Tensor methods for parameter estimation and bifurcation analysis of stochastic reaction networks

Shuohao Liao*,   Tomáš Vejchodský†,   Radek Erban*

April 12, 2015

Abstract

Stochastic modelling of gene regulatory networks provides an indispensable tool for understanding how random events at the molecular level influence cellular functions. A common challenge of stochastic models is to calibrate a large number of model parameters against the experimental data. Another difficulty is to study how the behaviour of a stochastic model depends on its parameters, i.e. whether a change in model parameters can lead to a significant qualitative change in model behaviour (bifurcation). In this paper, tensor-structured parametric analysis (TPA) is developed to address these computational challenges. It is based on recently proposed low-parametric tensor-structured representations of classical matrices and vectors. This approach enables simultaneous computation of the model properties for all parameter values within a parameter space. The TPA is illustrated by studying the parameter estimation, robustness, sensitivity and bifurcation structure in stochastic models of biochemical networks. A Matlab implementation of the TPA is available at `http://www.stobifan.org`.

1 Introduction

Many cellular processes are influenced by stochastic fluctuations at the molecular level which are often modelled using stochastic simulation algorithms for chemical reaction networks [1, 2]. For example, cell metabolism, signal transduction and cell cycle can be described by network structures of functionally separated modules of gene expression [3], the so called gene regulatory networks (GRNs).

Typical GRN models can have tens of variables and parameters. Traditionally, GRNs have been described using continuous deterministic models written as systems of ordinary differential equations (ODEs). Several methodologies for studying parametric properties of ODE systems, such as identifiability and bifurcation, have been developed in the literature [4, 5, 6, 7, 8]. Recently, experimental evidence has highlighted the significance of intrinsic randomness in GRNs, and stochastic models have been increasingly used [1, 9]. They are usually simulated using the Gillespie stochastic simulation algorithm (SSA) [10], or its equivalent formulations [11, 12]. However, methods for parametric analysis of ODEs cannot be directly applied to stochastic models. In this paper, we present a *tensor-structured parametric analysis* (TPA) which can be used to understand how molecular-level fluctuations influence the system-level behaviour of GRNs and its dependence on model parameters. We illustrate major application areas of the TPA by studying several biological models with increasing level of complexity.

The parametric analysis of GRN models is computationally intensive because both state space and parameter space are high-dimensional. The dimension of the state space, $\Omega_{\mathbf{x}}$, is equal to the number of reacting molecular species, denoted by N. When an algorithm, previously working with deterministic steady states, is extended to stochastic setting, its computational complexity is typically taken to the power N. Moreover, the exploration of the parameter space, $\Omega_{\mathbf{k}}$, introduces another multiplicative exponential complexity. Given

---

*Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom; e-mails: liao@maths.ox.ac.uk; erban@maths.ox.ac.uk

†Institute of Mathematics, Czech Academy of Sciences, Zitna 25, 115 67 Praha 1, Czech Republic; e-mail vejchod@math.cas.cz

a system that involves K parameters, the 'amount' of parameter combinations to be characterized scales equally with the volume of $\Omega_{\mathbf{k}}$, i.e. it is taken to the power K [13].

The TPA framework avoids the high computational cost of working in high-dimensional $\Omega_{\mathbf{x}}$ and $\Omega_{\mathbf{k}}$. The central idea is based on generalising the concept of separation of variables to parametric probability distributions [14]. The TPA framework can be divided into two main steps: a tensor-structured computation and a tensor-based analysis. First, the steady state distributions of stochastic models are simultaneously computed for all possible parameter combinations within a parameter space and stored in a tensor format, with smaller computational and memory requirements than in traditional approaches. The resulting tensor data are then analyzed using algebraic operations with computational complexity which scales linearly with dimension (i.e., linearly with N and K).

The rest of this paper is organised as follows. In Section 2, we discuss how the parametric steady state probability distribution can be presented and computed in tensor formats. We illustrate the data storage savings using tensor-structured simulations of four biological systems. The stored tensor data are then used as the input for the tensor-based analysis presented in the subsequent sections. In Section 3, we show that the existing procedures for parameter inference for deterministic models can be directly extended to the stochastic models using the computed tensor data. In Section 4, a direct visualisation of stochastic bifurcations in a high-dimensional state space is presented. The TPA of the robustness of the network to extrinsic noise is illustrated in Section 5. We conclude with a brief discussion in Section 6.

2 Tensor-structured computations

Considering a well-mixed chemically-reacting system of N distinct molecular species X_i, $i = 1, 2, \ldots, N$, inside a reactor (e.g. cell) of volume V, we denote its state vector by $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T$, where x_i is the number of molecules of the i-th chemical species X_i. In general, the volume V can be time dependent (for example, in cell cycle models which explicitly take into account cell growth), but we will focus in this paper on models with constant values of V. We assume that molecules interact through M reaction channels

$$\sum_{i=1}^{N} \nu_{j,i}^- X_i \xrightarrow{k_j} \sum_{i=1}^{N} \nu_{j,i}^+ X_i, \quad j = 1, 2, \ldots, M, \tag{1}$$
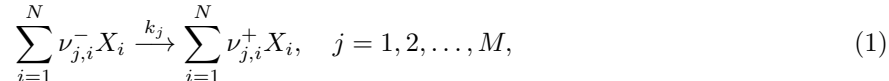
where $\nu_{j,i}^+$ and $\nu_{j,i}^-$ are the stoichiometric coefficients. The kinetic rate parameters, $\mathbf{k} = (k_1, k_2, \ldots, k_M)^T$, characterise the rate of the corresponding chemical reactions. We will treat $\mathbf{k}$ as auxiliary variables, and in other words, the parametric problem of (1) involves considering both $\mathbf{x} \in \Omega_{\mathbf{x}}$ and $\mathbf{k} \in \Omega_{\mathbf{k}}$. In this paper, we study problems where the dimension of the parameter space K is equal to M. We also consider cases where some rate constants are not varied in the parameter analysis, i.e. $K < M$. In this case, notation $\mathbf{k}$ will be used to denote K-dimensional vector of rate constants, $\mathbf{k} = (k_1, k_2, \ldots, k_K)^T$, which are considered during the TPA. The values of the remaining $(M - K)$ rate constants are fixed. In principle, the TPA could also be used to study models where $K > M$, i.e. when we consider additional parameters (e.g., system volume V).

Let $p(\mathbf{x}|\mathbf{k})$ be the steady state probability distribution that the state vector is $\mathbf{x}$ (if the system is observed for sufficiently long time) given the parameter values $\mathbf{k}$. The main idea of the TPA is to split $p(\mathbf{x}|\mathbf{k})$ in terms of coordinates as

$$p(\mathbf{x}|\mathbf{k}) = \sum_{\ell=1}^{R} \underbrace{f_1^\ell(x_1) \cdots f_N^\ell(x_N)}_{\Omega_{\mathbf{x}}} \underbrace{g_1^\ell(k_1) \cdots g_K^\ell(k_K)}_{\Omega_{\mathbf{k}}}, \tag{2}$$

where $\{f_i^\ell(x_i)\}_{i=1,\ldots,N}$ and $\{g_j^\ell(k_j)\}_{j=1,\ldots,K}$ are univariate functions that vary solely with a single state variable and parameter, respectively. The number of summands R, the so called *separation rank*, controls the accuracy of the decomposition (2). By increasing R, the separated expansion could theoretically achieve arbitrary accuracy.

The value of the separation rank R can be analytically computed for simple systems. For example, there are analytical formulas for the stationary distributions of first-order stochastic reaction networks [15]. They

Table 1: *Comparison of the matrix-based and tensor-structured methodologies.*

Biochemical system	Dimensionality			Matrix-based		Tensor-structured			
	N	K	$N+K$	$\text{Mem}^{\dagger}_{\text{CME}}$	$\text{Mem}^{\dagger}_{\text{CFPE}}$	$\text{Mem}_{\mathbf{A}}$	Mem_p	$T_{\mathbf{A}}[\text{sec}]$	$T_{\text{tot}}[\text{min}]$
Schlögl	1	4	5	2.68×10^{13}	2.74×10^{11}	4.01×10^{3}	2.07×10^{5}	1.2	30
Cell cycle	6	1	7	6.68×10^{17}	7.04×10^{13}	2.96×10^{4}	1.00×10^{7}	1.1	6433
FitzHugh-Nagumo	2	4	6	6.38×10^{14}	1.75×10^{13}	7.65×10^{4}	4.02×10^{5}	0.7	37
Reaction chain	20	0	20	1.20×10^{44}	1.53×10^{54}	9.26×10^{4}	7.28×10^{5}	15.6	283

† Estimated as the product of the number of discrete states and the number of parameter values.

are given in the form (2) with $R = 1$. Considering second-order stochastic reaction networks, there are no general analytical formulas for steady state distributions. They have to be approximated using computational methods. The main assumption of the TPA approach is that the parametric steady state distribution has a sufficiently accurate low-rank representation (2). In this paper, we show that this assumption is satisfied for realistic biological systems by applying the TPA to them and presenting computed (converged) results. The main consequence of low-rank representation (2) is that mathematical operations on the probability distribution $p(\mathbf{x}|\mathbf{k})$ in $N+K$ dimensions can be performed using combinations of one-dimensional operations, and the storage cost is bounded by $(N + K)R$. The rank R may also depend on $N + K$ and the size of the univariate functions in (2). Numerical experiments have shown a linear growth of R with respect to $N + K$ and a logarithmic growth with respect to the size of the univariate functions in the representation (2) [16, 17]. To find the representation (2), we solve the chemical Fokker-Planck equation (CFPE), as a (fully) continuous approximation to a (continuous time) discrete space Markov chain described by the corresponding chemical master equation (CME) [18, 19]. Specifically, we keep all the objects in the separated form of (2) during the computations, such that exponential scaling in complexity does not apply during any step of the TPA.

We refer to the representation (2) as tensor-structured, because computations are performed on $p(\mathbf{x}|\mathbf{k})$ as multidimensional arrays of real numbers, which we call tensors [20]. The (canonical) tensor decomposition [21], as a discrete counterpart of (2), then allows a multidimensional array to be approximated as a sum of tensor products of one-dimensional vectors. Within such a format, we can define standard algebraic operations similar to standard matrix operations such that the resulting tensor calculus enables efficient computation. The tensor-structured parametric steady state distribution (2) is approximated as the eigenfunction corresponding to the smallest eigenvalue of the parametric Fokker-Planck operator. The operator is constructed in a tensor separated representation as a sum of tensor products of one-dimensional operators. The eigenfunction is computed by adaptive shifted inverse power method, using minimum alternating energy method as the linear solver. We leave further discussion of technical computational details of the underlying methods to *Supporting Information (SI) Appendix* S1. The TPA has been implemented in MATLAB, and is part of the Stochastic Bifurcation Analyzer toolbox available at `http://www.stobifan.org`. The source code relies on the Tensor Train Toolbox [22].

2.1 Applications of the TPA to biological systems

We demonstrate the capabilities of the TPA framework by investigating four examples of stochastic reaction networks: a bistable switch in the 5-dimensional Schlögl model [23], oscillations in the 7-dimensional cell cycle model [24], neurons excitability in the 6-dimensional FitzHugh-Nagumo system [4] and a 20-dimensional reaction chain [25] (see *SI Appendix* S2 for more details of these models). Table 1 compares computational performance of the TPA with the traditional matrix-based methods for the computation of the parametric steady state distribution $p(\mathbf{x}|\mathbf{k})$. The minimum memory requirements of solving the CME and the CFPE using matrix-based methods, Mem_{CME} and Mem_{CFPE}, are estimated as products of numbers of discrete states times the total number of parameter combinations. They vary in ranges 10^{13}–10^{44} and 10^{11}–10^{54}, respectively, which are beyond the limits of the available hardware. In contrary, the TPA maintains affordable computational and memory requirements for all four problems considered, as we show in Table 1. The major

Table 2: *Parameter estimation for ODEs.*

(a1) Generate a candidate parameter vector $\mathbf{k}^* \in \Omega_{\mathbf{k}}$.

(a2) Compute model prediction $\mathbf{x}^*$ using the parameter vector $\mathbf{k}^*$.

(a3) Compare the simulated data $\mathbf{x}^*$ with the experimental evidence $\hat{\mathbf{x}}$, using distance function $d(\hat{\mathbf{x}}, \mathbf{x}^*)$ and tolerance ε. If $d(\hat{\mathbf{x}}, \mathbf{x}^*) < \varepsilon$, then accept $\mathbf{k}^*$. The tolerance $\varepsilon > 0$ is the desired level of agreement between $\hat{\mathbf{x}}$ and $\mathbf{x}^*$.

