# Multiway Ensemble Kalman Filter

**Yu Wang**
University of Michigan
`wayneyw@umich.edu`

**Alfred Hero**
University of Michigan
`hero@umich.edu`

## Abstract

In this work, we study the emergence of sparsity and multiway structures in second-order statistical characterizations of dynamical processes governed by partial differential equations (PDEs). We consider several state-of-the-art multiway covariance and inverse covariance (precision) matrix estimators and examine their pros and cons in terms of accuracy and interpretability in the context of physics-driven forecasting when incorporated into the ensemble Kalman filter (EnKF). In particular, we show that multiway data generated from the Poisson and the convection-diffusion types of PDEs can be accurately tracked via EnKF when integrated with appropriate covariance and precision matrix estimators.

## 1 Introduction

There has recently been a resurgence of interest in integrating machine learning with physics-based modeling. Much of the recent work has focused on black-box models such as deep neural networks [20, 14, 30, 24, 16, 25]. However, seeking shallower models that capture mechanism in a physically interpretable manner has been a recurring theme in both machine learning and physics [28]. In this paper, we introduce a high-dimensional statistical approach that naturally integrates physics and machine learning through Kronecker-structured Gaussian graphical models. The learned representation can then be incorporated into a high dimensional predictive model using the ensemble Kalman filtering framework.

**Multiway covariance/precision models.** High-dimensional multiway/tensor-variate data arise naturally in physical sciences. For example, weather satellites measure spatio-temporal climate variables such as temperature, wind velocity, sea level, pressure, etc. Due to the non-homogeneous nature of these data, second-order information that encodes (conditional) dependency structure within the data is of interest. Assuming the data are drawn from a tensor normal distribution, a straightforward way to estimate this structure is to vectorize the tensor and estimate the underlying Gaussian graphical model associated with the vector. However, such an approach ignores the tensor structure and requires estimating a rather high dimensional precision matrix, often with insufficient sample size. In many scientific applications the sample size can be as small as one when only a single tensor-valued measurement is available.

To address sample complexity challenges in learning second-order representations for multiway (tensor) data, sparsity is often imposed on the covariance $\Sigma$ or the inverse covariance $\Omega$. Such approaches include the sparse Kronecker product (KP) or Kronecker sum (KS) decomposition of $\Sigma$ or $\Omega$. Statistical models and corresponding learning algorithms can be derived using generative models or matrix approximations. The former include: KGlasso/Tlasso [23, 15] for estimating $\Omega = \mathbf{A} \otimes \mathbf{B}$, using a representation $\mathbf{AXB} = \mathbf{Z}$ for data $\mathbf{X}$ when $\mathbf{Z}$ is white noise. Another generative model is SyGlasso/SG-PALM [27, 26] that models the precision matrix as $\Omega = (\mathbf{A} \oplus \mathbf{B})^2$, which corresponds to assuming the data $\mathbf{X}$ obeys a Sylvester equation $\mathbf{XA} + \mathbf{BX} = \mathbf{Z}$. Matrix approximation methods include: KPCA [22, 8] that approximates the covariance matrix as $\Sigma = \sum_{i=1}^{r} \mathbf{A}_i \otimes \mathbf{B}_i$. Another matrix approximation method is the TeraLasso [9] that models the precision matrix as $\Omega = \mathbf{A} \oplus \mathbf{B}$.

TeraLasso is equivalent to approximation of the conditional dependency graph (encoded by the precision matrix) with a Cartesian product of smaller graphs [1].

**Multiway second-order characterization of dynamic processes.** Physical systems often exhibits sparsity and low-rank structures in their covariance or inverse covariance matrix. This is due to the fact that many physical systems are governed by differential equations, which are characterized by sparse differential operators. For instance, Wang and Hero [26] showed that for multivariate data generated by the Poisson equation, the discretized data has an inverse covariance matrix equal to a squared Kronecker sum of smaller sparse matrices. In related work, Lindgren et al. [13] elucidated a link between certain classes of Gaussian Fields (GF) and Gaussian Markov Random Fields (GMRF) via stochastic partial differential equations, and showed that efficient learning algorithms can be developed using the fact that GMRFs have sparse precision matrices.

For multiway/tensor-variate Gaussian data, the aforementioned multiway (inverse) covariance estimators have been shown to be statistically consistent in high-dimensional regimes when sample sizes ($N$) are much less than the dimensionality ($d$) of the covariates. An important question is whether these Kronecker structures can be integrated into the Kalman filter for tracking the states of a physical system that generates multiway data.

