Spectral Representation and Reduced Order Modeling of Stochastic Reaction Networks

Khachik Sargsyan, Bert Debusschere, Habib Najm
Sandia National Laboratories, 7011 East Ave., MS 9051, Livermore, CA 94550, USA
{ksargsy,bjdebus,hnnajm}@sandia.gov

***Abstract*— For many biochemical phenomena in cells the molecule count is low, leading to stochastic behavior that causes deterministic macroscale reaction models to fail. The main mathematical framework representing these phenomena is based on continuous-time, discrete-state Markov processes that model the underlying *stochastic reaction network*. Conventional dynamical analysis tools do not readily generalize to the stochastic setting due to non-differentiability and absence of explicit state evolution equations. We developed a reduced order methodology for dynamical analysis that relies on the Karhunen-Loève decomposition and polynomial chaos expansions. The methodology relies on adaptive data partitioning to obtain an accurate representation of the stochastic process, especially in the case of multimodal behavior. As a result, a mixture model is obtained that represents the reduced order dynamics of the system. The Schlögl model is used as a prototype bistable process that exhibits time-scale separation and leads to multimodality in the reduced order model.**

I. Introduction

The simplest description of chemical reaction processes is based on *rate equations*, i.e. ordinary differential equations (ODEs) for species concentrations. This macroscopic setting fails when the relevant volume or the species numbers are small because of the increased significance of stochastic noise due to random molecular collisions [9], [22]. *Stochastic reaction networks* (SRNs) account for intrinsic stochastic noise, and provide a general framework for chemical reaction models at the microscopic, molecular level. SRNs are generally governed by the Chemical Master Equation [7] (CME), which is a differential equation governing the time evolution of the Probability Density Function (PDF) of species numbers. The chemical master equation is obtained by modeling a SRN as a jump Markov process [23], [6], i.e. discrete-state, continuous-time stochastic processes with no memory. Since computing direct numerical solutions for CMEs is still challenging (for recent efforts, see [15] and references therein), simulation-based methods become the main analytical tools. In particular, Gillespie's Stochastic Simulation Algorithm [4], [5] (SSA) provides a simulation mechanism for the time-evolution of species numbers at the microscopic scale, thereby effectively sampling the CME solution. This allows determining useful statistical properties of the system by averaging *without* solving the CME itself.

In this work, we rely on Karhunen-Loève (KL) expansions [10], [14], [2] that represent the underlying stochastic processes in terms of orthonormal random variables, truncated to a reduced order model. This low-order representation is constructed based on the observed statistics of the stochastic process over a given period of time. With a truncated KL expansion, each realization of a stochastic process corresponds to a finite number of uncorrelated random variables, with non-standard distributions determined by the data. As a result, it is desirable to represent these random variables with polynomial chaos (PC) expansions [25] that enable computationally efficient estimation of system properties.

However, a global PC representation with a finite order and dimensionality does not accurately capture random variables that exhibit strong multimodalities [18]. Adaptive multi-wavelet [11], [12], [13] or PC [24] bases, both relying on stochastic domain decomposition, enable efficient analysis of such processes in the continuous deterministic setting. In this work, we extend the methodology proposed in [18] to obtain an adaptive, data-driven partitioning that captures the structure and modalities of intrinsic stochasticity. Our data partitioning algorithm, which involves a combination of clustering and data range bisection, leads to a mixture of PC expansions that properly represents multimodal distributions by taking advantage of the underlying data structure.