---

memory requirements of the TPA are $\text{Mem}_{\mathbf{A}}$ and Mem_p to store the discretized Fokker-Planck operator and the steady state distribution $p(\mathbf{x}|\mathbf{k})$, respectively (see *SI Appendix* S1 for detailed definitions). Similarly, $T_{\mathbf{A}}$ is the computational time to assemble the operator and T_{tot} is the total computational time.

Table 1 shows that the TPA can outperform standard matrix-based methods. It can also be less computationally intensive than stochastic simulations in some cases. For example, the total computational time is around 30 minutes for the TPA to simulate 64^4 different parameter combinations within the 4-dimensional parameter space of the Schlögl chemical system (see Table 1). If we wanted to compute the same result using the Gillespie SSA, we would have to run 64^4 different stochastic simulations. If they had to be all performed on one processor in 30 minutes, then we would only have 1.07×10^{-4} second per one stochastic simulation and it would not be possible to estimate the results with the same level of accuracy. In addition the TPA directly provides the steady state distribution $p(\mathbf{x}|\mathbf{k})$, which would be computationally intensive to obtain by stochastic simulations (with the same level of accuracy) for larger values of $N + K$.

3 Parameter estimation

Small uncertainties in the reaction rate values of stochastic reaction networks (1) are common in applications. Some model parameters are difficult to measure directly, and instead are estimated by fitting to time-course data. If GRNs are modelled using deterministic ODEs, there is a wide variety of tools available for parameter estimation. Many simple approaches are non-statistical [26], and the procedure usually, although not necessarily [27], follows the algorithm presented in Table 2. We tune uncertain parameters to minimize the distance measure $d(\hat{\mathbf{x}}, \mathbf{x}^*)$, while the rules to generate candidate parameters $\mathbf{k}^*$ in step (a1) and the definition of distance function along with stopping criteria in step (a3) may vary in different methods. In optimization-based methods, $\mathbf{k}^*$ may follow the gradient on the surface of the distance function [26]. In statistical methods, such distance measure is provided in the concept of likelihood, $\mathcal{L}(\mathbf{k}^*|\hat{\mathbf{x}}) = p(\hat{\mathbf{x}}|\mathbf{k}^*)$ [28]. In Bayesian methods, the candidate parameters $\mathbf{k}^*$ are generated from some prior information regarding uncertain parameters, $\pi(\mathbf{k})$, and form a posterior distribution rather than a single point estimate [29].

To extend the algorithm in Table 2 from deterministic ODEs to stochastic models requires substantial modifications [29]. One main obstacle is the step (a2) which requires repeatedly generating the likelihood function $\mathcal{L}(\mathbf{k}^*|\hat{\mathbf{x}})$, as the outcome of stochastic models. In this case, a modeller must either apply statistical analysis to approximate the likelihood [30], or use the Gillespie SSA to estimate it [31]. Consequently, the algorithms are computationally intensive and do not scale well to problems of realistic size and complexity. To avoid this problem, the TPA uses the tensor formalism to separate the simulation part from the parameter inference. The parameter estimation is performed on the tensor data obtained by methods described above (see Table 1). The algorithm used for the TPA parameter estimation is given in Table 3. The distance function $d(\hat{\mathbf{x}}, \mathbf{x}^*)$ is replaced with a distance between summary statistics, $\hat{S}$ and S^*, which describe the average behaviour and the characteristics of the system noise. The steps (b1) and (b3) are similar to steps (a1) and (a3) under the ODE settings, and a variety of existing methods can be extended directly to stochastic settings. The newly introduced step (b0) is executed only once during the parameter estimation. Steps (b1)–(b3) are then repeated until convergence. Step (b2) only requires manipulation of tensor data, of which the computational overhead is comparable to solving an ODE.

Table 3: *An algorithm for the tensor-structured parameter estimation.*

(b0) Compute the stationary distribution $p(\mathbf{x}|\mathbf{k})$ for all considered combinations of $\mathbf{x} \in \Omega_{\mathbf{x}}$ and $\mathbf{k} \in \Omega_{\mathbf{k}}$; and store $p(\mathbf{x}|\mathbf{k})$ in tensor data.

(b1) Generate a candidate parameter vector $\mathbf{k}^* \in \Omega_{\mathbf{k}}$.

(b2) Extract the stationary distribution $p(\mathbf{x}|\mathbf{k}^*)$ from the tensor-structured data $p(\mathbf{x}|\mathbf{k})$, and compute the summary statistics $S^* \equiv S^*(p)$.

(b3) Compare the model prediction S^* with the statistics $\hat{S}$ obtained from experimental data, using distance function $J(\hat{S}, S^*)$ and tolerance ε. If $J(\hat{S}, S^*) < \varepsilon$, then accept $\mathbf{k}^*$.

The tolerance $\varepsilon > 0$ is the desired level of agreement between $\hat{S}$ and S^*.

3.1 An example of parameter estimation

We consider that the distance measure $J(\hat{S}, S^*)$ in Table 3 is defined using a moment matching procedure [32, 33]:

$$J(\hat{S}, S^*) = \sum_{i_1,\ldots,i_N=1}^{L} \beta_{i_1,\ldots,i_N} \left(\frac{\hat{\mu}_{[i_1,\ldots,i_N]} - \mu_{[i_1,\ldots,i_N]}(\mathbf{k}^*)}{\hat{\mu}_{[i_1,\ldots,i_N]}} \right)^2, \tag{3}$$

where $\hat{\mu}_{[i_1,\ldots,i_N]}$ is the $(i_1,\ldots,i_N)$-th order empirical raw moment, $\mu_{[i_1,\ldots,i_N]}(\mathbf{k}^*)$ is the corresponding moment derived from $p(\mathbf{x}|\mathbf{k}^*)$ and L denotes the upper bound for the moment order. The weights, $\beta_{i_1,\ldots,i_N}$, can be chosen by modellers to attribute different relative importances to moments. Empirical moments are estimated from samples $\hat{x}_{d,\ell}$, $d = 1, 2, \ldots, N$, $\ell = 1, 2, \ldots, n_{\hat{\mu}}$, by

$$\hat{\mu}_{[i_1,\ldots,i_N]} = \frac{1}{n_{\hat{\mu}}} \sum_{\ell=1}^{n_{\hat{\mu}}} \hat{x}_{1,\ell}^{i_1} \cdots \hat{x}_{N,\ell}^{i_N}, \tag{4}$$
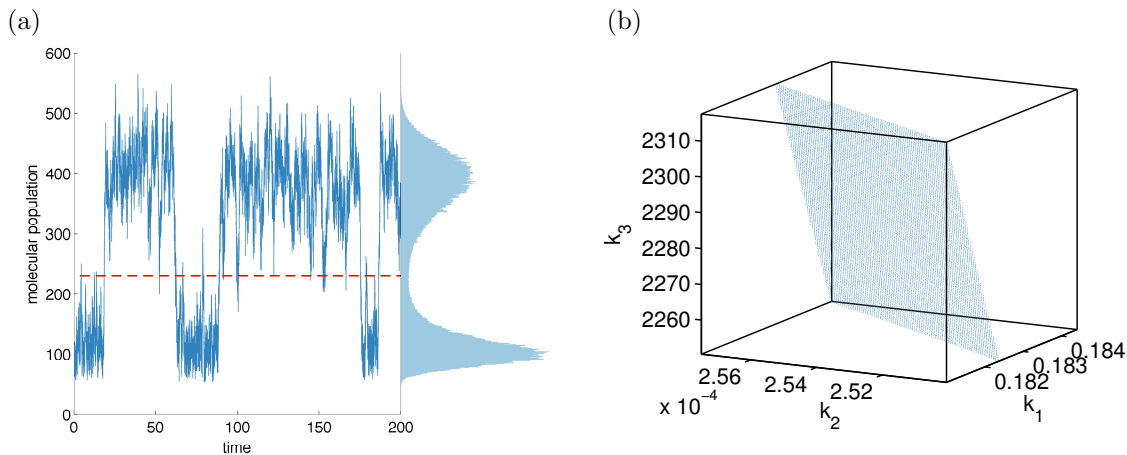
(a)   (b)

Figure 1: (a) *A short segment of the time series data and the histogram for the Schlögl reaction system generated by a long-time stochastic simulation. The dashed line corresponds to the threshold* 230 *which is used to separate the two macroscopic states of this bistable system.* (b) *The triplets of parameters* $[k_1, k_2, k_3]$ *for which the splitting probability* (6) *is equal to* $\hat{S} = 47.61\% \pm 5\%$. *The value of* k_4 *is fixed at its true value.*
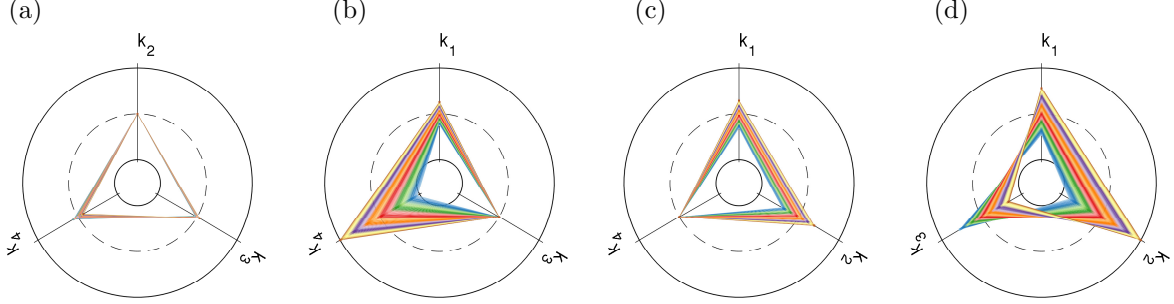
Figure 2: *Circular representation* [34] *of estimated parameter combinations for the Schlögl model. Each spoke represents the corresponding parameter range listed in Table* S4. *The true parameter values are specified by the intersection points between the spokes and the dashed circle. Each triangle (or polygon in general) of a fixed colour corresponds to one admissible parameter set with $\varepsilon = 0.25\%$. Each panel* (a)–(d) *shows the situation with one parameter fixed at its true value, namely:* (a) k_1 *is fixed;* (b) k_2 *is fixed;* (c) k_3 *is fixed;* (d) k_4 *is fixed.*

where $n_{\hat{\mu}}$ is the number of samples. Moments of the model output are computed as

$$\mu_{[i_1,\ldots,i_N]}(\mathbf{k}^*) = \int_{\Omega_{\mathbf{x}}} x_1^{i_1} \cdots x_N^{i_N} \, p\,(\mathbf{x}|\mathbf{k}^*)\,\mathrm{d}\mathbf{x}. \tag{5}$$

We show, in *SI Appendix* S1.4, that it is possible to directly compute different orders of moments, $\mu_{[i_1,\ldots,i_N]}(\mathbf{k}^*)$, using the representation (2) with $O(N)$ complexity.

We illustrate the tensor-structured parameter estimation using the Schlögl chemical system [23], which is written for $N = 1$ molecular species and has $M = 4$ reaction rate constants k_i, $i = 1, 2, 3, 4$. A detailed description of this system is provided in *SI Appendix* S2.1. We prescribe true parameter values as $k_1 = 2.5 \times 10^{-4}$, $k_2 = 0.18$, $k_3 = 2250$ and $k_4 = 37.5$, and use a long-time stochastic simulation to generate a time series as pseudo-experimental data (for a short segment, see Figure 1(a)). These pseudo-experimental data are then used for estimating the first three empirical moments $\hat{\mu}_i$, $i = 1, 2, 3$, using (4). While the moments of the model output, $\mu_i(\mathbf{k})$, $i = 1, 2, 3$, are derived from the tensor-structured data $p(\mathbf{x}|\mathbf{k})$, computed using (2). Moment matching is sensitive to the choice of weights [33]. However, for the sake of simplicity we choose the weights β_i, $i = 1, 2, 3$, in a way that the contributions of the different orders of moments are of similar magnitude within the parameter space. Having the stationary distribution stored in the tensor format (2), we can then efficiently iterate steps (b1)–(b3) in Table 3 to search for parameter values that produce adequate fit to the samples using the measure given in equation (3). We consider $\varepsilon = 0.25\%$ and visualise in Figure 2 the admissible parameter values satisfying $J(\hat{S}, S^*) < \varepsilon$.