## 2 Numerical experiments: ensemble Kalman filtering

The Kalman filter is a well-known technique to track the states of a linear system over time, and many variants have been proposed to deal with non-linear systems, such as the extended and ensemble Kalman filters. The ensemble Kalman filter (EnKF) is particularly effective when the dynamical system is complicated and its gradient is infeasible to calculate, which is often the case in physical systems [5, 1]. However, such systems are often high-dimensional and the EnKF operates in the regime where the number of ensemble members, $N$, is much less than the size of the state, $d$, suggesting that sparse inverse covariance models will be especially attractive. Hou et al. [11] introduced a sparsity-penalized EnKF, which uses an estimator of the forecasting covariance whose inverse is sparsity regularized. Here we propose incorporating the multiway covariance / inverse covariance models discussed in Section 1 into the EnKF of Hou et al. [11].

For motivation, we consider the Poisson equation, an elliptical PDE that governs many physical processes including electromagnetic induction, heat transfer, and convection [2]. On a rectangular region $\Omega = (0, d_1) \times (0, d_2)$ in the 2D Cartesian plane, the Poisson equation with homogeneous Dirichlet boundary condition is expressed as

$$\mathcal{D}u = (\partial_x^2 + \partial_y^2)u = f \quad \text{in } \Omega,$$
$$u = 0 \quad \text{on } \partial\Omega \tag{1}$$

where $f : \Omega \to \mathbb{R}$ is the given source function and $u : \Omega \to \mathbb{R}$ is the unknown. Using the finite difference method with a square mesh grid with unit spacing, the unknown and the source can be expressed as $d_1$-by-$d_2$ matrices, $\mathbf{U}$ and $\mathbf{F}$, respectively, that are related to each other via

$$U_{i+1,j} + U_{i-1,j} + U_{i,j+1} + U_{i,j-1} - 4U_{i,j} = F_{i,j} \tag{2}$$

for any interior grid point $(i, j)$. Defining $n$-by-$n$ square matrix

$$\mathbf{A}_n = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix},$$

the relation (2) can be expressed as the (vectorized) Sylvester equation with $K = 2$:

$$(\mathbf{A}_{d_1} \oplus \mathbf{A}_{d_2})\mathbf{u} = \mathbf{f}, \tag{3}$$

where $\mathbf{u} = \text{vec}(\mathbf{U})$, $\mathbf{f} = \text{vec}(\mathbf{F})$. Note that $\mathbf{A}$ is tridiagonal. In the case where $\mathbf{f}$ is white noise with variance $\sigma^2$, the inverse covariance matrix of $\mathbf{u}$ has the form $\text{cov}^{-1}(\mathbf{u}) = \sigma^{-2}(\mathbf{A}_{d_1} \oplus \mathbf{A}_{d_2})^T(\mathbf{A}_{d_1} \oplus \mathbf{A}_{d_2})$ and hence sparse.

---

[1] Note that Tlasso, TeraLasso, Syglasso/SG-PALM are generalizable to precision matrices of the form $\bigotimes_{k=1}^{K} \mathbf{A}_k$, $\bigoplus_{k=1}^{K} \mathbf{A}_k$, and $(\bigoplus_{k=1}^{K} \mathbf{A}_k)^2$, respectively, for $K \geq 2$.

Here, we discuss two ways to extend the spatial Poisson equation described above to incorporate temporal dynamics, and illustrate how multiway (inverse) covariance models can be used to track spatio-temporal systems. The first extension, which we call the Poisson-AR(1) process, imposes an autoregressive temporal model of order 1 on the source function $f$ in the Poisson equation (1). Specifically, we say a sequence of discretized spatial observations $\{\mathbf{U}^k \in \mathbb{R}^{d_1 \times d_2}\}_k$ indexed by time step $k = 1, \cdots, T$ is from a Poisson-AR(1) process if

$$(\mathbf{A}_{d_1} \oplus \mathbf{A}_{d_2}) \operatorname{vec}(\mathbf{U}^k) = \operatorname{vec}(\mathbf{Z}^k), \tag{4}$$

$$\operatorname{vec}(\mathbf{Z}^k) = a \operatorname{vec}(\mathbf{Z}^{k-1}) + \operatorname{vec}(\mathbf{W}^k), \quad |a| < 1, \tag{5}$$

where $\mathbf{Z}^0 \sim \mathcal{N}(\mathbf{0}, \sigma_z^2 \mathbf{I})$ and $\{\mathbf{W}^k \in \mathbb{R}^{d_1 \times d_2}\}_k$ is spatiotemporal white noise, i.e., $W_{i,j}^k \sim \mathcal{N}(0, \sigma_w^2)$, i.i.d.

