Physics-aware, data-driven discovery of slow and stable coarse-grained dynamics for high-dimensional multiscale systems

Sebastian Kaltenbach, Phaedon-Stelios Koutsourelakis
Professorship for Continuum Mechanics
Technical University of Munich
Munich, Germany
{sebastian.kaltenbach,p.s.koutsourelakis}@tum.de

Abstract

We propose a data-driven, hierarchical state-space model that provides a lower-dimensional, coarse-grained dynamical description of a high-dimensional, fine-grained, multiscale dynamical system. Given the expense associated with the generation of training data by physical simulators (e.g. molecular dynamics), we infuse domain knowledge with the introduction of an intermediate layer of physics-motivated, latent variables. We combine these with a flexible prior on the complex plane that facilitates the discovery of the latent, slow processes and ensures the long-term stability of the learned dynamics. In contrast to existing schemes, the proposed model does not require the a priori definition of projection operators from the fine-grained description and addresses simultaneously the tasks of dimensionality reduction and model estimation. We demonstrate its efficacy and accuracy in multiscale physical systems of particle dynamics where probabilistic, long-term predictions of phenomena not contained in the training data are produced.

1 Introduction

The present paper is concerned with the discovery of dynamic, coarse-grained models from fine-scale simulation data that provide insight into the latent, slow processes and enable multiscale modeling. Many problems in science and engineering are modeled by high-dimensional systems of deterministic or stochastic, (non)linear, microscopic evolution laws as in the atomistic simulation of materials [1], in complex flows [2] or in agent-based models [3]. Their solution is generally dominated by the smallest time scale involved even though the outputs of interest might pertain to time scales that are greater by several orders of magnitude [4]. The disparity of time-scales in combination with the high-dimensionality has motivated the development of coarse-grained descriptions which are lower-dimensional but are nevertheless predictive of the fine-grained system's evolution.

While generic state-space models for analyzing time-series data such as [5, 6, 7] can be employed in such problems, they struggle in the small data regime, the solutions obtained are not necessarily interpretable and can fail in out-of-sample predictions e.g. when new intial conditions are employed. In this paper we propose a novel physics-aware, probabilistic, coarse-graining framework with guaranteed stability that combines recent advances in statistical learning with a hierarchical architecture that promotes the discovery of interpretable, low-dimensional representations. We employ a generative state-space model with two layers of latent variables. The first describes the latent dynamics using a novel prior on the complex plane that guarantees stability and yields a clear distinction between fast and slow processes, the latter being responsible for the system's long-term evolution. The second layer involves physically-motivated latent variables which infuse inductive bias, enable connections

1st NeurIPS workshop on Interpretable Inductive Biases and Physically Structured Learning (2020), virtual.

with the very high-dimensional observables and reduce the data requirements for training. The probabilistic formulation adopted enables the quantification of a crucial, and often neglected, component in any model-order reduction process, i.e. the predictive uncertainty due to information loss.

This work shares similarities with approaches that combine physical knowledge with statistical learning. This includes methods based on dynamical invariants such as Hamiltonian [8, 9] and Lagrangian Dynamics [10] as well as approaches that use physical laws as regularization terms or for augmenting the loss function as in [11, 12, 13, 14]. In the context of molecular dynamics multiple schemes for coarse-graining which also guarantee long-term stability have been proposed by [15, 16, 17]. Another line of work close to ours pertains to the use of Koopman-operator theory [18] which attempts to identify appropriate transformations of the original coordinates that yield linear dynamics [19]. We note that these approaches [12, 20, 21, 22] require additionally the specification of an encoder i.e. a map from the original description to the reduced coordinates which we avoid in the generative formulation adopted.