II. Reduced Order Modeling via Karhunen-Loève Decomposition

As a reduced order model for a stochastic process $X(t,\theta)$, consider the L-truncated *Karhunen-Loève (KL) expansion* [10], [14]

$$X(t,\theta) \approx X_{KL}(t,\theta) = \bar{X}(t,\theta) + \sum_{i=1}^{L} \sqrt{\lambda_i} f_i(t) \xi_i, \quad (1)$$

where $\bar{X}(t,\theta)$ denotes the expectation with respect to the sample space element θ. In the above KL expansion, the λ_i are the eigenvalues of the covariance kernel with corresponding orthogonal eigenfunctions $f_i(t)$. The random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_L)$ consists of L jointly distributed and uncorrelated (but not independent) random variables. Essentially, the dynamics of the full process $X(t)$ is captured by a single random vector $\boldsymbol{\xi}$ (we will drop the argument θ for clarity, unless there is a need to put an emphasis on the intrinsic randomness).

As a benchmark process that exhibits a bimodal behavior, consider the Schlögl model [19], [6], [18], which is a SRN involving two reversible reactions and three species

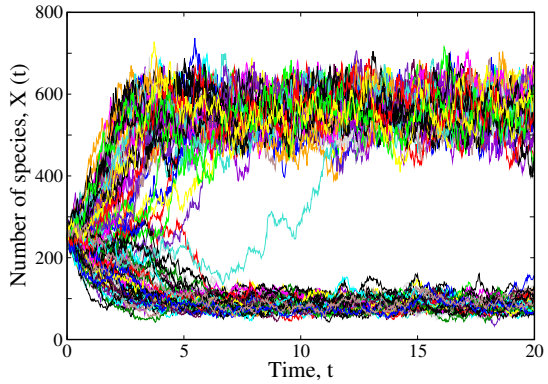

Fig. 1. Hundred SSA realizations of the Schlögl model with the nominal parameter set [17].

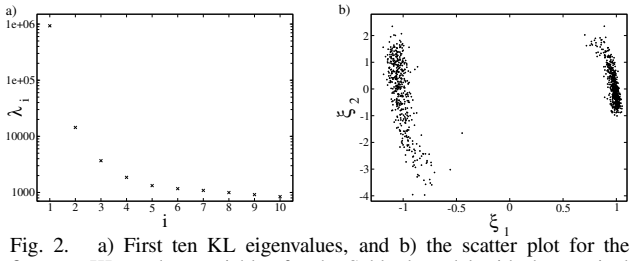

Fig. 2. a) First ten KL eigenvalues, and b) the scatter plot for the first two KL random variables for the Schlögl model with the nominal parameter set [17].

X, A and B, but with A and B assumed to be present in large and fixed numbers. We are interested in the number of species $X(t)$. With $X(0) = 250$ and the nominal set of rate constants, the system exhibits bistable behavior over a time window $t \in [0, 20]$, see Fig. 1. Furthermore, Fig. 2 illustrates the corresponding eigenvalue spectrum and the scatter plot of the projected samples of ξ_1 and ξ_2. The huge gap between the first two eigenvalues and bimodality along the ξ_1-dimension are direct results of the bimodality of the time-dependent process itself. Although the random vector $\boldsymbol{\xi}$ has uncorrelated components, it may have a complicated structure that is not known beforehand. We then turn to spectral expansions in order to properly represent the random vector $\boldsymbol{\xi}$.

III. Polynomial Chaos Expansion of the Reduced Order Model

We seek to approximate $\boldsymbol{\xi}$ with a random variable represented by a d-th order, L-dimensional PC expansion

$$\boldsymbol{\xi} = \sum_{k=0}^{P} \boldsymbol{c}_k \Psi_k(\zeta_1, \ldots, \zeta_L) \equiv \boldsymbol{g}\left(\boldsymbol{\zeta}; \boldsymbol{C}\right), \qquad (2)$$

with the number of terms $P+1 = \frac{(d+L)!}{d!L!}$ and multivariate orthogonal polynomials $\Psi_k(\boldsymbol{\zeta})$. The components of the random vector $\boldsymbol{\zeta}$ are standard i.i.d. random variables. In this work, we have used Hermite polynomials that are orthogonal with respect to the PDF of a standard normal random variable. Namely,

$$\begin{aligned} \langle \Psi_j(\boldsymbol{\zeta})\Psi_k(\boldsymbol{\zeta})\rangle &\equiv \int \Psi_j(\boldsymbol{\zeta})\Psi_k(\boldsymbol{\zeta}) \frac{e^{-\frac{\boldsymbol{\zeta}^T\boldsymbol{\zeta}}{2}}}{\sqrt{2\pi}} d\boldsymbol{\zeta} \\ &= \langle \Psi_k^2(\boldsymbol{\zeta})\rangle \delta_{jk}. \end{aligned} \qquad (3)$$