The summary statistics $\hat{S}$ are not restricted to lower order moments. The TPA can efficiently evaluate different choices of the summary statistics, because of the simplicity and generality of separable representation (2). For example, if one can experimentally measure the probability that the system stays in each of the two states of the bistable system, then distance measure $J(\hat{S}, S^*)$ can be based on the probability of finding the system within a particular part of the state space $\Omega_{\mathbf{x}}$. We show in *SI Appendix* S1.4 that such quantity can also be estimated in the tensor format efficiently with $O(N)$ complexity. Considering the Schlögl model, we estimate the probability that the system stays in the state with less molecules by

$$S = \mathbb{P}(x \leq 230), \tag{6}$$

where $\mathbb{P}$ denotes the probability and the threshold 230 separates the two macroscopic states of the Schlögl system, see the dashed line in Figure 1(a). The splitting probability (6) can be estimated using long time simulation of the Schlögl system as the fraction of states which are less or equal than 230 and is equal to $\hat{S} = 47.61\%$ for our true parameter values. Figure 1(b) shows the set of admissible parameters within the

parameter space $\Omega_{\mathbf{k}}$ whose values provide desired agreement on the splitting probability (6) with tolerance $\varepsilon = 5\%$, i.e. we use

$$J(\hat{S}, S^*) = |\hat{S} - S^*|$$

in the algorithm given in Table 3, where S^* is computed using (2) and (6).

3.2 Identifiability

One challenge of mathematical modelling of GRNs is whether unique parameter values can be determined from available data. This is known as the problem of identifiability. Inappropriate choice of the distance measure may yield ranges of parameter values with equally good fit, i.e. the parameters being not identifiable [35]. Here, we illustrate the tensor-structured identifiability analysis of the deterministic and stochastic models of the Schlögl chemical system. We plot the distance function against two parameter pairs, rate constants k_1-k_3 and k_2-k_4, in Figure 3. From the colour map, we see that the distance function (3) possesses a well distinguishable global minimum at the true values ($k_1 = 2.5 \times 10^{-4}$, $k_2 = 0.18$, $k_3 = 2250$ and $k_4 = 37.5$). This indicates that the stochastic model is identifiable in both cases. In the deterministic scenario, the Schlögl system loses its identifiability. When the distance function (3) only fits the mean concentration, the minimal values are attained on a curve in the 2D parameter space (the distance function is indicated by blue contour lines in Figure 3). Stochastic models are advantageous in model identifiability, because they can be parametrized using a wider class of statistical properties (typically, K quantities are needed to estimate K reaction rate constants for mass-action reaction systems). The TPA enables efficient and direct evaluation of $J(\hat{S}, S^*)$ all over the parameter space in a single computation by using the representation (2).

Figure 3 also reveals the differences between the model responses to parameter perturbations. The green contour lines show the landscape of $J(\hat{S}, S^*)$ for the stochastic model using only the mean values, i.e., $L = 1$ in (3). The minimum is attained on a straight line, representing another non-identifiable situation. This line (green) has a different direction than the line obtained for the deterministic model (blue). In particular, this example illustrates that the parameter values estimated from deterministic models do not give good approximation of both average behaviour and the noise level when they are used in stochastic models [36].

4 Bifurcation analysis

Bifurcation is defined as a qualitative transformation in the behaviour of the system as a result of the continuous change in model parameters. Bifurcation analysis of ODE systems has been used to understand the properties of deterministic models of biological systems, including models of cell cycle [37] and circadian rhythms [38]. Software packages, implementing numerical bifurcation methods for ODE systems, have also been presented in the literature [39, 40], but computational methods for bifurcation analysis of corresponding stochastic models are still in development [19]. Here, we use the tensor-structured data $p(\mathbf{x}|\mathbf{k})$ given by (2) for a model of fission yeast cell cycle control developed by Tyson [24], and perform the tensor-structured bifurcation analysis on the tensor data. The interaction of cyclin-cdc2 in the Tyson model is illustrated in Figure 4(a). Reactions and parameter values are given in *SI Appendix* S2.2.

The parameter k_1, indicating the breakdown of the active M-phase-promoting factor (MPF), is chosen as the bifurcation parameter. The analysis of the corresponding ODE model reveals that the system displays a stable steady state when k_1 is at its low values, which describes the metaphase arrest of unfertilized eggs [41]. On the other hand, the ODE model is driven into rapid cell cycling exhibiting oscillations when k_1 increases [24]. The ODE cell cycle model has a bifurcation point at $k_1 = 0.2694$, where a limit cycle appears [24]. In our TPA computations, we study the behaviour of the stochastic model for the values of k_1 which are close to the deterministic bifurcation point. We observe that the steady state distribution changes from a unimodal shape (Figure 4(b)) to a probability distribution with a "doughnut-shaped" region of high probability (Figure 4(c)) at $k_1 = 0.3032$. In particular, the stochastic bifurcation appears for higher values of k_1 than the deterministic bifurcation.

In Figure 5, we use the computed tensor-structured parametric probability distribution to visualise the stochastic bifurcation structure of the cell cycle model. As the bifurcation parameter k_1 increases, the
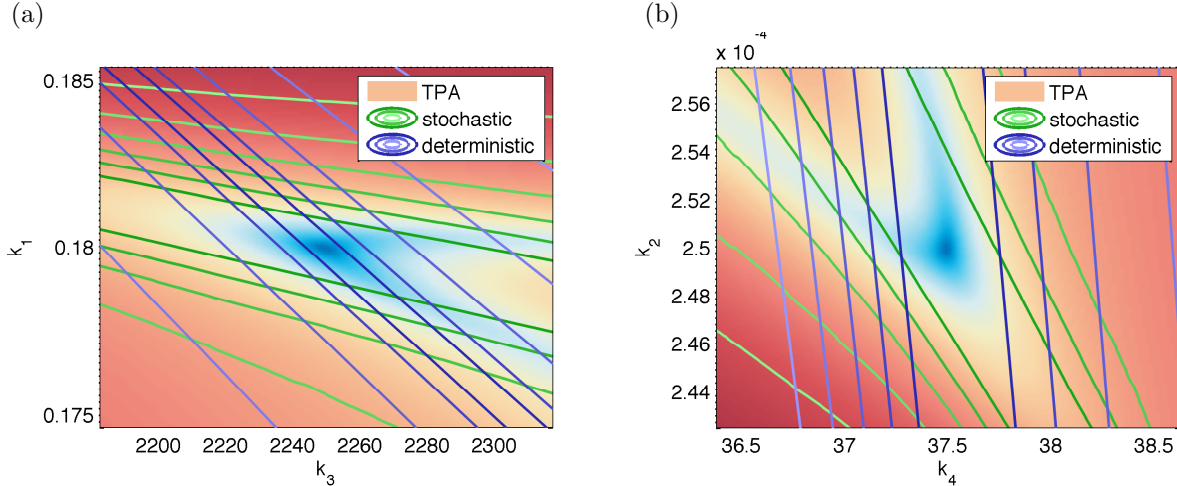
Figure 3: *Parameter identifiability analysis of the Schlögl reaction system.*
(a) *Estimation and identifiability of parameters k_1 and k_3 with k_2 and k_4 fixed at their true values. The colour scale corresponds to values of the distance function (3) with $L = 3$, i.e., the first three moments are compared. The green and blue contour lines indicate the distance functional (3) with the first moment (mean value) only. Green corresponds to the stochastic model and blue to the deterministic model.*
(b) *Estimation and identifiability of parameters k_2 and k_4 with k_1 and k_3 fixed at their true values. The same quantities as in panel (a) are plotted.*

expected oscillation tube is formed and amplified in the marginalised YP-pM-M state space (see panels (a)–(d) of Figure 5). In panels (e)–(h) of Figure 5, the marginal distribution in the Y-CP-pM subspace is plotted. We see that it changes from a unimodal (Figure 5(e)) to a bimodal distribution (Figure 5(f)). Cell cycle models have been studied in the deterministic context either as oscillatory [24] or bistable [42, 43] systems. In Figure 5, we see that the presented stochastic cell cycle model can appear to have both oscillations and bimodality, when diffferent subsets of observables are considered.

5 Robustness analysis

GRNs are subject to extrinsic noise which is manifested by fluctuations of parameter values [44]. This extrinsic noise originates from interactions of the modelled system with other stochastic processes in the cell or its surrounding environment. We can naturally include extrinsic fluctuations under the tensor-structured framework. For a GRN as in (1), we consider the copy numbers $X_1, X_2, \ldots, X_N$ as intrinsic variables and reaction rates $k_1, k_2, \ldots, k_M$ as extrinsic variables. Total stochasticity is quantified by the stationary distribution of the intrinsic variables, $p(\mathbf{x})$. We assume that the invariant probability density of extrinsic variables, $q(\mathbf{k})$, does not depend on the values of intrinsic variables $\mathbf{x}$. Then the law of total probability implies that the stationary probability distribution of intrinsic variables is given by

$$p(\mathbf{x}) = \int_{\Omega_{\mathbf{k}}} p(\mathbf{x}|\mathbf{k}) \, q(\mathbf{k}) \, d\mathbf{k}, \tag{7}$$

where $\Omega_{\mathbf{k}}$ is the parameter space and $p(\mathbf{x}|\mathbf{k})$ represents the invariant density of intrinsic variables conditioned on constant values of kinetic parameters, see the definition below equation (1). If distributions $q(\mathbf{k})$ of extrinsic variables can be determined from high quality experimental data then the stationary density can be computed directly by (7). If not, the TPA framework enables to test the behaviour of GRNs for different hypothesis about the distribution of the extrinsic variables. The advantage of the TPA is that it efficiently computes the high-dimensional integrals in (7), see *SI Appendix* S1.4.
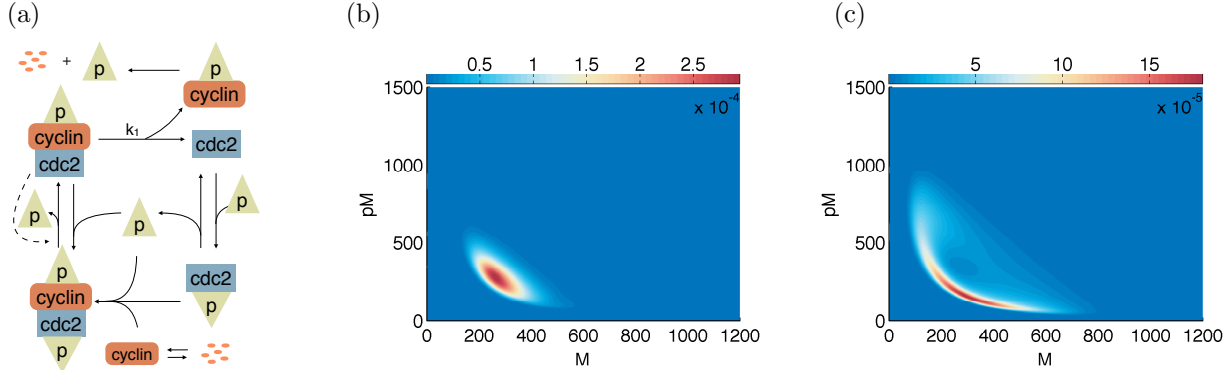
8

Figure 4: *Bifurcation analysis of the stochastic cell cycle model.*
(a) *Schematic description of the cyclin-cdc2 interactions. Free cyclin molecules combine rapidly with phosphorylated cdc2, to form the dimer MPF (cdc2-cyclin-p), which is immediately inactivated by phosphorylation process. The inactive MPF (p-cdc2-cyclin-p) can be converted to active MPF by autocatalytic dephosphorylation. The active MPF in excess breaks down into cdc2 molecules and phosphorylated cyclin, which is later subject to proteolysis. Finally, cdc2 is phosphorylated to repeat the cycle.*
(b) *Joint stationary distribution of cdc2-cyclin-p (M) and p-cdc2-cyclin-p (pM) plotted at the deterministic bifurcation point ($k_1 = 0.2694$).*
(c) *Joint stationary distribution of cdc2-cyclin-p (M) and p-cdc2-cyclin-p (pM) plotted for $k_1 = 0.3032$.*

5.1 Extrinsic noise in FitzHugh-Nagumo model

We consider the effect of extrinsic fluctuations on an activator-inhibitor oscillator with simple negative feedback loop: the FitzHugh-Nagumo neuron model which is presented in Figure 6(a). Self-autocatalytic positive feedback loop activates the X_1 molecules, which are further triggered by the external signal. The species X_2 is enhanced by the feedforward connection and it acts as an inhibitor that turns off the signalling [4]. We perform robustness analysis based on the simulated tensor data in Section 2.1 (summarized on the third line of Table 1). In our computational examples, we assume that $q(\mathbf{k}) = q_1(k_1) q_2(k_1) \ldots q_M(k_M)$, i.e. the invariant distributions of rate constants k_1, k_2, ..., k_M are independent. Then (7) reads as follows

$$p(\mathbf{x}) = \int_{\Omega_{\mathbf{k}}} p(\mathbf{x}|\mathbf{k}) \, q_1(k_1) \cdots q_M(k_M) \, \mathrm{d}\mathbf{k}. \tag{8}$$

Extrinsic variability in the FitzHugh-Nagumo system is studied in four prototypical cases of q_i, $i = 1, 2, \ldots, M$: (i) Dirac delta, (ii) normal, (iii) uniform, and (iv) bimodal distributions, as shown in Figure 6(b). Since these distributions have zero mean, the extrinsic noise is not biased. We can then use this information about extrinsic noise to simulate the stationary probability distribution of intrinsic variables by (8).