2 Physics-aware, Probabilistic Coarse-Graining

Given N high-dimensional observables $\{\boldsymbol{x}_{0:T}^{(i)}\}_{i=1}^{N}$ over T time-steps with $\boldsymbol{x} \in \mathcal{M} \subset \mathbb{R}^f$ ($f >> 1$) generated by a computational physics simulator, we seek a lower-dimensional set of of collective variables or latent generators $\boldsymbol{z}_t$ and their associated dynamics. Because of the difficulties associated with these tasks, we advocate the introduction of an intermediate layer of physically-motivated, lower-dimensional variables $\boldsymbol{X}_t$ (e.g. density or velocity fields) as well as a slowness-promoting prior for $\boldsymbol{z}_t$. Using $\boldsymbol{X}_t$ alone (without $\boldsymbol{z}_t$) would make it extremely difficult to enforce long-term stability while achieving sufficient complexity in the learned dynamics [23, 20]. Furthermore and even if the dynamics of $\boldsymbol{x}_t$ are Markovian, this is not necessarily the case for the coarse-grained $\boldsymbol{X}_t$ [24]. The latent variables $\boldsymbol{z}_t$ therefore effectively correspond to a nonlinear coordinate transformation that yields not only Markovian but also stable dynamics [21]. The general framework is summarized in Figure 1.

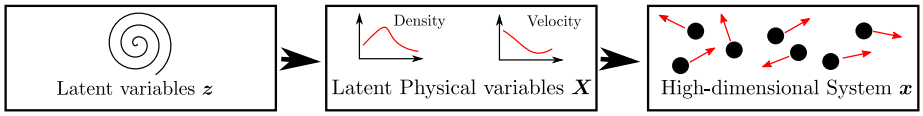

Figure 1: Visual summary of proposed framework. The low-dimensional variables $\boldsymbol{z}$ act as generators of an intermediate layer of latent, physically-motivated variables $\boldsymbol{X}$ that are able to reconstruct the high-dimensional system $\boldsymbol{x}$.

2.1 Model Structure

Our model consists of three levels. At the first level reside the latent variables $\boldsymbol{z}_t$ which are connected with $\boldsymbol{X}_t$ in the second layer through a probabilistic map $\boldsymbol{G}$. The physical variables $\boldsymbol{X}_t$ are finally connected to the high-dimensional observables through another probabilistic map $\boldsymbol{F}$. We parametrize $\boldsymbol{F}, \boldsymbol{G}$ with deep neural networks and denote by $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ the corresponding parameters. In particular, we postulate the following relations:

$$z_{t,j} = z_{t-1,j} \exp(\lambda_j) + \sigma_j \epsilon_{t,j} \quad \lambda_j \in \mathbb{C}, \quad \epsilon_{t,j} \sim \mathcal{CN}(0,1),\ j = 1, 2, \ldots, h \tag{1}$$

$$\boldsymbol{X}_t = \boldsymbol{G}(\boldsymbol{z}_t, \boldsymbol{\theta}_1) \tag{2}$$

$$\boldsymbol{x}_t = \boldsymbol{F}(\boldsymbol{X}_t, \boldsymbol{\theta}_2) \tag{3}$$

We assume that the latent variables $\boldsymbol{z}_t$ are complex-valued and a priori independent. We model their dynamics with a discretized Ornstein-Uhlenbeck process on the complex plane with initial conditions $z_{0,j} \sim \mathcal{CN}(0, \sigma_{0,j}^2)$. The parameters associated with this level are denoted summarily by $\boldsymbol{\theta}_0 = \{\sigma_{0,j}^2, \sigma_j^2, \lambda_j\}_{j=1}^{h}$.

2.1.1 Stable low-dimensional Dynamics

While the physical systems (e.g. molecular dynamics) of interest are highly non-stationary, they generally converge to equilibrium in the long-term. We enforce long-term stability here by ensuring

that the real-part of the λ_j's in Equation (1) is negative, i.e.:

$$\lambda_j = \Re(\lambda_j) + i\ \Im(\lambda_j) \text{ with } \Re(\lambda_j) < 0 \tag{4}$$

which guarantees first and second-order stability i.e. the mean as well as the variance are bounded at all time steps.