The above orthogonality relation leads to the projection formulas

$$\boldsymbol{c}_k = \frac{\langle \boldsymbol{\xi}\Psi_k(\boldsymbol{\zeta})\rangle}{\langle \Psi_k^2(\boldsymbol{\zeta})\rangle}. \qquad (4)$$

However, in order to compute the stochastic projection integral $\langle \boldsymbol{\xi}\Psi_k(\boldsymbol{\zeta})\rangle$, one needs an one-to-one correspondence between samples of $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$. To resolve this, we employ the *Rosenblatt transformation* [16], [17] that enables the projection (4) in the same, $\boldsymbol{\zeta}$-space. Finally, the full representation can be written as

$$\begin{aligned} X(t,\theta) &\approx X_{KLPC}(t,\theta) = \\ &= \bar{X}(t,\theta) + \sum_{i=1}^{L} \left(\sum_{k=0}^{P} c_{ik}\Psi_k(\boldsymbol{\zeta}) \right) \sqrt{\lambda_i} f_i(t), \end{aligned} \qquad (5)$$

i.e. the process $X(t,\theta)$ is described in terms of *deterministic* matrix elements c_{ik} and a random vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_L)$ of standard normal i.i.d. random variables. However, as shown in [18], the global PC representation is challenged if the random vector $\boldsymbol{\xi}$ has a multimodal character, which is certainly the case for the Schlögl model, see Fig. 2b.

IV. Adaptive Data Partitioning Algorithm

In order to tackle multimodalities, we analyzed various approaches of partitioning the data set of samples of $\boldsymbol{\xi}$, and introduced a novel, hybrid and adaptive strategy that involves approximate k-center clustering [8] to detect the bimodalities, followed by data-range bisection. The algorithm adaptively partitions the data set into subsets that are simpler to represent with low-order PC, until this representation is satisfactory in terms of the *Kullback-Leibler (K-L) divergence* or *relative entropy* [3], [1] between the PDFs of the data samples and the samples of the corresponding representation (i.e., $P(\cdot)$ and $Q(\cdot)$, respectively)

$$d(P||Q) = \int P(\boldsymbol{x}) \log \frac{P(\boldsymbol{x})}{Q(\boldsymbol{x})} d\boldsymbol{x}. \qquad (6)$$

Exact computation of the K-L divergence requires an integration that is extremely costly in multiple dimensions. Nevertheless, it can be estimated by Monte-Carlo integration in terms of the data samples that are available. Namely,

$$\begin{aligned} d(P||Q) &\approx \frac{1}{N}\sum_{n=1}^{N} \log \frac{P(\boldsymbol{\xi}^{(n)})}{Q(\boldsymbol{\xi}^{(n)})} \\ &= \frac{1}{N}\left(\sum_{n=1}^{N} \log P(\boldsymbol{\xi}^{(n)}) - \sum_{n=1}^{N} \log Q(\boldsymbol{\xi}^{(n)}) \right), \end{aligned} \qquad (7)$$

where $\boldsymbol{\xi}^{(n)}$ for $n = 1, 2, \ldots, N$ are the samples drawn from the distribution $P(\cdot)$, i.e. exactly the data samples

that are to be PC-represented. This approximation of the K-L divergence allows simple intuitive interpretation: the second sum is the log-probability of having the particular data set $\{\boldsymbol{\xi}^{(n)}\}_{n=1}^{N}$ *given* a model that leads to the PDF $Q(\cdot)$ (in other terms, the *likelihood* of the model), while the first sum is the likelihood *if* the model had the exact same PDF as the original data set (in a sense, a *target* likelihood). The PDFs in (7) are computed by standard KDE techniques [21], [20].