The second time-varying extension of the Poisson PDE model (1) is the convection-diffusion process [2]

$$\frac{\partial u}{\partial t} = \theta \sum_{i=1}^{2} \frac{\partial^2 u}{\partial x_i^2} - \epsilon \sum_{i=1}^{2} \frac{\partial u}{\partial x_i}. \tag{6}$$

Here, $\theta > 0$ is the diffusivity; and $\epsilon \in \mathbb{R}$ is the convection velocity of the quantity along each coordinate. Note that for simplicity of discussion here, we assume these coefficients do not change with space and time (see, Stocker [19], for example, for a detailed discussion). These equations are closely related to the Navier-Stokes equation commonly used in stochastic modeling for weather and climate prediction [2, 19]. Coupled with Maxwell's equations, these equations can be used to model magneto-hydrodynamics [17], which characterize solar activities including flares.

A solution of Equation (6) can be approximated similarly as in the Poisson equation case, through a finite difference approach. Denote the discrete spatial samples of $u(\mathbf{x}, t)$ at time $t_k$ as a matrix $\mathbf{U}^k \in \mathbb{R}^{d_1 \times d_2}$. We obtain a discretized update propagating $u(\mathbf{x}, t)$ in space and time, which locally satisfies

$$
\begin{aligned}
\frac{U_{i,j}^k - U_{i,j}^{k-1}}{\Delta t} &= \theta \left( \frac{U_{i+1,j}^k + U_{i-1,j}^k + U_{i,j+1}^k + U_{i,j-1}^k - 4U_{i,j}^k}{h^2} \right) \\
&\quad - \epsilon \left( \frac{U_{i+1,j}^k - U_{i-1,j}^k + U_{i,j+1}^k - U_{i,j-1}^k}{2h} \right),
\end{aligned}
\tag{7}
$$

where $\Delta t = t_{k+1} - t_k$ is the time step and $h$ is the mesh step (spatial grid spacing). For $\Delta t = 1$ and $h = 1$, $\mathbf{U}^k$ can be shown to obey the Sylvester matrix update equation [21] $\mathbf{A}_{\theta,\epsilon} \mathbf{U}^k + \mathbf{U}^k \mathbf{B}_{\theta,\epsilon}^T = \mathbf{U}^{k-1}$, or equivalently,

$$(\mathbf{B}_{\theta,\epsilon} \oplus \mathbf{A}_{\theta,\epsilon}) \operatorname{vec}(\mathbf{U}^k) = \operatorname{vec}(\mathbf{U}^{k-1}), \tag{8}$$

where $\mathbf{A}_{\theta,\epsilon} \in \mathbb{R}^{d_1 \times d_1}$ and $\mathbf{B}_{\theta,\epsilon} \in \mathbb{R}^{d_2 \times d_2}$ are symmetric tridiagonal matrices whose entries depend on $\theta, \epsilon$, and $h$ [7]. Note that, although $\{\mathbf{U}^k\}_k$ updates according to a sparse Kronecker sum, the inverse covariance of the marginal distribution that smooths over time is not sparse.

We turn Equation (8) into a state space model for the observations $\mathbf{X}$ by adding i.i.d. state noise $\mathbf{W}^k \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I})$, i.i.d. observation noise $\mathbf{V}^k \sim \mathcal{N}(0, \sigma_v^2 \mathbf{I})$ and a state measurement matrix $\mathbf{H}$:

$$\operatorname{vec}(\mathbf{X}^k) = \mathbf{H} \operatorname{vec}(\mathbf{U}^k) + \operatorname{vec}(\mathbf{V}^k), \tag{9}$$

$$\operatorname{vec}(\mathbf{U}^k) = (\mathbf{B}_{\theta,\epsilon} \oplus \mathbf{A}_{\theta,\epsilon})^{-1} \operatorname{vec}(\mathbf{U}^{k-1}) + \operatorname{vec}(\mathbf{W}^k). \tag{10}$$

We model the observed process $\mathbf{X}$ as an incomplete noisy version of the convection-diffusion state $\mathbf{U}_t$ obeying the Sylvester matrix update equation above discretization the convection-diffusion becomes assume a linear Gaussian state-space model for the observed process $\mathbf{X}_t$ governed by convection-diffusion dynamics:

$$
\begin{aligned}
\mathbf{A}_{\theta,\epsilon} \mathbf{U}_t + \mathbf{U}_t \mathbf{A}_{\theta,\epsilon} &= \mathbf{U}_{t-1}, \\
\mathbf{X}_t &= \mathbf{U}_t + \mathbf{V}_t,
\end{aligned}
$$

where $\mathbf{V}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is an i.i.d. Gaussian white noise. Note that in general one might not fully observe the state, leading to a partially observed measurement process

$$\operatorname{vec}(\mathbf{X}_t) = \mathbf{H} \operatorname{vec}(\mathbf{U}_t) + \operatorname{vec}(\mathbf{V}_t),$$