The transition density each process $\boldsymbol{z}_{t,j}$ is given by:

$$p\left(z_{t,j} \mid z_{t-1,j}\right) = \mathcal{N}\left(\begin{bmatrix}\Re(z_{t,j})\\ \Im(z_{t,j})\end{bmatrix} \mid \boldsymbol{\mu} = s_j\ \boldsymbol{R}_j \begin{bmatrix}\Re(z_{t-1,j})\\ \Im(z_{t-1,j})\end{bmatrix}, \boldsymbol{\Sigma} = \boldsymbol{I}\frac{\sigma_j^2}{2}\right) \tag{5}$$

with the orthogonal matrix $\boldsymbol{R}_j$ depending on the imaginary part of λ_j:

$$\boldsymbol{R}_j = \begin{bmatrix}\cos(\Im(\lambda_j)) & -\sin(\Im(\lambda_j))\\ \sin(\Im(\lambda_j)) & \cos(\Im(\lambda_j))\end{bmatrix} \tag{6}$$

The decay rate $s_j = \exp(\Re(\lambda_j))$ depends on the real part of λ_j, i.e. the closer to zero the latter is, the "slower" the evolution of the corresponding process is. As in probabilistic Slow Feature Analysis (SFA) [25, 26], we set $\sigma_j^2 = 1 - \exp(2\ \Re(\lambda_j)) = 1 - s_j^2$ and $\sigma_{0,j}^2 = 1$. This implies that a priori the latent dynamics are stationary and enables ranking of the slowness of each process j on the basis of $\Re(\lambda_j)$. Hence the only independent parameters are the λ_j. The imaginary part of λ_j can account for periodic effects in the latent dynamics.

2.1.2 Probabilistic Generative Mapping

We employ fully probabilistic maps between the different layers which involve two conditional densities based on Equations (2) and (3), i.e.:

$$p(\boldsymbol{x}_t \mid \boldsymbol{X}_t, \boldsymbol{\theta}_2) \quad \text{and} \quad p(\boldsymbol{X}_t \mid \boldsymbol{z}_t, \boldsymbol{\theta}_1) \tag{7}$$

In contrast to the majority of physics-motivated papers [27, 28, 24], we note that the generative structure adopted does not require the prescription of a restriction operator and the reduced variables need not be selected a priori but rather are adapted to best reconstruct the observables.

The splitting of the generative mapping into two parts through the introduction of the intermediate variables $\boldsymbol{X}_t$ has two main advantages. Firstly, $\boldsymbol{X}$ induce physical knowledge which reduces the complexity of the maps involved and the number of parameters that need to be learned. Secondly and once $\boldsymbol{X}$ have been inferred, additional insight is gained and the results are more easily interpretable.

2.2 Inference and Learning

Given the probabilistic relations above, our goal is to identify the state variables $\boldsymbol{X}_{0:T}^{(1:n)}, \boldsymbol{z}_{0:T}^{(1:n)}$ as well as all parameters $\boldsymbol{\theta}$ of the model. We follow a hybrid Bayesian approach in which the posterior of the state variables is approximated using structured Stochastic Variational Inference [29] and the MAP point estimates for $\boldsymbol{\theta} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$ are obtained.

We approximate the true posterior with $q_{\boldsymbol{\phi}}(\boldsymbol{X}_{0:T}^{(1:n)}, \boldsymbol{z}_{0:T}^{(1:n)}) = \prod_{i=1}^{N} q_{\boldsymbol{\phi}}(\boldsymbol{z}_{0:T}^{(i)} | \boldsymbol{X}_{0:T}^{(i)}) q_{\boldsymbol{\phi}}(\boldsymbol{X}_{0:T}^{(i)})$ where $\boldsymbol{\phi}$ are the parameters involved. Both of the associated densities are modeled with Gaussians with tridiagonal precision matrices [30, 31] in order to achieve scalability and the dependence between the variables is captured using deep neural networks. The optimal parameter values are found by maximizing the Evidence Lower Bound (ELBO) $\mathcal{F}(q_{\boldsymbol{\phi}}(\boldsymbol{X}_{0:T}^{(1:n)}, \boldsymbol{z}_{0:T}^{(1:n)}), \boldsymbol{\theta})$. We compute Monte Carlo estimates of the gradient of the ELBO with respect to ϕ and $\boldsymbol{\theta}$ with the help of the reparametrization trick [32] and perform stochastic optimization with the help of the ADAM algorithm [33].