We have analyzed various data partitioning schemes and found that the domain-based bisection approaches (specifically, data range bisection, data median bisection and data size bisection, see [17]) blindly split the data without detecting the modalities. Therefore, we enhanced the methodology with an initial clustering step (namely, an approximate version of the k-center clustering is implemented) that detects the modalities present in the data structure. After this initial step, it is shown that the data range bisection works better than the other approaches. It consists of finding and bisecting the data range in each direction simultaneously.

In order to find out whether an initial clustering is needed and what the optimal number of clusters is, we employ the *explained variance criterion*. The explained variance for a specific clustering is a variance of a data set that is obtained from the initial data set by replacing each sample with the mean of its cluster. The fraction of the explained variance over the total variance vanishes if there is only one cluster (the set itself) and is equal to one, if the number of clusters is the same as the number of data samples. We run several trial clustering cases for various fixed number of clusters and check the graph of the explained variance fraction versus the cluster number. This graph is generally increasing and concave down. If there is a well-seen 'elbow' in the graph, then its location corresponds to the optimal number of clusters. Otherwise, there is no need to proceed with the clustering, and the data is considered sufficiently unimodal [17].

The adaptive PC representation algorithm then proceeds as follows:

0. Obtain N SSA realizations $X(t)$.
1. Perform KL decomposition up to the eigenmode (dimension) L.
   1a. As a result, obtain a set of N data samples of the random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_L)$ and call it the current data set $S = \{\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \ldots, \boldsymbol{\xi}^{(N)}\}$.
   1b. *If* the explained variance criterion [17] detects modalities, cluster the data into the optimal number of clusters and proceed considering each cluster as a new data set. *Otherwise* proceed to Step 2.
2. Use the Rosenblatt transformation and quadrature evaluation of the projection integrals (4) to find a finite order PC representation for the current data samples: $\xi_i = \sum_{k=0}^{P} c_{ik}\Psi_k(\boldsymbol{\zeta})$, for $i = 1, 2, \ldots, L$.
   2a. Compute the K-L divergence between the data and the PC model using (7).
3. *If* the number of samples in the current data set exceeds the threshold N_{thr} *and* the K-L divergence is larger than the threshold d_{thr}, partition the current data set according to data range bisection, and recursively return to Step 2 for each new data set. *Else* keep the current PC representation and move to the next untreated data set.

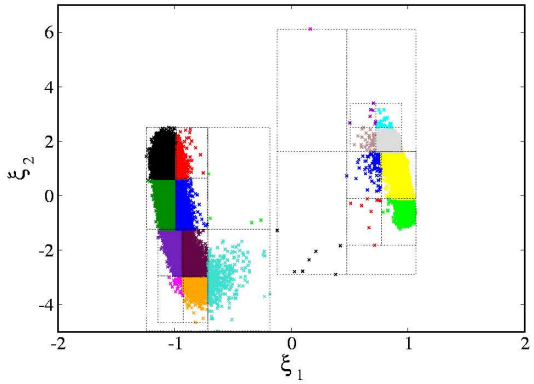

Fig. 3. The data partitions for the first two KL variables obtained from a KL projection of $N = 10^5$ realizations of the Schlögl process.

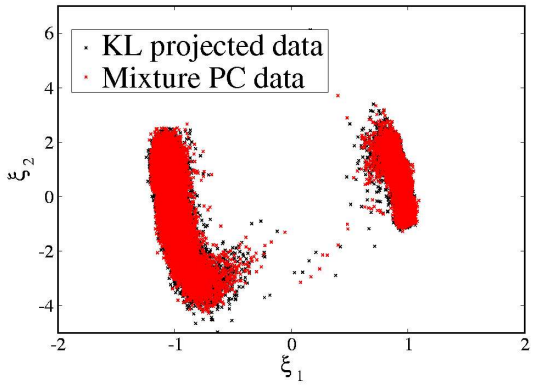

Fig. 4. The scatter plot of the original data set and the samples obtained from the mixture PC representation.

