\, \underbrace{\mathbb{E}[M_{\mathbf{off},a}(t)]}_{(*)} + (\gamma_2 + \hat{\tau}_{\mathbf{on}})\, a\right)\; p(G=\mathbf{off}, A=a;t),\end{aligned} \tag{5.1}$$

$$\begin{aligned}\frac{d\,p(G=\mathbf{on}, A=a;t)}{dt} &= \tau_{\mathbf{on}}\; p(G=\mathbf{off}, A=a;t) \\ &\quad + k_2\, \underbrace{\mathbb{E}[M_{\mathbf{on},a-1}(t)]}_{(*)}\; p(G=\mathbf{on}, A=a-1;t) \\ &\quad + \gamma_2\,(a+1)\;\; p(G=\mathbf{on}, A=a+1;t) \\ &\quad + \hat{\tau}_{\mathbf{on}}\,(a+1)\;\; p(G=\mathbf{off}, A=a+1;t) \\ &\quad - \left(\tau_{\mathbf{off}} + k_2\, \underbrace{\mathbb{E}[M_{\mathbf{on},a}(t)]}_{(*)} + \gamma_2\, a\right)\; p(G=\mathbf{on}, A=a;t).\end{aligned} \tag{5.2}$$

We see in the equations above that to solve for the marginal distribution, the terms marked by $(*)$, that is, the expectation of the $mRNA$ counts conditioned on the protein count a and the gene state **on** or **off**, need

to be estimated. We approximate the conditional expectation by a linear conditional expectation form from Lemma 3.1. That is, for $g \in \{\mathbf{off}: 0, \mathbf{on}: 1\}$ and $a \in \mathbb{Z}_+$, with

$$\alpha := \begin{bmatrix} \underbrace{\mathrm{cov}(G(t), M(t))}_{(**)} & \underbrace{\mathrm{cov}(A(t), M(t))}_{(**)} \end{bmatrix} \begin{pmatrix} \mathrm{cov}(G(t), G(t)) & \mathrm{cov}(G(t), A(t)) \\ \mathrm{cov}(G(t), A(t)) & \mathrm{cov}(A(t), A(t)) \end{pmatrix}^{-1},$$

we approximate,

$$\mathbb{E}[M_{g,a}(t)] \approx \alpha \left(\begin{bmatrix} g \\ a \end{bmatrix} - \begin{bmatrix} \mathbb{E}[G(t)] \\ \mathbb{E}[A(t)] \end{bmatrix} \right) + \underbrace{\mathbb{E}[M(t)]}_{(**)} \tag{5.3}$$

The terms $\mathrm{cov}(G(t), A(t))$, $\mathrm{cov}(G(t), G(t))$, $\mathrm{cov}(A(t), A(t))$, $\mathbb{E}[A(t)]$, and $\mathbb{E}[G(t)]$, can be computed using the marginal distribution $p(G = \cdot, A = \cdot; t)$. We note that after substituting the conditional expectation by a linear form in (5.3), all the terms we solve for in this section become approximations. However, for brevity we keep the same notation. A formal notational derivation is given in Appendix D. We now estimate the terms marked by $(**)$, their time derivatives were derived using [9, Lemma 2.1] or [33, Equation 11]:

$$\frac{d\,\mathbb{E}[M(t)]}{dt} = k_1\,\mathbb{E}[G(t)] - \gamma_1\,\mathbb{E}[M(t)]. \tag{5.4}$$

$$\begin{aligned} \frac{d\,\mathbb{E}[G\,M(t)]}{dt} &= \tau_{\mathbf{on}}\,(-\mathbb{E}[G\,M(t)] - \mathbb{E}[M(t)]) - \tau_{\mathbf{off}}\,\mathbb{E}[G\,M(t)] + k_1\,\mathbb{E}[G(t)] \\ &\quad - \gamma_1\,\mathbb{E}[G\,M(t)] + \hat{\tau}_{\mathbf{on}} \left(\mathbb{E}[M\,A(t)] - \underbrace{\mathbb{E}[G\,M\,A(t)]}_{(***)} \right). \end{aligned} \tag{5.5}$$

$$\begin{aligned} \frac{d\,\mathbb{E}[M\,A(t)]}{dt} &= k_1\,\mathbb{E}[G\,A(t)] - (\gamma_1 + \gamma_2)\,\mathbb{E}[M\,A(t)] + k_2\,\mathbb{E}[M^2(t)] \\ &\quad - \hat{\tau}_{\mathbf{on}} \left(\mathbb{E}[M\,A(t)] - \underbrace{\mathbb{E}[G\,M\,A(t)]}_{(***)} \right) \end{aligned} \tag{5.6}$$

$$\frac{d\,\mathbb{E}[M^2(t)]}{dt} = k_1\,\left(2\,\mathbb{E}[G\,M(t)] + \mathbb{E}[G(t)]\right) + \gamma_1\,\left(-2\,\mathbb{E}[M^2(t)] + \mathbb{E}[M(t)]\right). \tag{5.7}$$

To close the equations above, an estimate for the term $\mathbb{E}[G\,M\,A(t)]$, $(***)$, is needed. Hence, we construct a closure using the equations already known:

$$\begin{aligned} \mathbb{E}[G\,M\,A(t)] &= \sum_{a,m \in \mathbb{Z}_+} a\,m\;p(G = \mathbf{on}, M = m, A = a; t), \\ &= \sum_{a,m \in \mathbb{Z}_+} a\,m\;p(M = m \,|\, G = \mathbf{on}, A = a; t)\;p(G = \mathbf{on}, A = a; t), \\ &= \sum_{a \in \mathbb{Z}_+} \left(\sum_{m \in \mathbb{Z}_+} m\;p(M = m \,|\, G = \mathbf{on}, A = a; t) \right) a\;p(G = \mathbf{on}, A = a; t), \\ &= \sum_{a \in \mathbb{Z}_+} \underbrace{\mathbb{E}[M_{\mathbf{on},a}(t)]}_{(5.3)}\;a\;p(G = \mathbf{on}, A = a; t). \end{aligned} \tag{5.8}$$

Equations (5.1) to (5.8) form a closed system of equations. Since the ansatz was linear, we refer to the approximation found by solving these equations as a Linear ACE-Ansatz approximation. Note that approximations with a higher degree ansatz can also be constructed by simply increasing the number of moment equations. We will compare the quality of the Linear ACE-Ansatz approximation to the approximation constructed by the MCM method described in [14]. The joint distribution of the simple switch model can be solved using the *Optimal Finite State Projection method* (OFSP) [3]. We will use the OFSP approximation as a reference distribution for comparing both approximation methods. We see in Fig. 4B-D, that the ACE-Ansatz's marginal distribution approximation is fairly accurate in capturing the shape and bi-modality of the reference distribution. The MCM

[3]The `PyME` implementation of the OFSP method and the MCM were used in this work [36, 38]. It must noted that the MCM module in `PyME` is not optimised for speed. All code was run on an Intel i7 2.5GHz with 16GB of RAM.

method captures the bi-modalitiy but does not capture the shape of the marginal distribution in the **on** state. The linear ACE-Ansatz only used five equations to estimate the conditional expectations, whereas in the MCM, 277 equations were needed (see Table 2). Even though using more equations did not provide gains in the quality of the marginal distribution approximation, we see in Fig. 4A-C, that the MCM approximates the expectation of $mRNA$ and genes better than the ACE-Ansatz. Nevertheless, utilising the structure of a polynomial ansatz, we were able to construct an approximation which was nearly as accurate as that of the MCM; using only five equations.

Method	Num Equations	Error in Dist.	Error in Moments	Comp Time
ACE	344–5	0.0396	$(0.014, 0.134, 1.135)$	4 sec
MCM	277–277	0.526	$(0.005, 0.043, 0.267)$	156 sec
OFSP	39876–0	–	–	62 sec

Table 2: Performance comparison between the ACE-Ansatz and the MCM at $t = 10$. The error in distribution is the ℓ_1 error between the respective approximation and the OFSP solution. The error in moments is the ℓ_1 error between the respective moment approximation and the OFSP solution's moments. The entries in every line show errors in the approximation of the species (G, M, A), respectively.

In summary, different polynomial ansätze can be used to approximate the conditional expectation structures between dimensions. As an example, we performed dimension reduction using the linear ansatz. By applying Theorem 3.1, we could observe that the equations of motion needed to solve for the conditional expectation are simply the classical moment equations. A future research direction could be to investigate whether the structures we see in the polynomial forms can be extended to general basis functions, like radial basis functions, wavelets, etc.

5.2 ACE-Reconstruction

In the previous section, we demonstrated that the ACE-Ansatz could be used to accurately approximate the marginal distributions. We were able to demonstrate that the linear polynomial forms were sufficient approximations of the conditional expectations. This leads to the next question: "Can the ACE-Ansatz be used to estimate the conditional variance? If so, can both the conditional expectation and the conditional variance be used to reconstruct/approximate the conditional probability?" While this question warrants its own paper, we will give some preliminary insights using the theorems already established in this paper.