3 Particle Dynamics: Viscous Burgers' Equation

The high-dimensional, fine-grained model considered consists of f identical particles which can move in a bounded one-dimensional domain $s \in [-1,\ 1]$ (under periodic boundary conditions). The variables $\boldsymbol{x}_t$ consist therefore of the coordinates of the particles at each time instant t and the dimension of the system f is equal to the number of particles. We consider stochastic particle

dynamics that correspond to a viscous-Burgers'-type behaviour [34, 35, 36]. The mapping $\boldsymbol{F}$ between $\boldsymbol{X}_t$ and $\boldsymbol{x}_t$ is fixed to a multinomial distribution whereas $\boldsymbol{G}$ is represented with a neural network.

We train the model on $N = 64$ time-series of the positions of $f = 500 \times 10^3$ particles over $T = 20$ time-steps. The physical latent variables are associated with the particle density which is represented by discretizing the domain with 64 bins of equal size. Given the conservation of mass constraint, we make use of real-valued $\boldsymbol{X}_t$ which are then transformed with the softmax function in order to yield the particle density. Furthermore, we made use of $h = 5$ complex, latent processes $\boldsymbol{z}_t$.

In Figure 2 the learned values for the complex variables λ_j are plotted as well as the inferred and predicted time-evolution of 2 associated processes $z_{t,j}$ on the complex plane. We note the clear separation of time-scales in the first plot with three slow processes and two fast ones. This is also evident in the indicative trajectories on the complex plane.

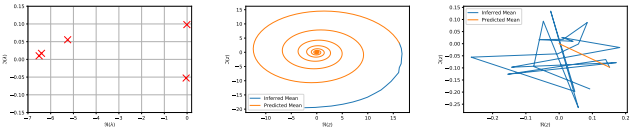

Figure 2: Complex eigenvalues λ (left) and the time evolution of indicative slow (middle) and fast (right) processes $z_{t,j}$ associated with them.

In Figure 3 we compare the true particle density with the one predicted by the trained reduced model. We note that the latter is computed by reconstructing the $\boldsymbol{x}_t$ futures. We observe that the model is able to accurately track first-order statistics well into the future.

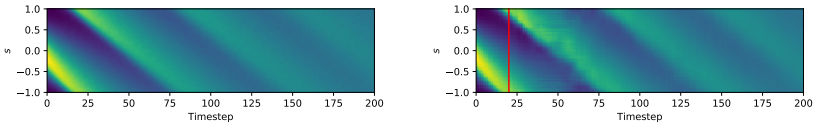

Figure 3: Particle density: Inferred and predicted posterior mean (right) in comparison with the ground truth (left). The red line divides inferred quantities from predicted ones.

A more detailed view of the predictive estimates with snapshots of the particle density at selected time instances is presented in Figure 4. Here, not only the posterior mean but also the associated uncertainty is displayed. The propagation and dissipation of the Burgers' type shock front is clearly captured. We want to emphasize the last Figure at $t = 1000$ when the steady state has been reached which clearly shows that our model is capable of converging to stable equilibrium.

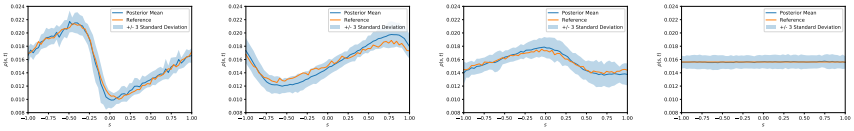

Figure 4: Predicted particle density profiles at $t = 40, 80, 120, 1000$ (from left to right).