When the extrinsic noise is omitted, the inhibited and excited states are linked by a volcano-shaped oscillatory probability distribution (Figure 6(c)). At the inhibited state, X_1 molecules first get activated from the positive feedback loop, and then excite X_2 molecules by feedforward control. The delay between the excitability of the two molecular species gives rise to the path (solid line) describing switching from the inhibited state to the excited state (Figure 6(c)). If the normal or uniform noise are introduced to the extrinsic variables, then the path becomes straighter (Figure 6(d) and Figure 6(e)). This suggests that, once X_1 molecules get excited or inhibited, X_2 molecules require less time to response.

GRNs with stronger negative feedback regulation gain higher potential to reduce the stochasticity. This argument has been both theoretically analysed [45, 46], and experimentally tested for a plasmid-borne system [47]. We have shown that the extrinsic noise reduces the delay caused by the feedback loop (Figure 6(d)). If we further increase the variability of the extrinsic noise, then the delay caused by the feedback loop is further reduced (Figure 6(e)). In the case of the bimodal distribution of extrinsic fluctuations, the most-likely path linking the inhibited and excited states even shrinks into an almost straight line (Figure 6(f)).
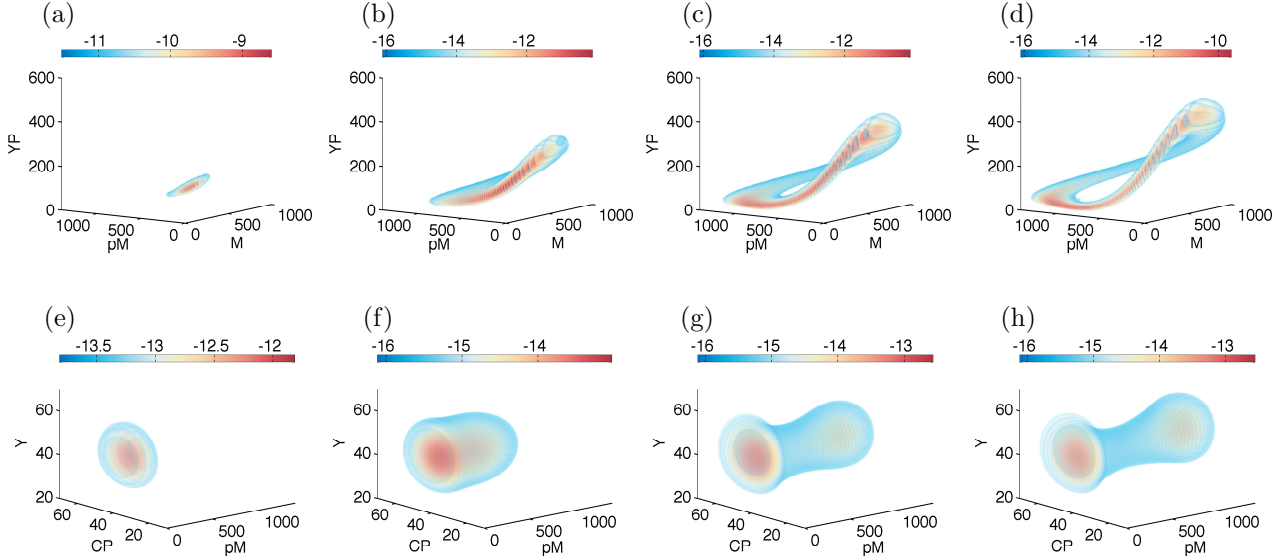
9

Figure 5: *Visualisation of the bifurcation structure of the stochastic cell cycle model.*
(a)–(d) *Marginal steady state distributions of the phosphorylated cyclin (YP), the inactive MPF (pM) and the active MPF (M).*
(e)–(h) *Marginal stationary distributions of the cyclin (Y), the phosphrylated cdc2 (CP) and the inactive MPF (pM).*
Each column corresponds to the same value of the bifurcation parameter k_1, which from the left to the right are 0.25, 0.3, 0.35 and 0.4, respectively. All figures show $\log_2$ of the marginal steady state distribution for better visualisation.

This means that, for the same level of the inhibitor X_2, the number of the activator X_1 is lower, i.e. the presented robustness analysis shows that the behaviour of stochastic GRNs with negative feedback regulation can benefit from the extrinsic noise.

6 Discussion

We have presented the TPA of stochastic reaction networks and illustrated that the TPA can (i) calculate and store the parametric steady state distributions; (ii) infer and analyse stochastic models of GRNs. To explore high dimensional state space $\Omega_{\mathbf{x}}$ and parameter space $\Omega_{\mathbf{k}}$, the TPA utilises a recently proposed low-parametric tensor-structured data format, as presented in equation (2). Tensor methods have been recently used to address the computational intensity of solving the CME [16, 48]. In this paper, we have extended these tensor-based approaches from solving the underlying equations to automated parametric analysis of the stochastic reaction networks. One notable advantage of the tensor approach lies in its ability to capture all probabilistic information of stochastic models all over the parameter space into one single tensor-formatted solution, in a way that allows linear scaling of basic operations with respect to the number of dimensions. Consequently, the existing algorithms commonly used in the deterministic framework can be directly utilized in stochastic models via the TPA. In this way, we can improve our understanding of parameters in stochastic models.

To overcome technical (numerical) challenges, we have introduced two main approaches for successful computation of the steady state distribution. First, we compute it using the CFPE approximation which provides additional flexibility in discretising the state space $\Omega_{\mathbf{x}}$. The CFPE admits larger grid sizes for numerical simulations than the unit grid size of the CME. In this way, the resulting discrete operator is better conditioned. We illustrate this using a 20-dimensional problem introduced in the last line of Table 1
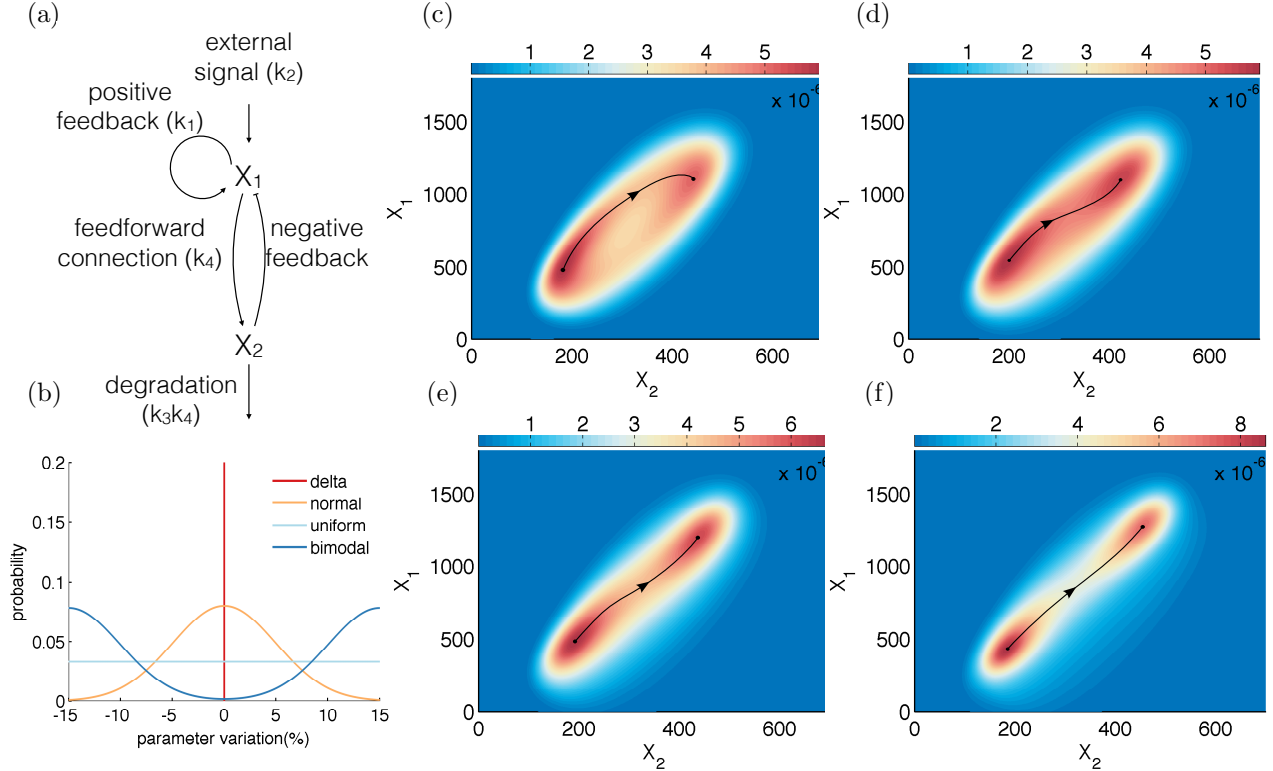
Figure 6: *Analysis of the FitzHugh-Nagumo system.* (a) *Schematic of the model.*
(b) *Four types of distributions of the extrinsic noise applied to model parameters.*
(c)–(f) *Steady state behaviour of the FitzHugh-Nagumo model with* (c) *constant reaction rates,* (d) *normal,*
(e) *uniform, and* (f) *bimodal distribution of the model parameters. The tensor-structured computational procedure follows formula* (8). *The directed black curves track the ridges of the stationary distribution, and depict the "most-likely" transition paths from the inhibited state to the excited state.*

and in *SI Appendix* S2.4. To compute the stationary distribution, a multilevel approach is implemented, where the steady state distribution is first approximated on a coarse grid, and then interpolated to a finer grid as the initial guess (see *SI Appendix* S1.3 for more details). The results are plotted in Figure 7. Second, we introduce the adaptive inverse power iteration scheme tailored to current tensor solvers of linear systems, see *SI Appendix* S1.3 for technical details. Since tensor linear solvers are less robust especially for ill-conditioned problems, it is necessary to carefully adapt the shift value during the inverse power iterations in order to balance the conditioning and sufficient speed of the convergence.

Techniques for the parameter inference and bifurcation analysis of stochastic models have been less studied in the literature than the corresponding methods for the ODE models. One of the reason for this is that the solution of the CME is more difficult to obtain than solutions of mean-field ODEs. This has been partially solved by the widely-used Monte Carlo methods, such as the Gillespie SSA, which can be used to estimate the required quantities [29]. Advantages of Monte Carlo methods are especially their relative simplicity and easy parallelisation. The TPA provides an alternative approach. The TPA uses more complex data structures and algorithms than the Gillespie SSA, but it enables to compute the whole probability distribution for all combinations of parameter values at once. The TPA stores this information in the tensor format. If the state and parameter spaces have higher number of dimensions, then the Monte Carlo methods would have problems with storing computed stationary distributions. Another advantage of the TPA is that it produces smooth data, see e.g. Figure 3 for the data over the parameter space and Figure 5 for the data in the state space. This is important for a stable convergence in the gradient-based optimization algorithms [49], and
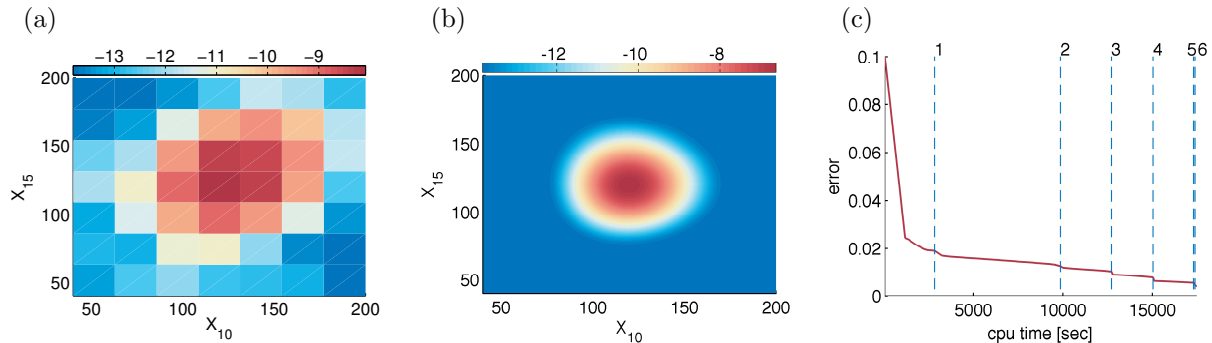
Figure 7: *The computation of the stationary distribution using the TPA for a 20-dimensional reaction chain. The CFPE is successively solved on seven grid levels with an increasing number of nodal points. The marginal stationary distribution in the X_{10}-X_{15} plane computed on* (a) *the initial coarsest level;*
(b) *the finest grid level.*
(c) *The convergence of the total error versus the computational time. The vertical dashed lines correspond to the grid levels. The grid size details on each grid level are given in SI Appendix S2.4 (Table S8).*

for reliable analysis of stochastic bifurcations. Monte Carlo methods provide necessarily noisy and hence nonsmooth data that may cause problems for these methods.