In this subsection, we will reconstruct two 2D distributions, both mono-modal with non-zero covariance, only using the marginal distribution and moments of the joint distribution.

5.2.1 Linear ACE-Reconstruction

Let us consider the distribution in Fig. 5A. We see that the distribution is mono-modal and has non-zero covariance. We continue referring to the dimension which is conditioned on as X and to the dimension being conditioned as Y. We see in Fig. 5B, that $\mathbb{E}[Y_x]$ has a linear form, and Fig. 5C shows that $\mathbb{E}[Y_x^2]$ has a quadratic form. We first compute the linear and quadratic polynomial approximations for $\mathbb{E}[Y_x]$ and $\mathbb{E}[Y_x^2]$, respectively (see Def. 3.1). We begin by solving for the coefficients of the linear approximation of the conditional expectation,

Moments of X	Moments of Y	Moments of mixed X and Y
$\mathbb{E}[X] = 33.25$	$\mathbb{E}[Y] = 94.26$	$\mathbb{E}[XY] = 3203.24$
$\mathbb{E}[X^2] = 1133.69$	$\mathbb{E}[Y^2] = 9211.15$	$\mathbb{E}[XY^2] = 318830.26$
$\mathbb{E}[X^3] = 39573.27$		$\mathbb{E}[X^2Y^2] = 1129342.88$
$\mathbb{E}[X^4] = 1412921.52$		

Table 3: Moments of the distribution in Fig. 5A.

$$\eta^1_{Y|X}(x) = \kappa_{11}\, x + \kappa_{10},$$

where the coefficients are found by solving

$$\begin{bmatrix} \mathbb{E}[X] & 1 \\ \mathbb{E}[X^2] & \mathbb{E}[X] \end{bmatrix} \bullet \begin{bmatrix} \kappa_{11} \\ \kappa_{10} \end{bmatrix} = \begin{bmatrix} \mathbb{E}[Y] \\ \mathbb{E}[XY] \end{bmatrix}.$$

Substituting in the terms from Table 3 and solving the above linear system of equations gives that: $\kappa_{11} = 2.343$ and $\kappa_{10} = 16.462$. For the expectation of Y^2 conditioned on X, we can use the same machinery. Let us now consider a quadratic algebraic form approximation:

$$\eta^2_{Y^2|X}(x) = \kappa_{22}\, x^2 + \kappa_{21}\, x + \kappa_{20},$$

where the coefficients are found by solving

$$\begin{bmatrix} \mathbb{E}[X^2] & \mathbb{E}[X] & 1 \\ \mathbb{E}[X^3] & \mathbb{E}[X^2] & \mathbb{E}[X] \\ \mathbb{E}[X^4] & \mathbb{E}[X^3] & \mathbb{E}[X^2] \end{bmatrix} \bullet \begin{bmatrix} \kappa_{22} \\ \kappa_{21} \\ \kappa_{20} \end{bmatrix} = \begin{bmatrix} \mathbb{E}[Y^2] \\ \mathbb{E}[XY^2] \\ \mathbb{E}[X^2Y^2] \end{bmatrix}.$$

Substituting in the terms from Table 3 and solving the above linear system of equations gives: $\kappa_{22} = 5.210, \kappa_{21} = 99.111$ and $\kappa_{10} = 9.251$. We have just derived an algebraic form for the first and second conditional moments. Hence, for each $x \in \Omega_X$, we have approximations of the first two moments of conditional probability $p(Y = \cdot \,|\, X = x)$. Now, we want to fit a distribution which approximates the conditional probability. To do this, we use the method of *Maximal Entropy*, which fits a distribution to some prescribed set of moments, such that the fitted distribution has maximal entropy among all candidate distributions with matching moments. (for further reading on the Maximal Entropy method please see [34]). In our case, we wish to fit only the first two moments. We denote the fitting as follows:

$$p(Y = y \,|X = x) \approx \Pi(\eta^1_{Y|X}(x), \eta^2_{Y^2|X}(x))(Y = y). \tag{5.9}$$

Then the approximation of the joint distribution using the conditional moment approximations is given by

$$p(X = x, Y = y) \approx \Pi(\eta^1_{Y|X}(x), \eta^2_{Y^2|X}(x))(Y = y)\ p(X = x). \tag{5.10}$$

If we compare the true joint distribution to the ACE-Reconstruction distribution (see Fig. 6), we see that the covariance structure and the range of probabilities are similar in both cases. In Fig. 6C, we see that the ACE-Reconstruction has maximal error near the mode of the distribution and it captures the tails of the distribution well. For comparison, we also fitted a Gaussian distribution to the mean and covariance of the original distribution [4] (see Fig. 6C). We can see that the Gaussian also captures the covariance structure and the probability range well. However, in contrast to the ACE-Reconstruction, the Gaussian reconstruction captures the mode of the original distribution well and loses accuracy in the tails. We see in Table 4, that for this example, the gain in computational accuracy of the ACE-Reconstruction over a Gaussian reconstruction does not merit the increased computational effort of the ACE-Reconstruction over the Gaussian reconstruction.

Method	Number of Moments	Error in Distribution
ACE	9	0.0497
Gaussian	5	0.0787

Table 4: Comparison between ACE-Reconstruction and Gaussian reconstruction. Error in distribution refers to the ℓ_1 error between the respective reconstruction and the true distribution.

5.2.2 Higher Degree ACE-Reconstruction

We saw in the simple linear case that ACE could capture the distribution very accurately. However, a simple Gaussian reconstruction was also comparably accurate and significantly cheaper computationally. Now we will consider a slightly more complex example. The distribution in Fig. 7A commonly occurs when the system's dynamics moves its distribution into the edge of the state space. In this particular case, the distribution was taken from an SIR model, at a time point where nearly all susceptible individuals are converted into infected individuals (see Appendix E for system parameters). We follow similar steps as in the linear example case, but extend a bit further by constructing a cubic polynomial– and a quartic polynomial approximation of the conditional moments. We start by writing down the moments which are necessary to approximate the conditional expectations. For a quartic polynomial form, moments up to degree eight are needed (see Table 5).

Moments of X		Moments of Y	Moments of mixed X and Y	
$\mathbb{E}[X] = 6.28$	$\mathbb{E}[X^5] = 38864736.15$	$\mathbb{E}[Y] = 131.17$	$\mathbb{E}[XY] = 821.69$	$\mathbb{E}[X^2Y^2] = 1141745.56$
$\mathbb{E}[X^2] = 73.09$	$\mathbb{E}[X^6] = 6861261393.8$	$\mathbb{E}[Y^2] = 17277.97$	$\mathbb{E}[X^2Y] = 8815.75$	$\mathbb{E}[X^3Y^2] = 21874741.71$
$\mathbb{E}[X^3] = 2453.18$	$\mathbb{E}[X^7] \approx 126.94 \times 10^{10}$		$\mathbb{E}[X^3Y] = 187486.58$	$\mathbb{E}[X^4Y^2] = 843134378.72$
$\mathbb{E}[X^4] = 248985.45$	$\mathbb{E}[X^8] \approx 240.45 \times 10^{12}$		$\mathbb{E}[XY^2] = 108453.02$	

Table 5: Moments of the distribution in Fig. 7A.

We now consider cubic and quartic polynomial approximations of the first two conditional moments, respectively:

$$\eta^1_{Y|X}(x) = \kappa_{13}\, x^3 + \kappa_{12}\, x^2 + \kappa_{11}\, x + \kappa_{10}, \tag{5.11}$$

[4] A Gaussian reconstruction in this context involves computing the Gaussian distribution over the discrete state space and then normalising to make the total mass one.

$$\eta^2_{Y^2|X}(x) = \kappa_{24}\, x^4 + \kappa_{23}\, x^3 + \kappa_{22}\, x^2 + \kappa_{21}\, x + \kappa_{20}. \tag{5.12}$$