Finally, in Figure 5 the ability of the proposed model to accurately predict second-order statistics is demonstrated.

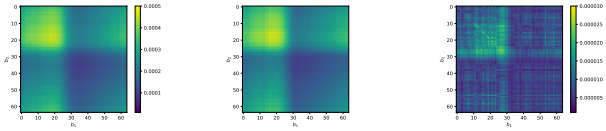

Figure 5: Two-point probability at $t = 40$, i.e. the probability of finding two particles simultaneously in two bins (b_1, b_2): Reference (left), posterior mean (middle), posterior standard deviation (right).

References

[1] Xiantao Li. "A coarse-grained molecular dynamics model for crystalline solids". en. In: *International Journal for Numerical Methods in Engineering* 83.8-9 (Aug. 2010), pp. 986–997.

[2] Andrew J. Majda and John Harlim. "Physics constrained nonlinear regression models for time series". en. In: *Nonlinearity* 26.1 (2013), p. 201.

[3] Spencer A. Thomas, David J. B. Lloyd, and Anne C. Skeldon. "Equation-free analysis of agent-based models and systematic parameter determination". In: *Physica A: Statistical Mechanics and its Applications* 464 (Dec. 2016), pp. 27–53.

[4] D. Givon, R. Kupferman, and A. Stuart. "Extracting Macroscopic Dynamics: Model Problems and Algorithms". In: *Nonlinearity* (2004).

[5] Maximilian Karl et al. "Deep variational bayes filters: Unsupervised learning of state space models from raw data". In: *arXiv preprint arXiv:1605.06432* (2016).

[6] Syama Sundar Rangapuram et al. "Deep state space models for time series forecasting". In: *Advances in neural information processing systems*. 2018, pp. 7785–7794.

[7] Longyuan Li et al. "Learning Interpretable Deep State Space Model for Probabilistic Time Series Forecasting." In: *IJCAI*. 2019, pp. 2901–2908.

[8] Samuel Greydanus, Misko Dzamba, and Jason Yosinski. "Hamiltonian neural networks". In: *Advances in Neural Information Processing Systems*. 2019, pp. 15379–15389.

[9] Peter Toth et al. "Hamiltonian generative networks". In: *arXiv preprint arXiv:1909.13789* (2019).

[10] Michael Lutter, Christian Ritter, and Jan Peters. "Deep lagrangian networks: Using physics as model prior for deep learning". In: *arXiv preprint arXiv:1907.04490* (2019).

[11] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In: *Journal of Computational Physics* 378 (2019), pp. 686–707.

[12] Bethany Lusch, J Nathan Kutz, and Steven L Brunton. "Deep learning for universal linear embeddings of nonlinear dynamics". In: *Nature communications* 9.1 (2018), pp. 1–10.

[13] Yinhao Zhu et al. "Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data". In: *Journal of Computational Physics* 394 (2019), pp. 56–81.

[14] Sebastian Kaltenbach and Phaedon-Stelios Koutsourelakis. "Incorporating physical constraints in a deep probabilistic machine learning framework for coarse-graining dynamical systems". In: *Journal of Computational Physics* 419 (2020).

[15] Frank Noé. "Machine learning for molecular dynamics on long timescales". In: *arXiv preprint arXiv:1812.07669* (2018).

[16] Hao Wu et al. "Variational Koopman models: slow collective variables and molecular kinetics from short off-equilibrium simulations". In: *The Journal of Chemical Physics* 146.15 (2017).

[17] Hao Wu et al. "Deep generative markov state models". In: *Advances in Neural Information Processing Systems*. 2018, pp. 3975–3984.

[18] B. O. Koopman. "Hamiltonian Systems and Transformations in Hilbert Space". In: *Proceedings of the National Academy of Sciences of the United States of America* 17.5 (1931), pp. 315–318. ISSN: 0027-8424. URL: `https://www.jstor.org/stable/86114` (visited on 12/30/2019).