Parameter inference of stochastic models can make use of various statistical measures, such as the variance and correlations. Monte Carlo approaches are widely used to compute these quantities, but they may be computationally expensive. The TPA provides an alternative approach. Once we compute the stationary distribution for the desired ranges of parameter values and store it in the tensor format, we can use the tensor operation techniques (see *SI Appendix* S1.4) to efficiently compute many different statistical measures from the same stationary distribution. If the results of the used statistical measure and chosen method are not satisfactory, we can modify or completely change both and try to infer the parameters again. Since the stationary distribution is stored, the modifications and changes can be done with low computational load. Namely, no stochastic simulations are needed. In addition, since the stationary distribution contains complete information about the stochastic steady state, it can be used to compute practically any quantity for comparison with experimental data. We have illustrated several different parametric studies in Figures 1(b), 2, and 3. All these results are based on a single tensor solution presented in Section 2.1 (see Table 1). Let us also note that it is relatively straightforward to use the TPA framework to study the parameter sensitivity of stochastic systems (i.e. to quantify the dependence of certain quantities of interest on continuous changes in model parameters). A systematic way for conducting the sensitivity analysis is illustrated in *SI Appendix* S1.5, using the fission yeast cell cycle model.

References

[1] Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nature Genetics* 31:69–73.

[2] Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences* 99:12795–12800.

[3] Li C, et al. (2010) Biomodels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology* 4:92.

[4] Tsai TYC, et al. (2008) Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* 321:126–129.

[5] Novák B, Tyson JJ (2008) Design principles of biochemical oscillators. *Nature Reviews Molecular Cell Biology* 9:981–991.

[6] Kitano H (2004) Biological robustness. *Nature Reviews Genetics* 5:826–837.

[7] Brophy JA, Voigt CA (2014) Principles of genetic circuit design. *Nature Methods* 11:508–520.

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

[9] McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences* 94:814–819.

[10] Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* 81:2340–2361.

[11] Cao Y, Gillespie DT, Petzold LR (2004) The slow-scale stochastic simulation algorithm. *Journal of Chemical Physics* 122:014116.

[12] Gibson MA, Bruck J (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A* 104:1876–1889.

[13] Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences* 104:2785–2790.

[14] Beylkin G, Mohlenkamp MJ (2002) Numerical operator calculus in higher dimensions. *Proceedings of the National Academy of Sciences* 99:10246–10251.

[15] Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of Mathematical Biology* 54:1–26.

[16] Dolgov SV, Khoromskij BN, Oseledets IV (2012) Fast solution of parabolic problems in the tensor train/quantized tensor train format with initial application to the Fokker–Planck Equation. *SIAM Journal on Scientific Computing* 34(6):A3016–A3038.

[17] Khoromskij BN, Schwab C (2011) Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs. *SIAM Journal on Scientific Computing* 33(1):364–385.

[18] Gillespie DT (1996) The multivariate Langevin and Fokker–Planck equations. *American Journal of Physics* 64:1246–1257.

[19] Erban R, Chapman SJ, Kevrekidis IG, Vejchodský T (2009) Analysis of a stochastic chemical system close to a SNIPER bifurcation of its mean-field model. *SIAM Journal on Applied Mathematics* 70:984–1016.

[20] Pereyra V, Scherer G (1973) Efficient computer manipulation of tensor products with applications to multidimensional approximation. *Mathematics of Computation* 27:595–605.

[21] Carroll JD, Chang JJ (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart–Young" decomposition. *Psychometrika* 35:283–319.

[22] Oseledets IV (2011) Tensor-train decomposition. *SIAM Journal on Scientific Computing* 33:2295–2317.

[23] Schlögl F (1972) Chemical reaction models for non-equilibrium phase transitions. *Zeitschrift für Physik* 253:147–161.

[24] Tyson JJ (1991) Modeling the cell division cycle: cdc2 and cyclin interactions. *Proceedings of the National Academy of Sciences* 88:7328–7332.

[25] Gillespie DT (2002) The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction. *Journal of Physical Chemistry A* 106:5063–5071.

[26] Moles CG, Mendes P, Banga JR. (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research* 13:2467–2474.

[27] Jaqaman K, Danuser G (2006) Linking data to models: data regression. *Nature Reviews Molecular Cell Biology* 7, 813–819.

[28] Pedersen A (1995) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics* 22:55–71.

[29] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface* 6(31):187–202.

[30] Reinker S, Altman RM, Timmer J (2006) Parameter estimation in stochastic biochemical reactions. *IEE Proceedings - Systems Biology* 153:168–178.

[31] Tian T, Xu S, Gao J, Burrage K (2007) Simulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics* 23:84–91.

[32] Lillacci G, Khammash M (2010) Parameter estimation and model selection in computational biology. *PLoS Computational Biology* 6(3):e1000696.

[33] Zechner C, Ruess J, Krenn P, Pelet S, Peter M, Lygeros J, Koeppl H (2012) Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences* 109(21):8340–8345.

[34] Von Dassow G, Meir E, Munro EM, Odell GM (2000) The segment polarity network is a robust developmental module. *Nature* 406:188–192.

[35] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology* 3(10):e189.

[36] Wilkinson DJ (2009) Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics* 10:122–133.

[37] Tyson JJ, Csikász-Nagy A, Novák B (2002) The dynamics of cell cycle regulation. *BioEssays* 24:1095–1109.

[38] olde Scheper T, Klinkenberg D, Pennartz C, van Pelt J (1999) A mathematical model for the intracellular circadian rhythm generator. *Journal of Neuroscience* 19(1):40–47.

[39] Doedel E, Champneys A, Fairgrieve T, Kuznetsov Y, Sandstede B, Wang X (1997) AUTO 97: Continuation And Bifurcation Software For Ordinary Differential Equations (with HomCont).

[40] Dhooge A, Govaerts W, Kuznetsov YA (2003) MATCONT: A MATLAB package for numerical bifurcation analysis of ODEs. *ACM Transactions on Mathematical Software* 29(2):141–164.

[41] Hunt T (1989) Under arrest in the cell cycle. *Nature* 342:483–484.

[42] Pomerening JR, Sontag ED, Ferrell JE (2003) Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nature Cell Biology* 5:346–351.

[43] Xiong W, Ferrell JE (2003) A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* 426:460–465.

[44] Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427:415–418.

[45] Thattai M, Van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences* 98:8614–8619.

[46] Swain PS (2004) Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *Journal of Molecular Biology* 344:965–976.

[47] Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. *Nature* 405:590–593.

[48] Kazeev V, Khammash M, Nip M, Schwab C (2014) Direct solution of the chemical master equation using quantized tensor trains. *PLoS Computational Biology* 10(3):e1003359.

[49] McGill JA, Ogunnaike BA, Vlachos DG (2012) Efficient gradient estimation using finite differencing and likelihood ratios for kinetic Monte Carlo simulations. *Journal of Computational Physics* 231(21), 7170–7186.

Supporting Information Appendix for:

Tensor methods for parameter estimation and bifurcation analysis of stochastic reaction networks

Shuohao Liao[*], Tomáš Vejchodský[†], Radek Erban[*]

S1 Methods

All models studied by the TPA are given in terms of well-mixed chemical systems where the system state changes according to the chemical reactions (1) (here, equation label (1) refers to the corresponding equation in the paper). Probability that a reaction occurs is determined by the propensity function

$$\alpha_j(\mathbf{x}, k_j) = k_j\, \tilde{\alpha}_j(\mathbf{x}), \quad j = 1, 2, \ldots, M, \tag{S1}$$

with the non-parametric part $\tilde{\alpha}_j(\mathbf{x})$ given by

$$\tilde{\alpha}_j(\mathbf{x}) = \exp\left[\left(1 - \sum_{i=1}^{N} \nu_{j,i}^-\right) \log V\right] \prod_{i=1}^{N} (\nu_{j,i}^-)! \binom{x_i}{\nu_{j,i}^-},$$

where V is the volume of the reactor. The stationary distribution $p(\mathbf{x}\,|\,\mathbf{k})$, where $\mathbf{k} = (k_1, k_2, \ldots, k_K)^T$ is the subset of kinetic rate constants for which the parameter analysis is considered, can be computed as the exact solution of the chemical master equation [50]. However, for the computational reasons, we will approximate it by the solution of the stationary chemical Fokker-Planck equation [18], which can be written as

$$\mathcal{A}(\mathbf{x}, \mathbf{k})\, p(\mathbf{x}\,|\,\mathbf{k}) = 0,$$

where

$$
\begin{aligned}
\mathcal{A}(\mathbf{x}, \mathbf{k})\, p(\mathbf{x}\,|\,\mathbf{k}) = &\; -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} \left(\sum_{j=1}^{M} \nu_{j,i}\, \alpha_j(\mathbf{x}, k_j)\, p(\mathbf{x}\,|\,\mathbf{k})\right) \\
&+ \frac{1}{2} \sum_{i,i'=1}^{N} \frac{\partial^2}{\partial x_i \partial x_{i'}} \left(\sum_{j=1}^{M} \nu_{j,i}\, \nu_{j,i'}\, \alpha_j(\mathbf{x}, k_j)\, p(\mathbf{x}\,|\,\mathbf{k})\right),
\end{aligned}
\tag{S2}
$$

is the parametric Fokker-Planck operator. We use the tensor structures to compute $p(\mathbf{x}\,|\,\mathbf{k})$ simultaneously for ranges of values of reaction rates $\mathbf{k}$. To achieve this, we split the model parameters from the state variables in a multiplicative way. Considering the definition of propensity functions (S1), we can split the parametric Fokker-Planck operator (S2) into M terms as

$$\mathcal{A}(\mathbf{x}, \mathbf{k}) = k_1 \mathcal{A}^{[1]}(\mathbf{x}) + \cdots + k_M \mathcal{A}^{[M]}(\mathbf{x}), \tag{S3}$$

where the non-parametric operator $\mathcal{A}^{[j]}(\mathbf{x})$ describes the normalised transition properties of the j-th reaction, and is defined (for any twice differentiable function f) by

$$\mathcal{A}^{[j]}(\mathbf{x})\, f(\mathbf{x}) = -\sum_{i=1}^{N} \nu_{j,i}\, \frac{\partial}{\partial x_i} \left(\tilde{\alpha}_j(\mathbf{x})\, f(\mathbf{x})\right) + \frac{1}{2} \sum_{i,i'=1}^{N} \nu_{j,i'}\, \nu_{j,i} \frac{\partial^2}{\partial x_i \partial x_{i'}} \left(\tilde{\alpha}_j(\mathbf{x})\, f(\mathbf{x})\right). \tag{S4}$$

[*]Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, United Kingdom; e-mails: liao@maths.ox.ac.uk; erban@maths.ox.ac.uk

[†]Institute of Mathematics, Czech Academy of Sciences, Zitna 25, 115 67 Praha 1, Czech Republic; e-mail vejchod@math.cas.cz

Let us note that the definition of propensity functions (S1) relies on the law of mass action. However, the TPA methodology is applicable even for more general definitions [51]. If the propensity functions depend nonlinearly on the kinetic rates, as in [52] for example, then the TPA methodology can be used provided a multiplicative splitting (S3) of the parametric Fokker-Planck operator is possible. Such splitting is always possible if the propensities can be written as a product of two terms, where the first term depends only on kinetic rates, and the second term on state variables.

The splitting of the parametric Fokker-Planck operator (S3) implies a perfect collinear relationship between the kinetic rate parameters. In the context of parameter estimation this means that a single constraint (like mean and variance) restricts the original K-dimensional parameter space to an $(K-1)$-dimensional subspace of parameter values that comply with the constraint. Therefore, if a direct comparison between a model and a sample is not possible, a necessary condition to statistically infer $K \leq M$ parameters of a stochastic system is to define at least K constraints.

S1.1 Tensorization

We consider the state variable $\mathbf{x}$ in a bounded domain $\Omega_{\mathbf{x}} \subset (0, \infty)^N$. Similarly, the kinetic rates $\mathbf{k}$, which are varied during the parametric analysis, are considered in $\Omega_{\mathbf{k}} \subset [0, \infty)^K$ where $K \leq M$. In order to utilize the tensor structures, we assume that $\Omega_{\mathbf{x}} = \mathcal{I}_1 \times \cdots \times \mathcal{I}_N$ and $\Omega_{\mathbf{k}} = \mathcal{J}_1 \times \cdots \times \mathcal{J}_K$, where $\mathcal{I}_d = (a_d^x, b_d^x)$, $d = 1, 2, \ldots, N$, are open intervals and $\mathcal{J}_\ell = [a_\ell^k, b_\ell^k]$, $l = 1, 2, \ldots, K$, are closed intervals.