The coefficients in the equations above can be found by solving the following system of equations:

$$\begin{bmatrix} \mathbb{E}[X^3] & \mathbb{E}[X^2] & \mathbb{E}[X] & 1 \\ \mathbb{E}[X^4] & \mathbb{E}[X^3] & \mathbb{E}[X^2] & \mathbb{E}[X] \\ \mathbb{E}[X^5] & \mathbb{E}[X^4] & \mathbb{E}[X^3] & \mathbb{E}[X^2] \\ \mathbb{E}[X^6] & \mathbb{E}[X^5] & \mathbb{E}[X^4] & \mathbb{E}[X^3] \end{bmatrix} \bullet \begin{bmatrix} \kappa_{13} \\ \kappa_{12} \\ \kappa_{11} \\ \kappa_{10} \end{bmatrix} = \begin{bmatrix} \mathbb{E}[Y] \\ \mathbb{E}[XY] \\ \mathbb{E}[X^2Y] \\ \mathbb{E}[X^3Y] \end{bmatrix},$$

and

$$\begin{bmatrix} \mathbb{E}[X^4] & \mathbb{E}[X^3] & \mathbb{E}[X^2] & \mathbb{E}[X] & 1 \\ \mathbb{E}[X^5] & \mathbb{E}[X^4] & \mathbb{E}[X^3] & \mathbb{E}[X^2] & \mathbb{E}[X] \\ \mathbb{E}[X^6] & \mathbb{E}[X^5] & \mathbb{E}[X^4] & \mathbb{E}[X^3] & \mathbb{E}[X^2] \\ \mathbb{E}[X^7] & \mathbb{E}[X^6] & \mathbb{E}[X^5] & \mathbb{E}[X^4] & \mathbb{E}[X^3] \\ \mathbb{E}[X^8] & \mathbb{E}[X^7] & \mathbb{E}[X^6] & \mathbb{E}[X^5] & \mathbb{E}[X^4] \end{bmatrix} \bullet \begin{bmatrix} \kappa_{24} \\ \kappa_{23} \\ \kappa_{22} \\ \kappa_{21} \\ \kappa_{20} \end{bmatrix} = \begin{bmatrix} \mathbb{E}[Y^2] \\ \mathbb{E}[XY^2] \\ \mathbb{E}[X^2Y^2] \\ \mathbb{E}[X^3Y^2] \\ \mathbb{E}[X^4Y^2] \end{bmatrix}.$$

We see in Fig. 7 that the conditional moment approximations do not fit as tightly as in the previous example, however, they do capture the right trend.

Like in the linear case, we use the Maximum Entropy method to approximate the conditional probability using the first two conditional moment approximations given in (5.11) and (5.12),

$$p(Y = y \mid X = x) \approx \Pi(\eta^1_{Y|X}(x), \eta^2_{Y^2|X}(x))(Y = y). \tag{5.13}$$

Then the reconstruction of the joint distribution is given by

$$p(X = x, Y = y) \approx \Pi(\eta^1_{Y|X}(x), \eta^2_{Y^2|X}(x))(Y = y)\ p(X = x). \tag{5.14}$$

We see in Fig. 8 that the ACE-Reconstruction performs much better than the Gaussian reconstruction. This is illustrated in Table 6, which shows that the expected error of the ACE-Reconstruction is approximately twelve times smaller than that of the Gaussian reconstruction. In general, distributions close to boundaries are difficult to compute, because the boundary forces the distributions to bend. This makes their approximation computationally challenging. The ACE-Reconstruction, which didn't give a perfect fit, nevertheless captured some of these dynamics by approximating the underlying conditional moments. It must be noted that the Gaussian reconstruction was done with far fewer moments than the ACE-Reconstruction, so it is not a fair comparison to only look at the shape. Nevertheless, the aim of this example was to demonstrate that the reconstruction of complex distribution shapes is possible; given that there is a smooth underlying manifold in which the conditional moments are embedded.

Method	Number of Moments	Error in Distribution
ACE	16	0.0378
Gaussian	5	0.4599

Table 6: Comparison between ACE-Reconstruction and Gaussian reconstruction. Error in distribution refers to the ℓ_1 error between the respective reconstruction and the true distribution.

We have given some preliminary evidence that distributions can be reconstructed using the conditional moments. However, there is much scope to improve this approach. For example, when the first and second conditional moments are approximated separately, two independent linear systems of equations are solved. While this is computationally simple, in many cases the positivity condition of the variance could be violated due to fitting errors. A better strategy which fixes this problem would be to set up a non-linear system of equations using Eve's Law. That is, we would trade a linear system of equations for a non-linear system of equations to preserve interpretability of the approximation.

6 Discussion

When presented with a high dimensional GRN, we are quick to reach for dimension reduction methods focused around principles of time scale separation and volume size expansion. These methods are effective at exploiting particular structures in the network, but are not an overarching framework to help decompose and understand GRNs. In this paper, we introduced an algebraic framework to describe the relationship between species in a general Kurtz process, and showed that conditional expectations are the key to understanding these relationships. We then proved that if the conditional expectation has an algebraic form, then the form can be inferred from the classical moments. In short, conditional expectations decompose the dimensional relationships,

and the moments decompose the conditional expectations, elucidating that the moments contain all critical information about the network. We then proved that GRNs inherently have algebraic forms between dimensions. Hence, one can translate the theory which we have developed to any general GRN.

To show that there are potential applications for the theorems developed in this work, we touched on two new methods: one to simulate/evolve GRNs using a polynomial conditional expectation ansatz; and one to reconstruct complex distribution shapes using conditional moments. Both cases gave positive preliminary results in favour of developing new numerical schemes using an algebraic conditional expectation ansatz.

This algebraic line of investigation can be extended into many aspects of GRNs. For example, we could investigate mappings which project the network onto a domain which yields lower degree conditional expectation forms (like the concept of linearisation in numerics). Conditional expectation could also be applied in model selection, where similarity metrics can be designed using the algebraic forms of the conditional expectations between dimensions.

In summary, conditional expectations are critical for understanding and decomposing GRNs. We proved that the algebraic perspective is a robust and intuitive framework for studying such networks. Future research down this line of thought is imperative.

References

[1] D. 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.

[2] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *The Annals of Applied Probability*, 16(4):1925–1961, 2006.

[3] J. Banasiak. *Positive semigroups with applications*. PhD thesis, University of KwaZulu-Natal, Durban, South Africa, 2014.

[4] N. Barkai and S. Leibler. Biological rhythms: Circadian clocks limited by noise. *Nature*, 403(6767):267–268, jan 2000.

[5] W. J. Blake, M. KÆrn, C. R. Cantor, and J. J. Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633–637, apr 2003.

[6] K. Burrage, S. MacNamara, and T. H. Tian. Accelerated leap methods for simulating discrete stochastic chemical kinetics. *Positive Systems, Proceedings*, 341:359–366, 2006.

[7] L. Cardelli, M. Kwiatkowska, and L. Laurenti. Stochastic analysis of Chemical Reaction Networks using Linear Noise Approximation. *BioSystems*, 149:26–33, 2016.

[8] K. Choudhary, S. Oehler, and A. Narang. Protein distributions from a stochastic model of the lac operon of E. coli with DNA looping: Analytical solution and comparison with experiments. *PLoS ONE*, 9(7), 2014.

[9] S. Engblom. Computing the moments of high dimensional solutions of the master equation. *Applied Mathematics and Computation*, 180(2):498–515, sep 2006.

[10] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, dec 1977.

[11] J. Goutsias. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *The Journal of Chemical Physics*, 122(18):184102, 2005.

[12] R. Grima, D. R. Schmidt, and T. J. Newman. Steady-state fluctuations of a genetic feedback loop: An exact solution. *Journal of Chemical Physics*, 137(3), 2012.

[13] E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *Journal of Chemical Physics*, 117(15):6959–6969, oct 2002.

[14] J. Hasenauer, V. Wolf, A. Kazeroonian, and F. J. Theis. Method of conditional moments (MCM) for the Chemical Master Equation. *Journal of Mathematical Biology*, pages 1–49, 2013.

[15] A. Hellander and P. Lötstedt. Hybrid method for the chemical master equation. *Journal of Computational Physics*, 227(1):100–122, 2007.

[16] T. A. Henzinger, L. Mikeev, M. Mateescu, and V. Wolf. Hybrid numerical solution of the chemical master equation. In *Proceedings of the 8th International Conference on Computational Methods in Systems Biology*, pages 55–65, Trento, Italy, 2010. ACM.

[17] D. J. Higham. Modeling and Simulating Chemical Reactions. *SIAM Review*, 50(2):347–368, 2008.

[18] T. Jahnke. On Reduced Models for the Chemical Master Equation. *Multiscale Modeling & Simulation*, 9(4):1646–1676, 2011.

[19] T. Jahnke and W. Huisinga. Solving the Chemical Master Equation For Monomolecular Reaction systems analytically. *Journal of Mathematical Biology*, 54:1–26, 2007.

[20] T. Jahnke and M. Kreim. Error bound for piecewise deterministic processes modeling stochastic reaction systems. *SIAM Multiscale Modeling and Simulation*, 10(4):1119–1147, 2012.

[21] T. Jahnke and V. Sunkara. Error Bound for Hybrid Models of Two-Scaled Stochastic Reaction Systems. In *Extraction of Quantifiable Information from Complex Systems*, volume 102 of *Lecture Notes in Computational Science and Engineering*, pages 303–319. Springer International Publishing, 2014.

[22] G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10):770–780, oct 2008.

[23] M. Khammash and B. Munsky. The finite state projection algorithm for the solution of the chemical master equation. *Journal of Chemical Physics*, 124(044104):1–12, 2006.

[24] T. G. Kurtz. Relationship Between Stochastic and Deterministic Models for Chemical Reactions. *J. Chem. Phys.*, 57(7):2976–2978, 1972.

[25] B. D. MacArthur, A. Ma'ayan, and I. R. Lemischka. Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol*, 10(10):672–681, oct 2009.