[19] Stefan Klus et al. "Data-Driven Model Reduction and Transfer Operator Approximation". en. In: *Journal of Nonlinear Science* 28.3 (June 2018), pp. 985–1010. ISSN: 1432-1467. DOI: `10.1007/s00332-017-9437-7`. URL: `https://doi.org/10.1007/s00332-017-9437-7` (visited on 12/13/2019).

[20] Kathleen P Champion, Steven L Brunton, and J Nathan Kutz. "Discovery of nonlinear multiscale systems: Sampling strategies and embeddings". In: *SIAM Journal on Applied Dynamical Systems* 18.1 (2019), pp. 312–333.

[21] Craig Gin et al. "Deep Learning Models for Global Coordinate Transformations that Linearize PDEs". In: *arXiv preprint arXiv:1911.02710* (2019).

[22] Kookjin Lee and Kevin T Carlberg. "Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders". In: *Journal of Computational Physics* 404 (2020).

[23] L Felsberger and PS Koutsourelakis. "Physics-constrained, data-driven discovery of coarse-grained dynamics". In: *Communications in Computational Physics* 25.5 (2019), pp. 1259–1301. DOI: 10.4208/cicp.OA-2018-0174.

[24] Alexandre Chorin and Panagiotis Stinis. "Problem reduction, renormalization, and memory". In: *Communications in Applied Mathematics and Computational Science* 1.1 (2007), pp. 1–27.

[25] Richard Turner and Maneesh Sahani. "A Maximum-Likelihood Interpretation for Slow Feature Analysis". In: *Neural Computation* 19.4 (Apr. 2007), pp. 1022–1038. ISSN: 0899-7667. DOI: 10.1162/neco.2007.19.4.1022. URL: https://doi.org/10.1162/neco.2007.19.4.1022 (visited on 02/06/2020).

[26] Lazaros Zafeiriou et al. "Probabilistic slow features for behavior analysis". In: *IEEE transactions on neural networks and learning systems* 27.5 (2015), pp. 1034–1048.

[27] Robert Zwanzig. "Nonlinear generalized Langevin equations". In: *Journal of Statistical Physics* 9.3 (1973), pp. 215–220.

[28] Hazime Mori. "Transport, collective motion, and Brownian motion". In: *Progress of theoretical physics* 33.3 (1965), pp. 423–455.

[29] Matthew D Hoffman et al. "Stochastic variational inference". In: *The Journal of Machine Learning Research* 14.1 (2013), pp. 1303–1347.

[30] Evan Archer et al. "Black box variational inference for state space models". In: *arXiv preprint arXiv:1511.07367* (2015).

[31] Robert Bamler and Stephan Mandt. "Structured black box variational inference for latent time series models". In: *arXiv preprint arXiv:1707.01069* (2017).

[32] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[33] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[34] Stephen Roberts. "Convergence of a Random Walk Method for the Burgers Equation". In: *Mathematics of Computation* 52.186 (1989), pp. 647–673. ISSN: 0025-5718. DOI: 10.2307/2008486. URL: http://www.jstor.org/stable/2008486.

[35] Alina Chertock and Doron Levy. "Particle Methods for Dispersive Equations". In: *Journal of Computational Physics* 171.2 (Aug. 2001), pp. 708–730. ISSN: 0021-9991. DOI: 10.1006/jcph.2001.6803. URL: http://www.sciencedirect.com/science/article/pii/%20S0021999101968032.

[36] Ju Li et al. "Deciding the Nature of the Coarse Equation Through Microscopic Simulations: The Baby-Bathwater Scheme". In: *SIAM Rev.* 49.3 (July 2007), pp. 469–487. ISSN: 0036-1445. DOI: 10.1137/070692303. URL: http://dx.doi.org/10.1137/070692303 (visited on 03/07/2017).

Broader Impact

This paper deals with computational and mathematical modeling of multiscale dynamical systems of certain physical systems. Anticipated impact pertains primarily to the scientific community. Use in real-life applications would require additional adjustments and expert input.