The final representation then corresponds to a PDF that is a *mixture* of PDFs of PC representations of each of the K subsets. Namely,

$$PDF_{\boldsymbol{\xi}_{PC}}(\boldsymbol{y}) = \sum_{j=1}^{K} p_j PDF_{\boldsymbol{g}\left(\boldsymbol{\zeta};\boldsymbol{C}^{(j)}\right)}(\boldsymbol{y}), \qquad (8)$$

where p_j is the fraction of data samples in the j-th partition.

Fig. 3 shows the final partitions for a two-dimensional data set for a random vector $\boldsymbol{\xi} = (\xi_1, \xi_2)$ that is obtained by a KL projection of $N = 10^5$ realizations of the Schlögl process. The data set itself and the samples of its third order mixture PC model representation are shown in Fig. 4. The K-L convergence analysis of our hybrid methodology and other data partitioning strategies is illustrated in Fig. 5. Although for this particular data set - it is bimodal along the first dimension only - the plain

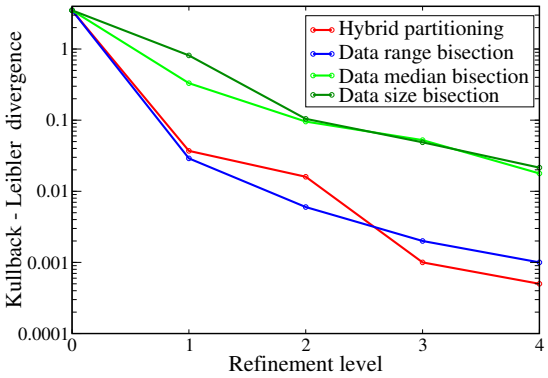

Fig. 5. Convergence of the mixture PC representation as the partitioning refinement level increases. Various data partitioning strategies are compared [17]. The zeroth refinement levels correspond to the global representation, while the first level is simply the clustering for the hybrid partitioning. The third refinement levels correspond to the illustration from Fig. 3.

data range bisection is as efficient as the hybrid approach, it is shown [17] that the hybrid methodology is more robust for *general* data sets with no *a priori* knowledge of the data structure available.

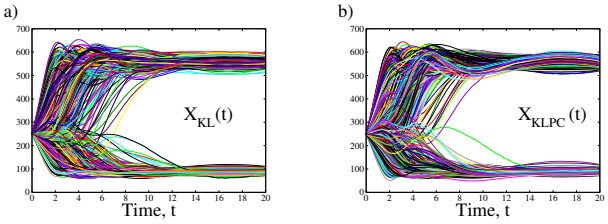

Fig. 6. a) The 5-mode KL truncated sum for the Schlögl process. b) The final representation obtained from mixture PC expansions of the underlying five-dimensional KL random vector. Both expansions are obtained with $N = 10^5$ realizations with only every hundredth realization shown for illustration purposes.

Finally, Fig. 6 illustrates the 5-mode KL-truncated sum

$$X_{KL}(t) = \bar{X}(t) + \sum_{i=1}^{5} \xi_i \sqrt{\lambda_i} f_i(t) \qquad (9)$$

of the underlying Schlögl process as well as the process, recovered from the third order mixture PC representation of the KL-projected variables, i.e.

$$X_{KLPC}(t) = \bar{X}(t) + \sum_{i=1}^{5} (\xi_{PC})_i \sqrt{\lambda_i} f_i(t). \qquad (10)$$

Clearly, the stochastic process $X(t)$, first reduced to $X_{KL}(t)$ (described by a random vector $\boldsymbol{\xi}$) by the KL projection, is further reduced to $X_{KLPC}(t)$ (described by a set of deterministic matrices $\{\boldsymbol{C}^{(j)}\}_{j=1}^{K}$, one for each partition of the data samples of $\boldsymbol{\xi}$) by our mixture PC representation while preserving the skeleton of the dynamics of the original process for further analysis of the system, or as a reduced order model.