[26] S. MacNamara, A. M. Bersani, K. Burrage, and R. B. Sidje. Stochastic chemical kinetics and the total quasi-steady-state assumption: application to the stochastic simulation algorithm and chemical master equation. *Journal of Chemical Physics*, 129(095105):1–13, 2008.

[27] S. Menz, J. Latorre, C. Schütte, and W. Huisinga. Hybrid Stochastic–Deterministic Solution of the Chemical Master Equation. *Multiscale Modeling & Simulation*, 10(4):1232–1262, 2012.

[28] W. Nagel and R. Steyer. *Probability and Conditional Expectation*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd, Oxford, UK, apr 2017.

[29] M. Pájaro, A. A. Alonso, I. Otero-Muras, and C. Vázquez. Stochastic modeling and numerical simulation of gene regulatory networks with protein bursting. *Journal of Theoretical Biology*, 421:51–70, 2017.

[30] C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *The Journal of Chemical Physics*, 118(11):4999–5010, 2003.

[31] G. A. F. Seber and A. J. Lee. *Linear Regression Analysis*. John Wiley & Sons, Inc., Hoboken, NJ, USA, jan 2003.

[32] A. Singh and J. P. Hespanha. Models for multi-specie chemical reactions using polynomial stochastic hybrid systems. *Decision and Control, 2005 and 2005 ...*, pages 2969–2974, 2005.

[33] P. Smadbeck and Y. N. Kaznessis. Efficient moment matrix generation for arbitrary chemical networks. *Chemical Engineering Science*, 84:612–618, 2012.

[34] P. Smadbeck and Y. N. Kaznessis. A closure scheme for chemical master equations. *Proceedings of the National Academy of Sciences*, 110(35):14261–14265, 2013.

[35] R. Srinivastiv, L. You, J. Summers, and J. Yin. Stochastic vs. Deterministic Modeling of Intracellular Viral Kinetics. *Journal of Theoretical Biology*, 218(3):309–321, 2002.

[36] V. Sunkara. PyME (Python Solver for The Chemical Master Equation).

[37] V. Sunkara. *Analysis and Numerics of the Chemical Master Equation*. PhD thesis, Australian National University, 2013.

[38] V. Sunkara and M. Hegland. An optimal Finite State Projection Method. *Procedia Computer Science*, 1(1):1579–1586, may 2010.

[39] P. Thomas, N. Popovi , and R. Grima. Phenotypic switching in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 111(19):6994–6999, 2014.

[40] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland, 3 edition, 2007.

[41] J. M. G. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9):5988–5992, 2002.

[42] D. J. Wilkinson. *Stochastic modelling for systems biology*. Mathematical and computational biology series. Chapman & Hall / CRC, Boca Raton, Fla. [u.a.], 2006.
99

A Proofs

Proof of Lemma 3.1-3. We substitute the conditional expectation form into Eve's law (Law of Total Variance) and then reduce.

Eve's Law states that

$$\operatorname{cov}(Y, Y) = \mathbb{E}[\operatorname{cov}(Y_x, Y_x)] + \operatorname{cov}(\mathbb{E}[Y_x], \mathbb{E}[Y_x]).$$

Verbosely, the total variation of Y is the sum of the expectation of the conditional variances and the variance of the conditional expectation. We begin by reducing the covariance of the conditional expectations:

$$\operatorname{cov}(\mathbb{E}[Y_x], \mathbb{E}[Y_x]) := \sum_{x \in \Omega_X} \left[(\mathbb{E}[Y_x] - \mathbb{E}[Y]) (\mathbb{E}[Y_x] - \mathbb{E}[Y])^T \right] p(X = x),$$

substituting the linear conditional expectation form and the expanding gives us

$$\begin{aligned}
&= \sum_{x \in \Omega_X} \left[(\alpha\, x + \beta - \mathbb{E}[Y]) (\alpha\, x + \beta - \mathbb{E}[Y])^T \right] p(X = x), \\
&= \sum_{x \in \Omega_X} \left[(\alpha\, x + \mathbb{E}[Y] - \alpha\, \mathbb{E}[X] - \mathbb{E}[Y]) (\alpha\, x + \mathbb{E}[Y] - \alpha\, \mathbb{E}[X] - \mathbb{E}[Y])^T \right] p(X = x), \\
&= \sum_{x \in \Omega_X} \left[(\alpha\, x - \alpha\, \mathbb{E}[X]) (\alpha\, x - \alpha\, \mathbb{E}[X])^T \right] p(X = x), \\
&= \sum_{x \in \Omega_X} \alpha \left[(x - \mathbb{E}[X]) (x - \mathbb{E}[X])^T \right] \alpha^T\, p(X = x), \\
&= \alpha \left[\sum_{x \in \Omega_X} (x - \mathbb{E}[X]) (x - \mathbb{E}[X])^T\, p(X = x) \right] \alpha^T,
\end{aligned}$$

substituting the definition of a covariance gives

$$= \alpha \operatorname{cov}(X, X)\, \alpha^T.$$

Substituting this term above into Eve's law gives us that,

$$\mathbb{E}[\operatorname{cov}(Y_x, Y_x)] = \operatorname{cov}(Y, Y) - \alpha \operatorname{cov}(X, X)\, \alpha^T.$$

□

B Parameters of the three models

#	Reaction	Coefficient	Stoichiometry	Description
1	$\emptyset \xrightarrow{c_1} X$	$c_1 = 50.0$	$(1, 0)$	Birth of X
2	$X \xrightarrow{c_2} \emptyset$	$c_2 = 1.0$	$(-1, 0)$	Death of X
3	$\emptyset \xrightarrow{c_3} Y$	$c_3 = 20.0$	$(0, 1)$	Birth of Y
4	$Y \xrightarrow{c_4} \emptyset$	$c_4 = 1.0$	$(0, -1)$	Death of Y

Table 7: **Model 1** system parameters

#	Reaction	Coefficient	Stoichiometry	Description
1	$\emptyset \xrightarrow{c_1} X$	$c_1 = 10.0$	$(1, 0)$	Birth of X
2	$X \xrightarrow{c_2} \emptyset$	$c_2 = 1.0$	$(-1, 0)$	Death of X
3	$X \xrightarrow{c_3} Y + X$	$c_3 = 4.0$	$(0, 1)$	Autocatalytic production of Y using X
4	$Y \xrightarrow{c_4} \emptyset$	$c_4 = 0.5$	$(0, -1)$	Death of Y

Table 8: **Model 2** system parameters

#	Reaction	Coefficient	Stoichiometry	Description
1	$\emptyset \xrightarrow{c_1} X$	$c_1 = 50.0$	$(1, 0)$	Birth of X
2	$X \xrightarrow{c_2} \emptyset$	$c_2 = 0.01$	$(-1, 0)$	Death of X
3	$2X \xrightarrow{c_3} Y + 2X$	$c_3 = 0.4$	$(0, 1)$	Autocatalytic production of Y using two X
4	$Y \xrightarrow{c_4} \emptyset$	$c_4 = 1.0$	$(0, -1)$	Death of Y

Table 9: **Model 3** system parameters

C Simple Gene Switch Derivations

C.1 Chemical Master Equation

$$\begin{aligned}\frac{d\,p(G=\mathbf{off},M=m,A=a;t)}{dt} &= \tau_{\mathbf{off}}\,p(G=\mathbf{on},M=m,A=a;t)\\ &+\gamma_1\,(m+1)\,p(G=\mathbf{off},M=m+1,A=a;t)\\ &+\kappa_2\,m\,p(G=\mathbf{off},M=m,A=a-1;t)\\ &+\gamma_2\,(a+1)\,p(G=\mathbf{off},M=m,A=a+1;t)\\ &-\left[\tau_{\mathbf{on}}+(\gamma_1+\kappa_2)\,m+(\gamma_2+\hat{\tau}_{\mathbf{on}})\,a\right]p(G=\mathbf{off},M=m,A=a;t).\end{aligned}$$

$$\begin{aligned}\frac{d\,p(G=\mathbf{on},M=m,A=a;t)}{dt} &= \tau_{\mathbf{on}}\,p(G=\mathbf{off},M=m,A=a;t)\\ &+\kappa_1\,p(G=\mathbf{on},M=m-1,A=a;t)\\ &+\gamma_1\,(m+1)\,p(G=\mathbf{on},M=m+1,A=a;t)\\ &+\kappa_2\,m\,p(G=\mathbf{on},M=m,A=a-1;t)\\ &+\gamma_2\,(a+1)\,p(G=\mathbf{on},M=m,A=a+1;t)\\ &+\hat{\tau}_{\mathbf{on}}\,(a+1)\,p(G=\mathbf{off},M=m,A=a+1;t)\\ &-\left\{\tau_{\mathbf{off}}+\kappa_1+(\gamma_1+\kappa_2)\,m+\gamma_2\,a\right\}p(G=\mathbf{on},M=m,A=a;t)\end{aligned}$$

C.2 Marginal Distributions

We follow the same steps as in the generalised form (see §2.2). Deriving the CME for the marginal distribution of the gene and the proteins involves the following two steps:

- substituting $p(G=\cdot,M=\cdot,A=\cdot;t) = p(M=\cdot\,|\,G=\cdot,A=\cdot;t)\;p(G=\cdot,A=\cdot;t)$,
- suming over all $m \in \Omega_M$ and then collating all conditional probability terms.