We consider homogeneous Dirichlet boundary conditions on the boundary of $\Omega_{\mathbf{x}}$. We approximate the stationary distribution by the (normalized) eigenfunction of $\mathcal{A}(\mathbf{x}, \mathbf{k})$ corresponding to the eigenvalue closest to zero. Since the Fokker-Planck operator $\mathcal{A}(\mathbf{x}, \mathbf{k})$ is an elliptic operator, the largest eigenvalue converges to 0 from below as the size of $\Omega_{\mathbf{x}}$ increases to infinity. In particular, if we choose a sufficiently large computational domain $\Omega_{\mathbf{x}}$, then the largest eigenvalue will be close to zero and the Dirichlet boundary conditions will not cause any substantial error.

The chemical Fokker-Planck operator (S4) is discretized in $\Omega_{\mathbf{x}}$ by the finite difference method. We consider tensor grids [53] in both $\Omega_{\mathbf{x}}$ and $\Omega_{\mathbf{k}}$. The tensor grid in $\Omega_{\mathbf{x}}$ has nodes $(x_{1,i_1}, \ldots, x_{N,i_N})$, $i_d = 1, 2, \ldots, n_d$, $d = 1, 2, \ldots, N$. There are n_d points $x_{d,i_d} = a_d^x + i_d h_d^x$, $i_d = 1, 2, \ldots, n_d$, in every $\mathcal{I}_d$ with the grid size $h_d^x = (b_d^x - a_d^x)/(n_d + 1)$, $d = 1, 2, \ldots, N$. Similarly, we define tensor grid $(k_{1,j_1}, \ldots, k_{K,j_K})$ in $\Omega_{\mathbf{k}}$, where $k_{\ell,j_\ell} = a_\ell^k + (j_\ell - 1)h_\ell^k$, $j_\ell = 1, 2, \ldots, m_\ell$, form a uniform partition of $\mathcal{J}_\ell$ with the grid size $h_\ell^k = (b_\ell^k - a_\ell^k)/(m_\ell - 1)$, $\ell = 1, 2, \ldots, K$. Note that the boundary points a_d^x and b_d^x, $d = 1, 2, \ldots, N$, are not present in the tensor grid due to the Dirichlet boundary conditions.

The values of the stationary distribution $p(\mathbf{x} \,|\, \mathbf{k})$ at the nodal points are organized as an $(N + K)$-dimensional tensor $\bar{\mathbf{p}} \in \mathbb{R}^{n_1 \times \cdots \times n_N \times m_1 \times \cdots \times m_K}$ with entries

$$\bar{\mathbf{p}}_{i_1, \ldots, i_N, j_1 \ldots, j_K} = p(x_{1,i_1}, \ldots, x_{N,i_N} | k_{1,j_1}, \ldots, k_{K,j_K}). \tag{S5}$$

In the traditional matrix-vector approach, we would organize the entries of $\bar{\mathbf{p}}$ into a long vector. However, the tensor structure is more natural, because it corresponds to the original physical position of the nodes within the state and parameter space [54]. Finally, let us note that if $n = n_1 = \cdots = n_N$ and $m = m_1 = \cdots = m_K$ then there is $n^N m^K$ entries in the tensor $\bar{\mathbf{p}}$. Thus, the number of memory places to store the tensor $\bar{\mathbf{p}}$ grows exponentially with N and K. In the next subsection, we present the main idea of the separated representation of tensors that allows to solve this problem.

S1.2 Separation of dimensions

The main idea of the separated (or low-parametric) representation is to approximate tensor $\bar{\mathbf{p}}$ by the following sum of rank-one tensors:

$$\bar{\mathbf{p}} \approx \sum_{r=1}^{R} \underbrace{\phi_1^{[r]} \otimes \cdots \otimes \phi_N^{[r]}}_{\text{state space}} \otimes \underbrace{\psi_1^{[r]} \otimes \cdots \otimes \psi_K^{[r]}}_{\text{parameter space}}, \tag{S6}$$

where $\phi_d^{[r]} \in \mathbb{R}^{n_d}$, $d = 1, 2, \ldots, N$, and $\psi_\ell^{[r]} \in \mathbb{R}^{m_\ell}$, $\ell = 1, 2, \ldots, K$, are factor vectors, R is known as the separation rank, and symbol $\otimes$ denotes the tensor product of vectors [55]. Let us recall that the tensor product $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_N$ of vectors $\mathbf{v}_d \in \mathbb{R}^{n_d}$, $d = 1, 2, \ldots, N$, is defined as a tensor $\bar{\mathbf{v}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ with entries $v_{i_1, i_2, \ldots, i_N} = v_{1, i_1} v_{2, i_2} \cdots v_{N, i_N}$.

Representation (S6) has the potential to solve high-dimensional problems. Indeed, if we consider for simplicity $n = n_1 = \cdots = n_N$ and $m = m_1 = \cdots = m_K$ then the representation (S6) requires to store $(nN + mK)R$ numbers only. For moderate values of R this is substantially less than the number of entries of $\bar{\mathbf{p}}$. Moreover, low-parametric representations such as (S6) enable to perform algebraic operations in an efficient way, see Section S1.4.

The accuracy of the separated representation (S6) depends on the choice of the factor vectors and on the size of the tensor rank R. Clearly, the higher rank enables higher accuracy, but requires higher computational and storage costs. In practical computations, the rank R is dynamically controlled using algorithms for tensor truncation, see Section S1.3. Let us note that the representation (S6) is known as the canonical polyadic decomposition [56]. However, due to reasons connected with the stability of the tensor truncation algorithms, it is not suitable for actual computation and more stable tensor formats have to be employed [22]. We have introduced the canonical polyadic decomposition (S6) due to its simplicity to illustrate the main idea of the separate representation of tensors.

For certain simple problems, like birth-death process, the separable representation of the stationary distribution can be derived explicitly. However, in general, we have to compute the stationary distribution in the form (S6). To achieve this, we need to express the discretized Fokker-Planck operator in a separable form as well. Based on the structure of $\mathcal{A}(\mathbf{x}|\mathbf{k})$ in (S3), the discretization of the parametric Fokker-Planck operator can be divided into two steps: decomposing the non-parametric part (see Section S1.2.1) and the parametric part (see Section S1.2.2).

S1.2.1 Decomposition of the non-parametric part

We use the finite differences to discretize the derivatives in the non-parametric operators $\mathcal{A}^{[j]}(\mathbf{x})$ in (S4), see e.g [57]. The separated tensor representation does not require high-dimensional difference stencils. Instead, just one-dimensional differences are needed. Further, since the standard finite difference discretizations of differential operators yield matrices, we organize their entries naturally into tensors. In this situation we speak about tensor matrices and denote them in capital bold font. The idea is exactly the same as in (S5), where we organized a long vector into a tensor.

Thus, the finite difference matrix approximating the non-parametric operator $\mathcal{A}^{[j]}(\mathbf{x})$ in (S4) can be expressed as the following tensor matrix:

$$\mathbf{A}^{[j]} = -\sum_{i=1}^{N} \nu_{j,i} \mathbf{G}^{[i;j]} + \frac{1}{2} \sum_{i,i'=1}^{N} \nu_{j,i} \nu_{j,i'} \mathbf{F}^{[i,i';j]}, \quad j = 1, 2, \ldots, M, \tag{S7}$$

where tensor matrices $\mathbf{G}^{[i;j]}$ and $\mathbf{F}^{[i,i';j]}$ refer to tensor-structured discretizations of the summands in the first and second sums in (S4), respectively, and are determined by

$$\mathbf{G}^{[i;j]} = \tilde{v}_j \, H_1^{[j]} \otimes \cdots \otimes D_i H_i^{[j]} \otimes \cdots \otimes H_N^{[j]},$$
$$\mathbf{F}^{[i,i';j]} = \tilde{v}_j \, H_1^{[j]} \otimes \cdots \otimes D_i H_i^{[j]} \otimes \cdots \otimes D_{i'} H_{i'}^{[j]} \otimes \cdots \otimes H_N^{[j]}, \quad \text{for} \quad i < i',$$
$$\mathbf{F}^{[i,i;j]} = \tilde{v}_j \, H_1^{[j]} \otimes \cdots \otimes D_i D_i H_i^{[j]} \otimes \cdots \otimes H_N^{[j]},$$

where the volume scaling coefficient is $\tilde{v}_j = \exp\left[(1 - \sum_{i=1}^{N} \nu_{j,i}^-) \log V\right]$. Here, $H_i^{[j]} \in \mathbb{R}^{n_i \times n_i}$ and $D_i \in \mathbb{R}^{n_i \times n_i}$ for $i = 1, \ldots, N$ and $j = 1, \ldots, M$ are matrices and, thus, the tensor product $\otimes$ works in the same way as the Kronecker product. Matrix D_i is the central difference matrix with entries $-1/(2h_i^x)$ and $1/(2h_i^x)$ distributed along its super- and sub-diagonal, respectively. Matrix $H_i^{[j]}$ is diagonal with diagonal entries

$$H_i^{[j]}(\ell, \ell) = (\nu_{j,i}^-)! \binom{x_{i,\ell}}{\nu_{j,i}^-} \quad \text{for} \quad \ell = 1, 2, \ldots, n_i.$$

3

We observe that tensor matrices $\mathbf{G}^{[i;j]}$ and $\mathbf{F}^{[i,i';j]}$ are expressed in a separated representation similar to (S6) with the separation rank $R = 1$. Consequently, the non-parametric operator $\mathbf{A}^{[j]}$ in (S7) admits separable representation of rank $R = N(N+1)/2 + N = N^2/2 + 3N/2$. Thus, any further algebraic operation on $\mathbf{A}^{[j]}$ would contribute to the overall complexity growing quadratically in terms of number of chemical species.

S1.2.2 Decomposition of the parametric part

Having the low-parametric discrete tensor-structured representations (S7) of the non-parametric operators $\mathbf{A}^{[j]}$, we write a discrete tensor-structured representation of the parametric Fokker-Planck operator (S3) as

$$\mathbf{A} = \mathbf{A}^{[1]} \otimes K_1 \otimes I_2 \otimes \cdots \otimes I_M + \mathbf{A}^{[2]} \otimes I_1 \otimes K_2 \otimes \cdots \otimes I_M + \cdots + \mathbf{A}^{[M]} \otimes I_1 \otimes I_2 \otimes \cdots \otimes K_M, \quad \text{(S8)}$$

where $K_j \in \mathbb{R}^{m_j \times m_j}$ denotes a diagonal matrix whose diagonal entries correspond to the grid nodes of the j-th parameter, i.e., $K_j(\ell, \ell) = k_{j,\ell}$ for $\ell = 1, 2, \ldots, m_j$ and $j = 1, \ldots, M$.

Equation (S8) is a low-parametric tensor representation of the discretized parametric Fokker-Planck operator with separation rank $M(N^2/2 + 3N/2)$. This rank grows linearly with the number of chemical reactions M and quadratically with the number of chemical species N. Then, the parametric steady state distribution of the form (S6) is solved as the eigenvector of $\mathbf{A}$ corresponding to the eigenvalue closest to zero (see Section S1.3).

S1.3 Solving the stationary CFPE in tensor format

Let $\mathbf{A}$ be the tensor-structured parametric Fokker-Planck operator assembled in (S8). Our goal is to approximate the stationary distribution by the eigenvector $\bar{\mathbf{p}}$ corresponding to the eigenvalue $\lambda_{\min}$ which is closest to zero, i.e.

$$\mathbf{A}\bar{\mathbf{p}} = \lambda_{\min}\bar{\mathbf{p}}. \quad \text{(S9)}$$

A standard method to find the required eigenpair of $\mathbf{A}$ is the inverse power method, and here we modify the original algorithm for better implementations in tensor-structured computations.

Adaptive inverse power algorithm. The main building block is the fact that, beginning with an initial guess $\bar{\mathbf{p}}_0$ and given a shift value σ, the inverse power scheme,

$$(\mathbf{A} - \sigma\mathbf{I})\bar{\mathbf{p}}_{k+1} = \frac{\bar{\mathbf{p}}_k}{\|\bar{\mathbf{p}}_k\|}, \quad k = 0, 1, \ldots \quad \text{(S10)}$$

would converge to the eigenvector corresponding to the eigenvalue closest to the chosen shift σ, provided that the eigenvalue is of multiplicity one. Since all eigenvalues of the Fokker-Planck operator have negative real parts, we choose $\sigma \geq 0$. We do it adaptively based on the performance of the tensor linear solver, i.e. $\sigma \equiv \sigma_k$ in (10).