ACKNOWLEDGMENT

This work was supported by the U.S. Department of Energy Office of Science through the Applied Mathematics program in the Office of Advanced Scientific Computing Research (ASCR) under contract 07-012783 with Sandia National Laboratories. Sandia National Laboratories is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy under contract No. DE-AC04-94-AL85000.

REFERENCES

[1] M. Arnst and R. Ghanem. Probabilistic equivalence and stochastic model reduction in multiscale analysis. *Comput. Methods Appl. Mech. and Engrg.*, 197:3584–3592, 2002.
[2] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer Verlag, New York, 1991.
[3] A. Gibbs and F. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70:419–436, 2002.
[4] D. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.
[5] D. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
[6] D. Gillespie. *Markov Processes: An Introduction for Physical Scientists*. Academic Press, San Diego, CA, 1992.
[7] D. Gillespie. A rigorous derivation of the chemical master equation. *Phys. A*, 188:404–425, 1992.
[8] T. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theor. Comp. Science*, 38:293–306, 1985.
[9] W. Horsthemke and R. Lefever. *Noise-Induced Transitions*. Springer-Verlag, Berlin, 1984.
[10] K. Karhunen. Zur spektraltheorie stochastischer prozesse. *Ann. Acad. Sci. Fennicae*, 37, 1946.
[11] O. Le Maître, R. Ghanem, O. Knio, and H. Najm. Uncertainty propagation using Wiener-Haar expansions. *J. Comput. Phys.*, 197(1):28–57, 2004.
[12] O. Le Maître, H. Najm, R. Ghanem, and O. Knio. Multi-resolution analysis of Wiener-type uncertainty propagation schemes. *J. Comput. Phys.*, 197:502–531, 2004.
[13] O. Le Maître, H. Najm, P. Pébay, R. Ghanem, and O. Knio. Multi-resolution-analysis scheme for uncertainty quantification in chemical systems. *SIAM J. Sci. Comput.*, 29(2):864–889, 2007.
[14] M. Loève. *Probability Theory*. Van Nostrand, Princeton, NJ, 1955.
[15] S. MacNamara, A. Bersani, K. Burrage, and R. 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(9):095105–1–13, 2008.
[16] M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3):470 – 472, 1952.
[17] K. Sargsyan, B. Debusschere, H. Najm, and O. L. Maître. Spectral representation and reduced order modeling of the dynamics of stochastic reaction networks via adaptive data partitioning. *SIAM Journal on Scientific Computing*, submitted.
[18] K. Sargsyan, B. Debusschere, H. Najm, and Y. Marzouk. Bayesian inference of spectral expansions for predictability assessment in stochastic reaction networks. *Journal of Computational and Theoretical Nanoscience*, in press.
[19] F. Schlögl. On thermodynamics near a steady state. *Z. Phys.*, 248:446–458, 1971.
[20] D. Scott. *Multivariate Density Estimation. Theory, Practice and Visualization*. Wiley, New York, 1992.
[21] B. Silverman. *Density Estimation*. Chapman and Hall, London, 1986.
[22] R. Srivastava, L. You, J. Summers, and J. Yin. Stochastic vs. deterministic modeling of intracellular viral kinetics. *Journal of Theoretical Biology*, 218(3):309 – 321, 7 2002.
[23] N. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier Science, Amsterdam, 1992.
[24] X. Wan and G. E. Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *J. Comput. Phys.*, 209:617–642, 2005.
[25] N. Wiener. The homogeneous chaos. *Am. J. Math.*, 60:897–936, 1938.