C.2.1 Step 1:

$$\begin{aligned}\frac{d\,p(G=\mathbf{off},M=m,A=a;t)}{dt} &= \tau_{\mathbf{off}}\,p(M=m\,|\,G=\mathbf{on},A=a;t)\;p(G=\mathbf{on},A=a;t)\\ &+\gamma_1\,(m+1)\,p(M=m+1\,|\,G=\mathbf{off},A=a;t)\;p(G=\mathbf{off},A=a;t)\\ &+\kappa_2\,m\,p(M=m\,|\,G=\mathbf{off},A=a-1;t)\;p(G=\mathbf{off},A=a-1;t)\\ &+\gamma_2\,(a+1)\,p(M=m\,|\,G=\mathbf{off},A=a+1;t)\;p(G=\mathbf{off},A=a+1;t)\\ &-\left[\tau_{\mathbf{on}}+(\gamma_1+\kappa_2)\,m+(\gamma_2+\hat{\tau}_{\mathbf{on}})\,a\right]p(M=m\,|\,G=\mathbf{off},A=a;t)\;p(G=\cdot,A=\cdot;t).\end{aligned}$$

$$\begin{aligned}\frac{d\,p(G=\mathbf{on},M=m,A=a;t)}{dt} &= \tau_{\mathbf{on}}\,p(M=m\,|\,G=\mathbf{off},A=a;t)\;p(G=\mathbf{off},A=a;t)\\ &+\kappa_1\,p(M=m-1\,|\,G=\mathbf{on},A=a;t)\;p(G=\mathbf{on},A=a;t)\\ &+\gamma_1\,(m+1)\,p(M=m+1\,|\,G=\mathbf{on},A=a;t)\;p(G=\mathbf{on},A=a;t)\\ &+\kappa_2\,m\,p(M=m\,|\,G=\mathbf{on},A=a-1;t)\;p(G=\mathbf{on},A=a-1;t)\\ &+\gamma_2\,(a+1)\,p(M=m\,|\,G=\mathbf{on},A=a+1;t)\;p(G=\mathbf{on},A=a+1;t)\\ &+\hat{\tau}_{\mathbf{on}}\,(a+1)\,p(M=m\,|\,G=\mathbf{off},A=a+1;t)\;p(G=\mathbf{off},A=a+1;t)\\ &-\left\{\tau_{\mathbf{off}}+\kappa_1+(\gamma_1+\kappa_2)\,m+\gamma_2\,a\right\}p(M=m\,|\,G=\mathbf{on},A=a;t)\;p(G=\mathbf{on},A=a;t)\end{aligned}$$

C.2.2 Step 2:

$$\begin{aligned}\sum_m \frac{d\,p(G=\mathbf{off},M=m,A=a;t)}{dt} &= \tau_{\mathbf{off}}\left(\sum_m p(M=m\,|\,G=\mathbf{on},A=a;t)\right)\;p(G=\mathbf{on},A=a;t)\\ &+\gamma_1\left(\sum_m (m+1)\,p(M=m+1\,|\,G=\mathbf{off},A=a;t)\right)\;p(G=\mathbf{off},A=a;t)\\ &+\kappa_2\left(\sum_m m\,p(M=m\,|\,G=\mathbf{off},A=a-1;t)\right)\;p(G=\mathbf{off},A=a-1;t)\\ &+\gamma_2\,(a+1)\left(\sum_m p(M=m\,|\,G=\mathbf{off},A=a+1;t)\right)\;p(G=\mathbf{off},A=a+1;t)\\ &-\left[\tau_{\mathbf{on}}+\left(\sum_m (\gamma_1+\kappa_2)\,m\;p(M=m\,|\,G=\mathbf{off},A=a;t)\right)+(\gamma_2+\hat{\tau}_{\mathbf{on}})\,a\right]\;p(G=\mathbf{off},A=a;t).\end{aligned}$$

$$
\begin{aligned}
\sum_m \frac{d\,p(G=\mathbf{on},M=m,A=a;t)}{dt} \;=\;& \tau_{\mathbf{on}}\left(\sum_m p(M=m\mid G=\mathbf{off},A=a;t)\right)p(G=\mathbf{off},A=a;t)\\
&+\kappa_1\left(\sum_m p(M=m-1\mid G=\mathbf{on},A=a;t)\right)p(G=\mathbf{on},A=a;t)\\
&+\gamma_1\left(\sum_m (m+1)\,p(M=m+1\mid G=\mathbf{on},A=a;t)\right)p(G=\mathbf{on},A=a;t)\\
&+\kappa_2\left(\sum_m m\,p(M=m\mid G=\mathbf{on},A=a-1;t)\right)p(G=\mathbf{on},A=a-1;t)\\
&+\gamma_2\,(a+1)\left(\sum_m p(M=m\mid G=\mathbf{on},A=a+1;t)\right)p(G=\mathbf{on},A=a+1;t)\\
&+\hat{\tau}_{\mathbf{on}}\,(a+1)\left(\sum_m p(M=m\mid G=\mathbf{off},A=a+1;t)\right)p(G=\mathbf{off},A=a+1;t)\\
&-\left[\tau_{\mathbf{off}}+\kappa_1+\left(\sum_m(\gamma_1+\kappa_2)\,m\,p(M=m\mid G=\mathbf{on},A=a;t)\right)+\gamma_2\,a\right]p(G=\mathbf{on},A=a;t)
\end{aligned}
$$

D Formal ACE-Ansatz approximation derivation

Before we begin the derivation, it is important to discuss Assumption 2.1-3. We state that the joint distribution needs to have non-zero probability over all of the state space through all time. We can easily violate this condition by starting the Kurtz process with the initial probability distribution which is non-zero over only a subset of the entire state space (e.g. a single state). However, the CME generator (2.4) has the feature that regardless of the initial condition, in an infinitesimal time, all the states have non-zero probability. Hence, numerically, if the processes does start at a single state, we can evolve it forward by a small time step using OFSP, and then use this time point for the initial condition in the dimension reduction methods. In the case of the Simple Gene Switch example in §5.1.2, we used $t=1$ as the starting point for all dimension reduction methods.

We use the following notational convention: the approximation of the probability measure $p(G=g,A=a;t)$ is denoted by the function $w(g,a,t)$, furthermore, the approximation for the expectation operator $\mathbb{E}[\bullet(t)]$ is denoted by the function $\eta_\bullet(t)$. Then the formal derivation of equation (5.1) to (5.8) are given by equations (D.1) to (D.12).

$$
\begin{aligned}
\frac{d\,w(\mathbf{off},a,t)}{dt} =\;& \tau_{\mathbf{off}}\,w(\mathbf{on},a,t)\\
&+k_2\,\eta_{M|}(\mathbf{off},a-1,t)\,w(\mathbf{off},a-1,t)\\
&+\gamma_2\,(a+1)\;w(\mathbf{off},a+1,t)\\
&-\left(\tau_{\mathbf{on}}+k_2\,\eta_{M|}(\mathbf{off},a,t)+(\gamma_2+\hat{\tau}_{\mathbf{on}})\,a\right)\;w(\mathbf{off},a,t),
\end{aligned}
\tag{D.1}
$$

$$
\begin{aligned}
\frac{d\,w(\mathbf{on},a,t)}{dt} =\;& \tau_{\mathbf{on}}\,w(\mathbf{off},a,t)\\
&+k_2\,\eta_{M|}(\mathbf{on},a-1,t)\,w(\mathbf{on},a-1,t)\\
&+\gamma_2\,(a+1)\,w(\mathbf{on},a+1,t)\\
&+\hat{\tau}_{\mathbf{on}}\,(a+1)\,w(\mathbf{off},a+1,t)\\
&-\left(\tau_{\mathbf{off}}+k_2\,\eta_{M|}(\mathbf{on},a,t)+\;\gamma_2\,a\right)\;w(\mathbf{on},a,t).
\end{aligned}
\tag{D.2}
$$

$$
\eta_{M|}(g,a,t)=\alpha\left(\begin{bmatrix} g\\ a\end{bmatrix}-\begin{bmatrix}\eta_G(t)\\ \eta_A(t)\end{bmatrix}\right)+\eta_M(t)
\tag{D.3}
$$

$$
\frac{d\,\eta_M(t)}{dt}=k_1\,\eta_G(t)-\gamma_1\,\eta_M(t).
\tag{D.4}
$$

$$
\begin{aligned}
\frac{d\,\eta_{G\,M}(t)}{dt} =\;& \tau_{\mathbf{on}}\;(-\eta_{G\,M}(t)-\eta_M(t))-\tau_{\mathbf{off}}\,\eta_{G\,M}(t)+k_1\,\eta_G(t)\\
&-\gamma_1\,\eta_{G\,M}(t)+\hat{\tau}_{\mathbf{on}}\;(\eta_{M\,A}(t)-\eta_{G\,M\,A}(t))\,.
\end{aligned}
\tag{D.5}
$$