We apply the alternating minimum energy method (AMEN) [58] to solve the linear system (S10). Given an initial N-dimensional tensor $\bar{\mathbf{p}}$, the AMEN method minimises the residual in a single dimension at a time with other dimensions fixed, and alternates the dimensions from 1 to N. The entire *sweep* repeats until a convergence criterion is satisfied. Typically, smaller shift σ makes the whole inverse power method converge faster to the steady state solution, however, within each inverse iteration (S10), the AMEN may require many sweeps to achieve a reasonable tolerance. Thus, our strategy is to double the shift value σ when the solver reach certain upper threshold, and half σ to seek for better convergence for the whole procedure when the AMEN converges with only a few sweeps.

Another extension arises from a feature of tensor-structured data format. The tensor separation rank R can increase rapidly over successive algebraic operations, making the representation untenable. To avoid uncontrollable growth of the separation rank throughout the computation, we need to reduce it by adaptively changing the involving factor vectors while maintaining the required accuracy. This procedure is usually called tensor truncation:

$$\bar{\mathbf{p}}^* = \Gamma(\bar{\mathbf{p}}), \quad \text{(S11)}$$

4

where operator Γ is the truncation operator, and $\text{rank}(\bar{\mathbf{p}}^*) < \text{rank}(\bar{\mathbf{p}})$. Although finding the optimal tensor separation rank is still an open question of the ongoing research, the tensor train format, together with its SVD-based tensor truncation algorithm, is a stable and useful prototype for our implementations, and we refer the readers to [22] for further details.

Consequently, the adaptive inverse power method used in the TPA is summarised as follows:

Step 0. Initialize: initial guess $\bar{\mathbf{p}}_0$; shift value $\sigma = \sigma_0$; stopping criteria ε; maximum number of AMEN sweeps in each inverse iteration N_{max}; thresholds to increase (N_{in}) and decrease (N_{de}) the shift value.

Step 1. Solve the k-th tensor-structured inverse iteration (S10) up to N_{max} sweeps.

Step 2. Check the number of sweep N_{comp} for the AMEN solver to converge:

    2a. If $N_{comp} > N_{in}$, let $\sigma = 2\sigma$ and jump back to Step 1.

    2b. If $N_{de} < N_{comp} \leq N_{in}$, go to Step 3.

    2c. If $N_{comp} \leq N_{de}$, $\sigma = \sigma/2$ and go to Step 3.

Step 3. Truncate the tensor separation rank as in (S11).

Step 4. Check the stopping criteria:

    4a. If $\|\bar{\mathbf{p}}_{k+1} - \bar{\mathbf{p}}_k\| > \varepsilon$, let $\bar{\mathbf{p}}_{k+1} = \bar{\mathbf{p}}_k$ and $k = k + 1$, and jump to Step 1.

    4b. If $\|\bar{\mathbf{p}}_{k+1} - \bar{\mathbf{p}}_k\| \leq \varepsilon$, return $\bar{\mathbf{p}}_{k+1}$ and exit.

Multi-level acceleration. When the dimensionality of the problem is large, the adaptive scheme discussed above may converge slowly, because on a fixed grid size, the AMEN requires very large shift value σ to solve (S10). Thus, the TPA makes use of a multilevel scheme to accelerate the solution process for high-dimensional problems. The system (S9) is first solved on a coarse grid with grid size $2h$. The approximated stationary solution is then interpolated to a fine grid with grid size h and used as an initial guess. The method continues to solve the system on finer grids until some convergence criteria are achieved.

A key step in the multilevel approach is the interpolation, or prolongation, matrix that transfers the solution on a coarse grid to a fine grid. The prolongation operator has a rank-one tensor structure. Let N-dimensional tensor $\bar{\mathbf{p}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ contain the function values on an N-dimensional tensor grid with n_k, $k = 1, 2, \ldots, N$, grid points along each direction. The prolongation operator $\mathbf{P}^{[k]}$ to the k-th dimension is then defined as

$$\mathbf{P}^{[k]} = I \otimes \cdots \otimes I \otimes \underbrace{P_{n_k}^{2n_k}}_{k\text{-th mode}} \otimes I \otimes \cdots \otimes I, \tag{S12}$$

where $P_{n_k}^{2n_k} \in \mathbb{R}^{2n_k \times n_k}$ is the one-dimensional interpolation matrix defined by

$$P_{n_k}^{2n_k} = \frac{1}{2} \begin{pmatrix} 2 & & & \\ 1 & 1 & & \\ & 2 & & \\ & 1 & 1 & \\ & & & \ddots \end{pmatrix}.$$

If tensor $\bar{\mathbf{p}}$ has the rank-R separated representation as (S6) with $n = n_1 = n_2 = \cdots = n_N$, the complexity to interpolate a single dimension is $\mathcal{O}(n)$, and the total complexity of a full interpolation over N-dimensional tensor grid is $\mathcal{O}(nN)$. We summarise the multi-level accelerated adaptive inverse power method as follows:

Step 0. Initialize: initial grid size on the coarsest grid h_1; initial guess $\bar{\mathbf{p}}_0^{(1)}$; initial error tolerance $\varepsilon^{(1)}$; maximum number of grid levels L_{max}; and let $\ell = 1$.

Step 1. Solve the eigenvalue problem (S9) on the ℓ-th level with initial guess $\bar{\mathbf{p}}_0^{(\ell)}$, using the adaptive inverse power method. Return the solution $\bar{\mathbf{p}}_k^{(\ell)}$ that satisfies the error tolerance $\|\bar{\mathbf{p}}_k^{(\ell)} - \bar{\mathbf{p}}_{k-1}^{(\ell)}\| \leq \varepsilon^{(\ell)}$.

Step 2. If $\ell < L_{max}$, interpolate the solution $\bar{\mathbf{p}}_k^{(\ell)}$ to a finer grid by successive application of the prolongation operator $\mathbf{P}^{[k]}$ in (S12) to each dimension. Let $\bar{\mathbf{p}}_0^{(\ell+1)} = \bar{\mathbf{p}}_k^{(\ell)}$, $\varepsilon^{(\ell+1)} = \varepsilon^{(\ell)}/2$, $\ell = \ell + 1$. Go to Step 1.

Step 3. If $\ell = L_{max}$, return the solution $\bar{\mathbf{p}}_k^{(\ell)}$ and exit.

Multilevel approach is used in Section S2.4 to analyse the 20-dimensional chemical system (S17). CPU times for each grid size are shown in Table S8. In general, the operators of both the CME and CFPE are non-symmetric and ill-conditioned, and challenging to handle using tensor-structured solvers [59, 60]. Although it can be improved by shortening the time-step [48], the CFPE has a distinctive advantage over the CME for its flexibility in choosing the grid size, which enables to control the accuracy and use of acceleration strategies, such as the presented multilevel approach.

Implementation. The TPA, implemented in MATLAB, is included in the Stochastic Bifurcation Analyzer toolbox available at http://www.stobifan.org. The source code relies on the Tensor Train Toolbox [22]. Simulations are performed on a 64-bit Linux desktop equipped with Quad-Core AMD Opteron™ Processor 8356 × 16 and 63 GB RAM.

S1.4 Elementary tensor operations

The computation of the tensor-structured parametric solution $\bar{\mathbf{p}}$ has been described in Section S1.3. In this section, we discuss computational details of post-processing the solution in the form (S6) for parametric analysis. This analysis is based on high-dimensional integration, implemented using the k-mode product described below.

Tensor multiplication: the k-mode product [61]. Let $\bar{\mathbf{p}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_N}$ be an N-dimensional tensor, the k-mode product of $\bar{\mathbf{p}}$ with a vector $\mathbf{q} \in \mathbb{R}^{n_k}$ is denoted by $\bar{\mathbf{p}} \times_k \mathbf{q}$ and is a tensor of size $n_1 \times n_{k-1} \times 1 \times n_{k+1} \times \cdots \times n_N$. Elementwise, we have

$$(\bar{\mathbf{p}} \times_k \mathbf{q})_{i_1,\ldots,i_{k-1},1,i_{k+1},\ldots,i_N} = \sum_{j_k=1}^{n_k} \bar{\mathbf{p}}_{i_1,\ldots,i_{k-1},j_k,i_{k+1},\ldots,i_N} \, q_{j_k}. \tag{S13}$$

Further, if $\bar{\mathbf{p}}$ can be written as a rank-R tensor, i.e., $\bar{\mathbf{p}} = \sum_{r=1}^{R} \phi_1^{[r]} \otimes \cdots \otimes \phi_N^{[r]}$, then the k-mode product can be evaluated through R one-dimensional inner products:

$$\bar{\mathbf{p}} \times_k \mathbf{q} = \sum_{r=1}^{R} \phi_1^{[r]} \otimes \cdots \otimes \phi_{k-1}^{[r]} \otimes \langle \phi_k^{[r]}, \mathbf{q} \rangle \otimes \phi_{k+1}^{[r]} \otimes \cdots \otimes \phi_N^{[r]}.$$

The $(i_1, i_2, \ldots, i_N)$-th order moment computation in equation (5). For tensor-structured parametric solution in (S6), integral (5) can be simultaneously approximated for all parameter sets through successive application of the mode product introduced in (S13) as

$$\mu_{[i_1,\ldots,i_N]}(\mathbf{k}^*) \approx h_1^x h_2^x \cdots h_N^x \left(\bar{\mathbf{p}} \times_1 \mathbf{x}_1^{i_1} \times_2 \mathbf{x}_2^{i_2} \times_3 \cdots \times_N \mathbf{x}_N^{i_N} \right), \tag{S14}$$

where $\mathbf{x}_d^{i_d} = (x_{d,1}^{i_d}, x_{d,1}^{i_d}, \ldots, x_{d,n_d}^{i_d})^T$ and h_d^x is the grid size, defined in Section S1.1. The computational complexity of (S14) is $\mathcal{O}(nNR)$, where $n = \max\{n_1, n_2, \ldots, n_d\}$.

6

Computation of integral in equation (8) in the paper. Given the distributions of parameters $q_j(k_j)$ for $j = 1, \ldots, K$, the integral in equation (8) can be efficiently computed using the tensor-structured solution (S6) by

$$\bar{\mathbf{p}}_{\mathbf{x}} = h_1^k h_2^k \ldots h_K^k \left(\bar{\mathbf{p}} \times_{N+1} \mathbf{q}_1 \times_{N+2} \mathbf{q}_2 \times_{N_3} \cdots \times_{N+K} \mathbf{q}_K \right), \tag{S15}$$

where the entries of vectors $\mathbf{q}_j$ for $j = 1, \ldots, K$, represent the values of $q_j(k_j)$ at the discrete node points $k_{j,1}, k_{j,2}, \ldots, k_{j,m_j}$ and $h_1^k, h_2^k, \ldots, h_K^k$ are grid sizes in the parameter space, defined in Section S1.1. If $m = m_1 = m_2 = \cdots = m_K = m$ then the complexity of evaluating the approximation (S15) of the K-dimensional integral (8) is $\mathcal{O}(mKR)$, which scales linearly with the separation rank R, the number of parameters K, and the number of grid nodes m along each dimension in the parameter space.

Computing transition probability and oscillation amplitude. In the parameter estimation (Figure 2) and sensitivity analysis (Figure S1), we illustrate the results based on the transition probability and oscillation amplitude that are extracted from tensor-structured parametric solution. For example, the probability that, in steady state distribution, the ℓ-th chemical species stays below a certain threshold $\tilde{x}_\ell$, is estimated as follows. We first integrate out all the other dimensions in the state space, and integrate the ℓ-th dimension up to $\tilde{x}_\ell$, i.e.,

$$p(x_\ell \leq \tilde{x}_\ell \,|\, \mathbf{k}) = \int_{a_1^x}^{b_1^x} \cdots \int_{a_{\ell-1}^x}^{b_{\ell-1}^x} \int_{a_\ell^x}^{\tilde{x}_\ell} \int_{a_{\ell+1}^x}^{b_{\ell+1}^x} \cdots \int_{a_N^x}^{b_N^x} p(\mathbf{x} \,|\, \mathbf{k}) \, \mathrm{d}\mathbf{x}.$$

In tensor structure, we use N-mode products to compute $p(x_\ell \leq \tilde{x}_\ell \,|\, \mathbf{k})$ simultaneously for all parameter combinations by

$$h_1^x h_2^x \cdots h_N^x \left(\bar{\mathbf{p}} \times_1 \mathbf{1} \times_2 \mathbf{1} \times_3 \cdots \times_{\ell-1} \mathbf{1} \times_\ell \mathbf{1}_{\tilde{x}_\ell} \times_{\ell+1} \mathbf{1} \times_{\ell+2} \cdots \times_N \mathbf{1} \right),$$

where $\mathbf{1}$ denotes a vector of all ones. Entries of $\mathbf{1}_{\tilde{x}_\ell}$ are equal to 1 if the corresponding grid point is smaller or equal to $\tilde{x}_\ell$, while its other entries are zero.