$$
\begin{aligned}
\frac{d\,\eta_{M\,A}(t)}{dt} =\;& k_1\,\eta_{G\,A}(t)-(\gamma_1+\gamma_2)\,\eta_{M\,A}(t)+k_2\,\eta_{M^2}(t)\\
&-\hat{\tau}_{\mathbf{on}}\;(\eta_{M\,A}(t)-\eta_{G\,M\,A}(t))
\end{aligned}
\tag{D.6}
$$

$$\frac{d\,\eta_{M^2}(t)]}{dt} = k_1\ (2\,\eta_{G\,M}(t) + \eta_G(t)) + \gamma_1\ (-2\,\eta_{M^2}(t) + \eta_M(t))\,. \tag{D.7}$$

$$\eta_{G\,M\,A}(t) = \sum_{a\in\mathbb{Z}_+} \eta_{M|}(\mathbf{on}, a, t)\, a\, w(\mathbf{on}, a, t). \tag{D.8}$$

$$\eta_A(t) = \sum_{a\in\mathbb{Z}_+} a\ [w(\mathbf{on}, a, t) + w(\mathbf{off}, a, t)]\,. \tag{D.9}$$

$$\eta_{A^2}(t) = \sum_{a\in\mathbb{Z}_+} a^2\ [w(\mathbf{on}, a, t) + w(\mathbf{off}, a, t)]\,. \tag{D.10}$$

$$\eta_{G^2}(t) = \eta_G(t). \tag{D.11}$$

$$\alpha := \begin{bmatrix} \eta_{G\,M}(t) - \eta_G(t)\,\eta_M(t) & \eta_{M\,A}(t) - \eta_M(t)\,\eta_A(t) \end{bmatrix} \begin{pmatrix} \eta_{G^2}(t) - \eta_G(t)^2 & \eta_{G\,A}(t) - \eta_G(t)\,\eta_A(t) \\ \eta_{G\,A}(t) - \eta_G(t)\,\eta_A(t) & \eta_{A^2}(t) - \eta_A(t)^2 \end{pmatrix}^{-1}. \tag{D.12}$$

E SIR system parameters

The initial starting population was set to $(S(0) = 200, I(0) = 4)$. The OFSP method was configured to have a global error of 10^{-6}, with compression performed every 10 steps where each time step was of length 0.002. The distribution is the snapshot of the system at $t = 0.15$. We also omit the recovered state since the total population is conserved, that is, $S(t) + I(t) + R(t) = 204$ for all time.

#	Reaction	Coefficient	Stoichiometry	Description
1	$S + I \xrightarrow{c_1} 2I$	$c_1 = 0.3$	$(-1, 1)$	Susceptible person becomes an infected person
2	$I \xrightarrow{c_2} \emptyset$	$c_2 = 5.5$	$(0, -1)$	Infected person leaves the system

Table 10: SIR system parameters

List of Figures

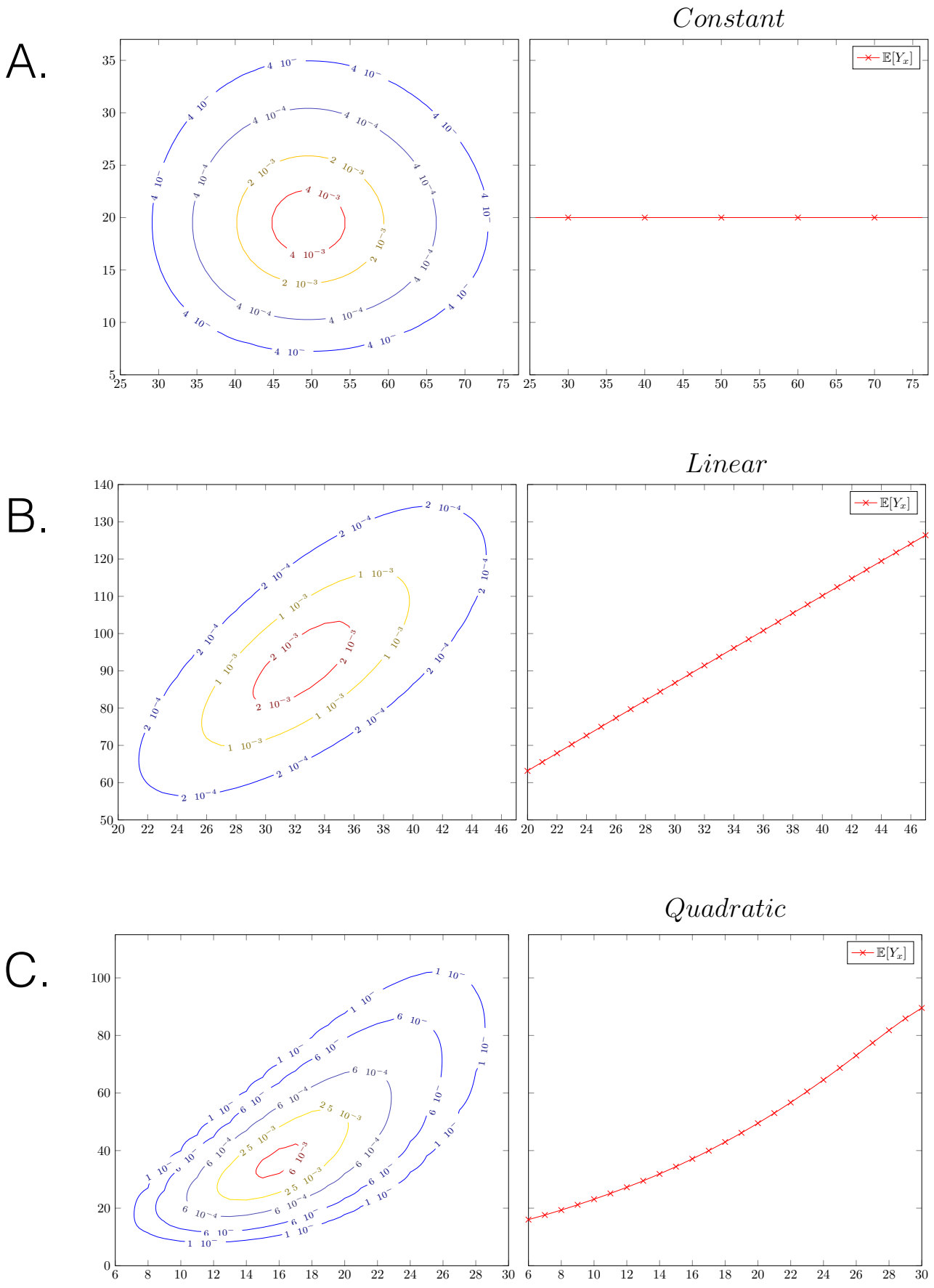

Figure 1: **A–C.** (Left) Contour plots describing the joint probability distributions with constant, linear, and quadratic conditional expectation forms, respectively. (Right) The conditional expectations of the distributions in the left figure. (see Appendix B for system parameters)

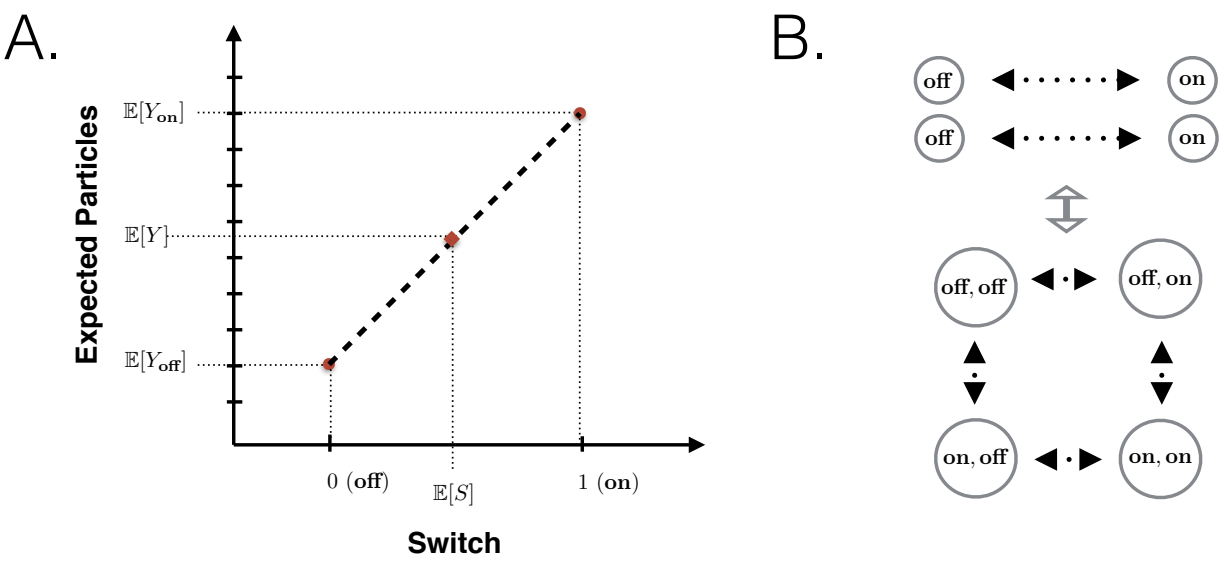

Figure 2: **A.** Cartoon illustrating the consequence of Theorem 4.1. **B.** Cartoon illustrating the decomposition of a coupled two switch problem into an uncoupled four switch problem

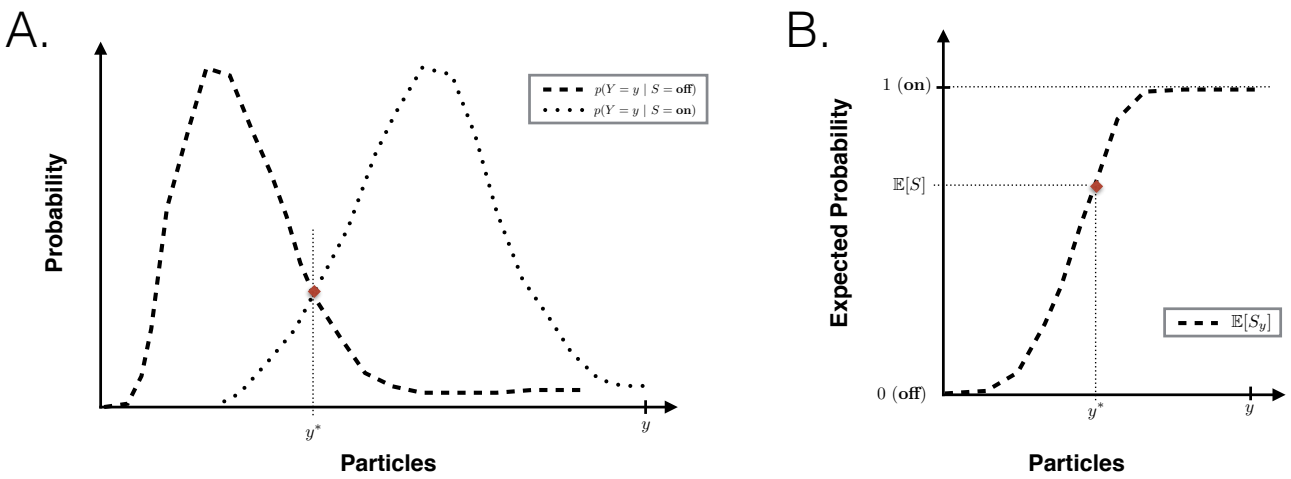

Figure 3: **A–B.** Cartoons illustrating the consequence of Theorem 4.2.

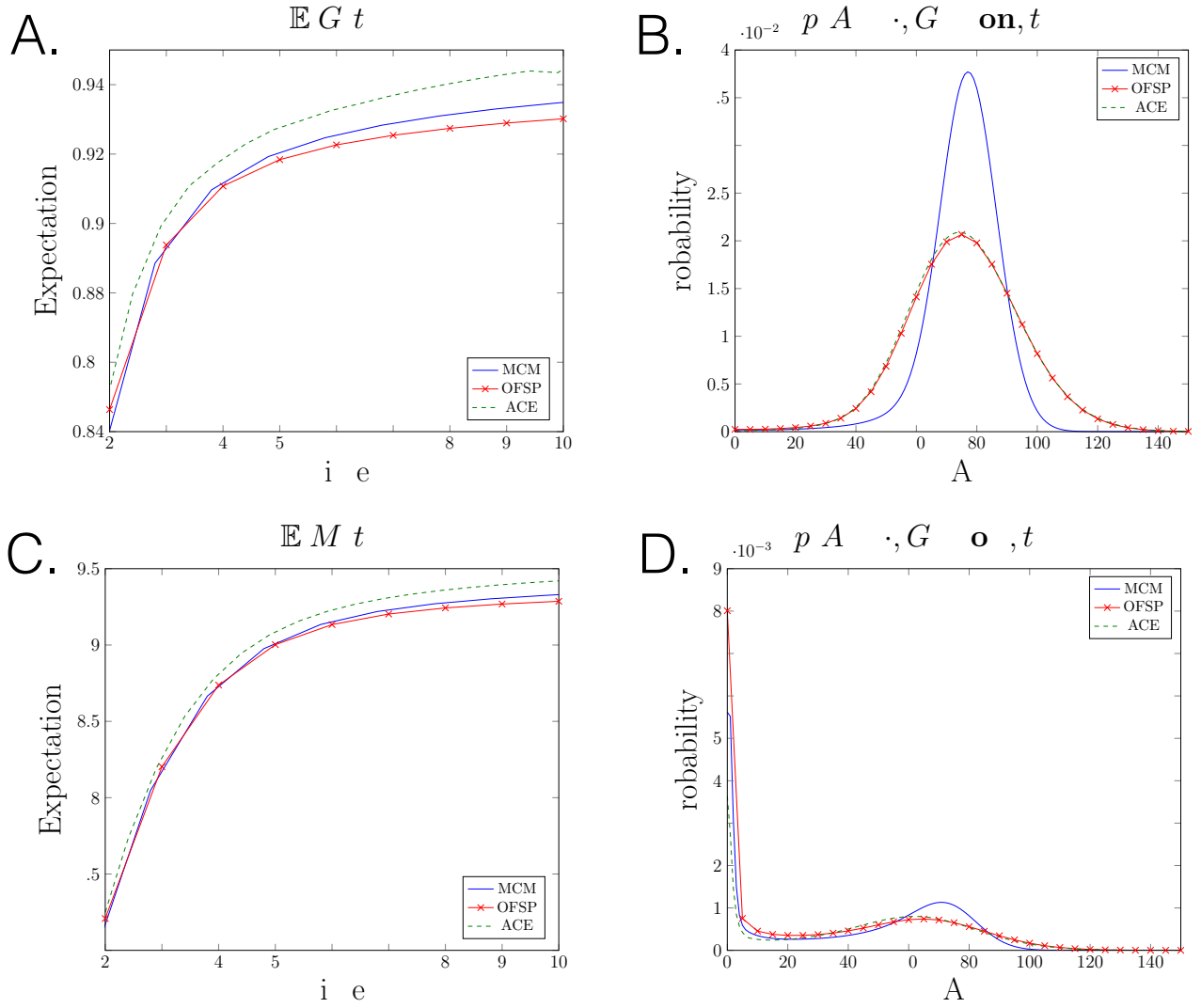

Figure 4: Comparison of different method's approximation of the Simple Gene Switch Model (§5.1.1). **A**–**D.** (−−) Double dashed green line is representative of the ACE-Ansatz approximation, (–) solid blue line is representative of the MCM approximation, and (-×-) the crossed red line is represenative of the OFSP reference solution. **A.** The expectation of the gene being in **on** state at time t. **B.** The probability distribution of being in an **on** state at time $t = 10$. **C.** The expectation of the population of $mRNA$ in the system at time t. **D.** The probability distribution of being in an **off** state at time $t = 10$.

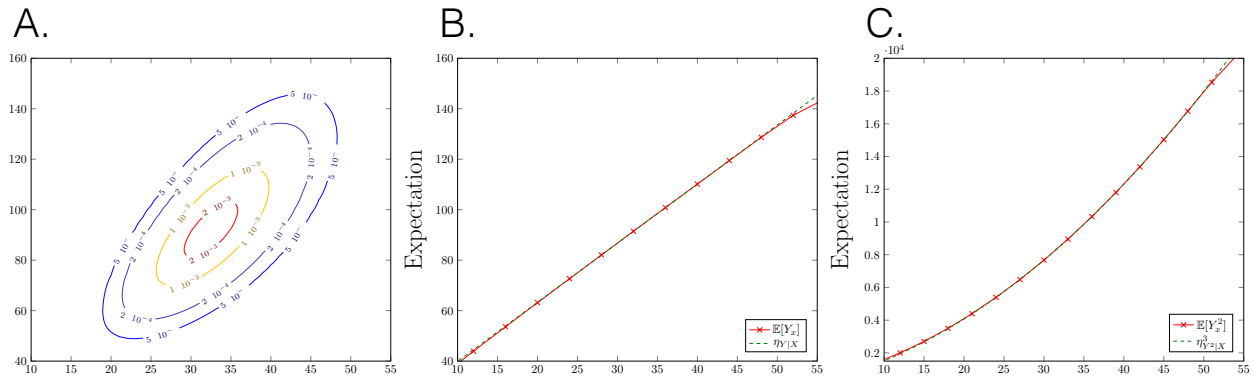

Figure 5: **A.** Joint probability distribution of **Model 2** (Example 3.1) plotted via a contour plot. **B.** The conditional expectation of the distribution in (A.) illustrated with a red crossed line (-×-) and the linear ACE approximation of the conditional expectation illustrated with a dashed green line (−−). **C.** The squared conditional expectation of the distribution in (A.) illustrated with a red crossed line (-×-) and the quadratic ACE approximation of the conditional expectation illustrated with a dashed green line (−−).