S1.5 Sensitivity analysis

The sensitivity indicator for an observable quantity Θ with respect to a parameter k is often computed as a finite difference [62]

$$S(\Theta) \approx \frac{\|\Theta(k + \Delta k) - \Theta(k)\|}{\Delta k} \frac{k}{\|\Theta(k)\|}, \tag{S16}$$

where $\|\cdot\|$ represent a suitable norm, and Δk is a change in the value of k. The model is sensitive to the parameter, if a small Δk yields a large value of $S(\Theta)$. For deterministic models, the observable Θ is usually the steady-state mean concentration. In stochastic setting, we have more options.

For example, let us consider the cell cycle model described in Figure 4(a). We will study the sensitivity with respect to the parameter k_1 for the following three observables Θ: mean concentration of the MPF (Θ_m), the oscillation amplitude (Θ_a) and the steady state distribution (Θ_p). In the case of the oscillation amplitude, we quantify Θ_a as the probability that the molecular population of the active MPF exceeds 400.

In the TPA framework, Θ_m and Θ_a are evaluated for all considered values of k_1 with computational cost scaling linearly with N. More importantly, the tensor-structured data enable direct comparison of two steady state probabilities in the 6-dimensional state space. Namely, the norm $\|\Theta_p(k_1 + \Delta k_1) - \Theta_p(k_1)\|$, needed in (S16), can be directly computed. The results are plotted in Figure S1. We observe that, within the considered range of k_1 (see Table S5), the sensitivity in the steady state distribution (blue curve) dominates in magnitude over $S(\Theta_m)$ and $S(\Theta_a)$. The steady state distribution contains a global information about the system and is more sensitive to parameter changes than derived quantities, like Θ_m and Θ_a.

S2 Description of models used in the illustrative TPA computations

S2.1 Schlögl model

The Schlögl system is defined by chemical reactions listed in Table S1. This table also shows the true values of parameters k_1, k_2, k_3 and k_4. Using this system, we illustrate the capabilities of the TPA for the parameter estimation and identifiablity. Table S2 provides the values of the first three statistical moments and the corresponding weights. The moments have been computed from a time series obtained by a long-time stochastic simulation with the values of parameters given in Table S1. A short segment of the time series is illustrated in Figure 1(a). We then assume that the values of parameters k_1, k_2, k_3 and k_4 are unknown. The TPA enables to evaluate the moment matching distance function given in the paper in equation (3), for all values of parameters within the parameter space given in Table S3. The resulting data are stored in the tensor format which enables efficient manipulations and post-processing.

Having the values of the distance function stored in the tensor format, we can easily and quickly obtain further pieces of information. For example, we can find those parameter values which do not fit the moments exactly, but with certain accuracy. More precisely, we consider tolerance $J_{\mathrm{TOL}} = 0.25\,\%$ and visualize in Figures 2(a)–(d) parameter values with distance function J less than J_{TOL}. Alternatively, if the values of moments are not available, we can utilize the TPA for different experimental data – see Figure 1(b) and equation (6) in the paper.

Note that all four values of parameters k_1, k_2, k_3 and k_4 cannot be estimated solely from the steady state distribution, because it does not inform us how fast the system reaches the steady state. In particular, if $\{k_i\}_{i=1,2,3,4}$ fit the pseudo-experimental data, then $\{Ck_i\}_{i=1,2,3,4}$ for any $C > 0$ fit these data as well. Therefore, in Figures 1(b) and 2, we fix one of the parameters at its true value and estimate values of the other three.

Table S1: *Overview of kinetic reactions of the Schlögl model.*

Index	Reaction	Kinetic rate[a]	True value
1	$3X \to 2X$	k_1/V^2	$k_1 = 2.5 \times 10^{-4}$
2	$2X \to 3X$	k_2/V	$k_2 = 0.18$
3	$\emptyset \to X$	$k_3 \times V$	$k_3 = 2250$
4	$X \to \emptyset$	k_4	$k_4 = 37.5$

[a]The reacting volume is set to $V = 1$ unit.

Table S2: *Moments estimated from stochastic simulation of the Schlögl model.*

Moment order	Value	Weight
1	$\hat{\mu}_1 = 261.32$	$\beta_1 = 1$
2	$\hat{\mu}_2 = 2.03 \times 10^4$	$\beta_2 = 100$
3	$\hat{\mu}_3 = -2.04 \times 10^5$	$\beta_3 = 0.001$

S2.2 Cell cycle model

The cell cycle model consists of nine chemical reactions and six chemical species as listed in Tables S4 and S5, see also Figure 4(a) in the paper. We have used this model to show how the TPA can be used to analyse bifurcations for high-dimensional problems, see Figures 4(b), 4(c) and 5 in the paper. In addition, we have used this system to discuss the sensitivity of various quantities in the stochastic model, see Figure S1 and Section S1.5.

Table S3: *Properties of molecular and rate variables in the Schlögl model.*

Type	Notation	Range	No. of nodes
Species	X	$[0, 1000]$	1024
Rate	k_1	$[2.43 \times 10^{-4}, 2.58 \times 10^{-4}]$	128
Rate	k_2	$[0.17, 0.19]$	128
Rate	k_3	$[2134, 2266]$	128
Rate	k_4	$[36.08, 38.63]$	128

Table S4: *Overview of kinetic reactions of the cell cycle model.*

Index	Reaction	Kinetic rate[a]	Parameter(s)
1	$M \rightarrow C_2 + YP$	k_1	bifurcation parameter
2	$\emptyset \rightarrow Y$	$k_2 \times V$	$k_2 = 0.015$
3	$CP + Y \rightarrow pM$	k_3/V	$k_3 = 200$
4	$pM \rightarrow M$	$k_4' + k_4 \left(M/V \right)^2$	$k_4 = 180, k_4' = 0.018$
5	$M \rightarrow pM$	$k_5 \times tP$	$k_5 = 0, tP = 0.001$
6	$Y \rightarrow \emptyset$	k_6	$k_6 = 0$
7	$YP \rightarrow \emptyset$	k_7	$k_7 = 0.6$
8	$C_2 \rightarrow CP$	$k_8 \times tP$	$k_8 = 1000$
9	$CP \rightarrow C_2$	k_9	$k_9 = 1000$

[a]The volume corresponds to a single cell and is set to $V = 5000$ units.

Table S5: *Properties of molecular and rate variables in the cell cycle model.*

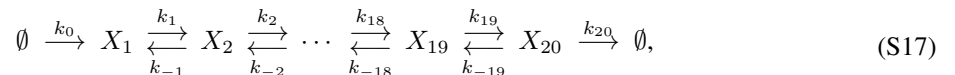| Type	Name	Notation	Range	No. of nodes
Species	cdc2	C2	$[2230, 4990]$	N/A[a]
Species	cdc2-P	CP	$[10, 70]$	256
Species	p-cyclin-cdc2-p	pM	$[0, 1500]$	256
Species	p-cyclin-cdc2	M	$[0, 1200]$	256
Species	cyclin	Y	$[20, 70]$	256
Species	p-cyclin	YP	$[0, 700]$	256
Rate	degradation rate of active MPF	k_1	$[0.25, 0.4]$	64

[a]Discretisation of cdc2 is not applicable here, since this variable is eliminated by the conservation law of cdc2 assumed by the original author.

S2.3   FitzHugh-Nagumo model

The FitzHugh-Nagumo model consists of five chemical reactions between two chemical species. It is illustrated in Figure 6(a) in the paper and the parameter ranges and mean values are provided in Tables S6 and S7. This model is used to show how the TPA can assess the influence of the extrinsic noise, see Figure 6 in the paper.

S2.4   A chemical reaction system in 20 dimensions

We consider a reaction chain of 20 molecular species:

$$\emptyset \xrightarrow{k_0} X_1 \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} X_2 \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} \cdots \underset{k_{-18}}{\overset{k_{18}}{\rightleftarrows}} X_{19} \underset{k_{-19}}{\overset{k_{19}}{\rightleftarrows}} X_{20} \xrightarrow{k_{20}} \emptyset, \tag{S17}$$
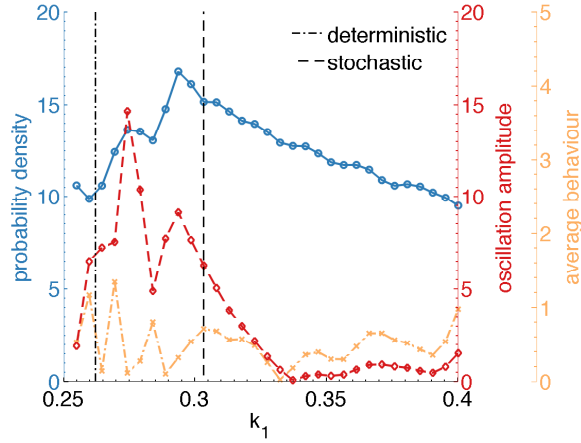
Figure S1: *Sensitivity indicators $S(\Theta)$ calculated by (S16) for 32 equidistant nodes within the range $[0.25, 0.4]$ of parameter k_1 (see Table S5 and Section S1.5). Three observables are considered: stationary distribution Θ_p (blue), oscillation amplitude Θ_a (red) and average number Θ_m of active MPF (orange). The dot-dashed and dashed lines indicate parameter values for which Figures 4(b) and 4(c) in the paper were computed, i.e. $k_1 = 0.2694$ (deterministic bifurcation point) and $k_1 = 0.3032$, respectively.*

Table S6: *Properties of molecular and rate variables in the FitzHugh-Nagumo model.*

Type	Notation	Range	No. of nodes
Species	X_1	$[0, 1800]$	256
Species	X_2	$[0, 700]$	256
Rate	k_1	$[0.17, 0.23]$	128
Rate	k_2	$[0.952, 0.1288]$	128
Rate	k_3	$[2.125, 2.875]$	128
Rate	k_4	$[0.0892, 0.1207]$	128

Table S7: *Overview of kinetic reactions of the FitzHugh-Nagumo model.*

Index	Reaction	Kinetic rate[a]	Mean value
1	$X_1 \rightarrow 2X_1$	$(X_1 - k_1 \times V)(V - X_1)$	$k_1 = 0.2$
2	$X_1 \rightarrow \emptyset$	X_2	N/A
3	$\emptyset \rightarrow X_1$	$k_2 \times V$	$k_2 = 0.112$
4	$X_2 \rightarrow \emptyset$	$k_3 \times k_4$	$k_3 = 2.5$
5	$X_1 \rightarrow X_1 + X_2$	k_4	$k_4 = 0.105$

[a]The system volume is $V = 2000$ units.

where $k_0 = 12$, $k_i = 0.2$ for $i = 1, \ldots, 20$, and $k_{-j} = 0.1$ for $j = 1, \ldots, 19$. A multilevel approach is implemented to solve the underlying CFPE, where the steady state distribution is first approximated on a coarse grid, and then interpolated to a finer grid (see Section S1.3). The results are plotted in Figure 7.

Table S8: *Multilevel discretisation for the* 20-*dimensional reaction chain* (S17).

Level	1	2	3	4	5	6	7
No. of nodes n	8	16	32	64	128	256	512
Grid size h	20	10	5	2.5	1.25	0.625	0.3125
CPU time ($\times 10^3$ sec)	2.82	7.02	2.84	2.3	2.25	0.08	0.10

References

[50] Gillespie DT (1992) A rigorous derivation of the chemical master equation. *Physica A* 188:404–425.

[51] Wu J, Vidakovic B, Voit EO (2011) Constructing stochastic models from deterministic process equations by propensity adjustment. *BMC Systems Biology* 5:187.

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

[53] Nobile F, Tempone R, Webster CG (2008) A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis* 46:2309–2345.

[54] Donoho DL, et al. (2000) High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture* pp 1–32.

[55] Lynch RE, Rice JR, Thomas DH (1964) Direct solution of partial difference equations by tensor product methods. *Numerische Mathematik* 6:185–199.

[56] Kruskal JB (1977) Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Its Applications* 18:95–138.

[57] Smith GD (1985) *Numerical solution of partial differential equations: finite difference methods* (Oxford University Press).

[58] Dolgov SV, Savostyanov DV (2013) Alternating minimal energy methods for linear systems in higher dimensions. Part I: SPD systems. (arXiv:1301.6068 [math.NA]).

[59] Oseledets IV, Tyrtyshnikov EE (2009) Breaking the curse of dimensionality, or how to use SVD in many dimensions. *SIAM Journal on Scientific Computing* 31:3744–3759.

[60] Oseledets IV (2010) Approximation of 2^d\times2^d matrices using tensor decomposition. *SIAM Journal on Matrix Analysis and Applications* 31:2130–2145.

[61] Kolda TG, Bader BW (2009) Tensor decompositions and applications. *SIAM review* 51:455–500.

[62] Savageau MA (1971) Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature* 229:542–544.