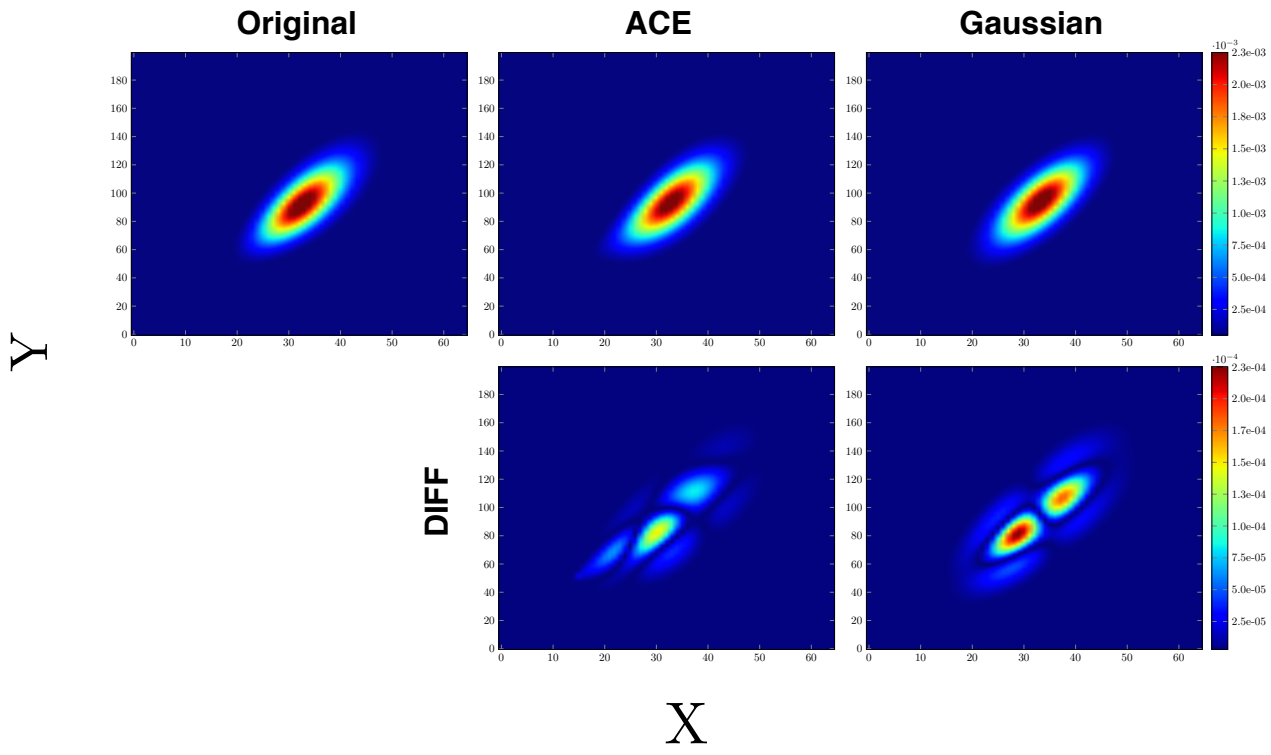

Figure 6: (**Top row**) Heat maps showing the original distribution to be approximated on the left, then the ACE-Reconstruction and Gaussian reconstruction in the adjacent heat maps. The intensities are corresponding to the colour bar given on the right. (**Bottom row**) The heat maps show the pointwise absolute difference between the reconstructed and the original distribution, the reconstruction method is given in the column heading. The intensities are corresponding to the colour bar on the right.

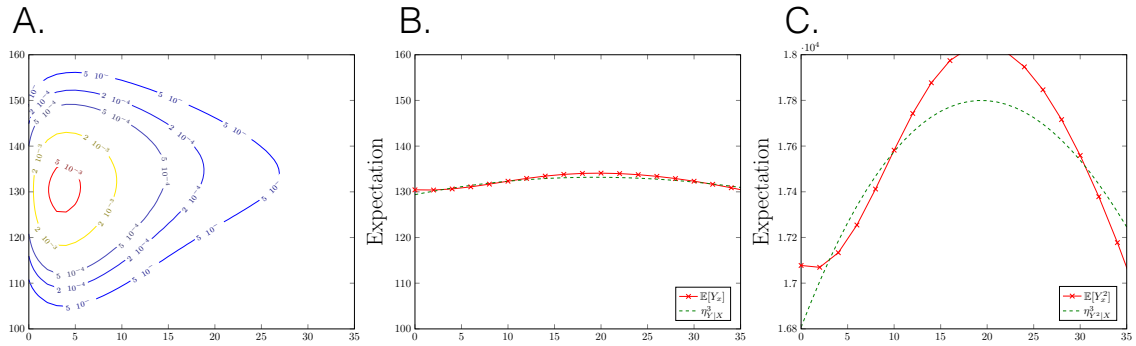

Figure 7: **A.** Joint probability distribution of the SIR system (Appendix E) plotted via a contour plot. **B.** The conditional expectation of the distribution in (A.) illustrated with a red crossed line (-×-) and the cubic ACE approximation of the conditional expectation illustrated with a dashed green line (−−). **C.** The squared conditional expectation of the distribution in (A.) illustrated with a red crossed line (-×-) and the quartic ACE approximation of the conditional expectation illustrated with a dashed green line (−−).

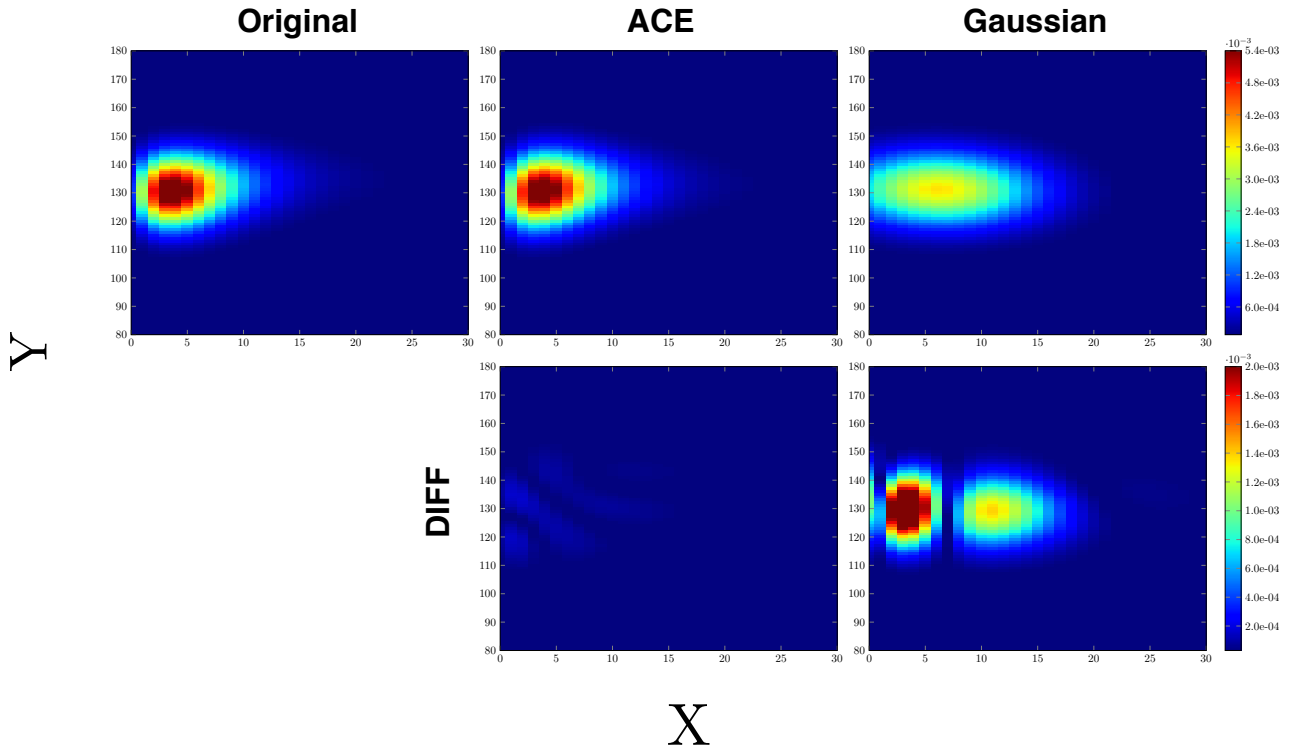

Figure 8: (**Top row**) Heat maps showing the original distribution to be approximated on the left, then the ACE-Reconstruction and Gaussian reconstruction in the adjacent heat maps. The intensities are corresponding to the colour bar given on the right. (**Bottom row**) The heat maps show the pointwise absolute difference between the reconstructed and the original distribution, the reconstruction method is given in the column heading. The intensities are corresponding to the colour bar on the right.