3

where $\mathbf{H}$ is a measurement matrix that can incorporate effects such as unobserved, masked, or superposed states. It is also possible to take into account the error in the dynamic process, i.e.,

$$\text{vec}(\mathbf{U}_t) = (\mathbf{A}_{\theta,\epsilon} \oplus \mathbf{A}_{\theta,\epsilon})^{-1} \text{vec}(\mathbf{U}_{t-1}) + \text{vec}(\mathbf{W}_t,)$$

where $\text{vec}(\mathbf{W}_t)$ is assumed to be white noise. Under the case of perfect observation ($\mathbf{H}$ is the identity matrix and $\mathbf{W}_t = \mathbf{0}$), Note that although the state variable evolves via a Sylvester equation, similar to the Poisson equation case, the state (inverse) covariance matrix at time step $t_k$ admits different structures. Specifically, the state precision matrix $\mathbf{\Omega}^k = \text{cov}^{-1}(\text{vec}(\mathbf{U}^k)) \in \mathbb{R}^{d_1 d_2 \times d_1 d_2}$ evolves as $\mathbf{\Omega}^k = (\mathbf{B}_{\theta,\epsilon} \oplus \mathbf{A}_{\theta,\epsilon})\mathbf{\Omega}^{k-1}(\mathbf{B}_{\theta,\epsilon} \oplus \mathbf{A}_{\theta,\epsilon}) + \sigma_w^{-2}\mathbf{I}$ (see Katzfuss et al. [12], for example). This matrix is not necessarily sparse for finite $k$ but, assuming that the eigenvalues of the matrix $\mathbf{B}_{\theta,\epsilon} \oplus \mathbf{A}_{\theta,\epsilon}$ are in $(-1, 1)$, the limiting precision matrix $\mathbf{\Omega}^\infty = \lim_{k\to\infty} \mathbf{\Omega}^k$ is $\mathbf{\Omega}^\infty = (\mathbf{B}_{\theta,\epsilon} \oplus \mathbf{A}_{\theta,\epsilon})\mathbf{\Omega}^\infty(\mathbf{B}_{\theta,\epsilon} \oplus \mathbf{A}_{\theta,\epsilon}) + \sigma^{-2}\mathbf{I}$. The $\mathbf{\Omega}^\infty$ matrix is sparse because $\mathbf{A}_{\theta,\epsilon}$ and $\mathbf{B}_{\theta,\epsilon}$ are both tridiagonal.

To illustrate, we consider a 2D spatio-temporal process of dimension $64 \times 64$ where only half of the entries are observed, which leads to a measurement matrix $\mathbf{H} \in \{0, 1\}^{2048 \times 4069}$. We generated the true states and the corresponding observations according to Convection-Diffusion and Poisson dynamics for $T = 20$ time steps. Several realizations of the true state variables are shown in Figure 1. At each time step, we generated an ensemble of size $N = 25$ and estimated the state covariance / inverse covariance using several sparse (multiway) inverse covariance estimation methods, including Glasso [6], KPCA [8], KGlasso [23], TeraLasso [9], SG-PALM [26]. Figure 2 shows evolution of the computed root mean squared errors (RMSEs) for the estimated states across all ensemble members. In Figure 3 we show the true and estimated (inverse) covariance matrices obtained for the last time step. The Poisson process is time-invariant, and at each time step the EnKF involves estimation of a sparse Kronecker sum squared inverse covariance matrix. Hence, the SG-PALM method operates under the correct model assumption in this situation. On the other hand, the inverse covariance structure under the convection-diffusion dynamics model is dense due to the smoothing effect of Kalman filtering and the nature of the temporal dynamics. But, its steady-state covariance has low-dimensional structures. The KPCA in this case was able to approximate this structure, as illustrated in Figure 3.
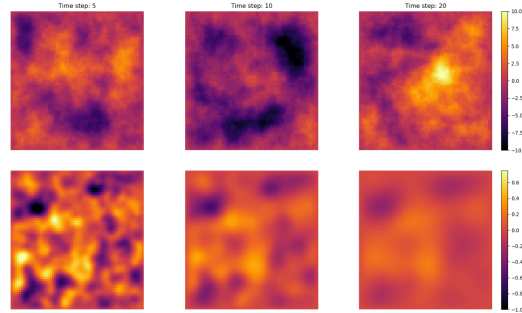


Figure 1: 2D convection-diffusion (bottom) and Poisson (top) state variables at three different time stamps. Note that there is temporal correlation exists in the convection-diffusion states while the Poisson states are temporally independent.

*Remark.* The proposed multiway EnKF is able to track systems governed by elliptic (e.g., Poisson) parabolic, and hyperbolic (e.g., convection-diffusion) PDEs. There are indeed other important PDEs / dynamical systems that cannot be modelled by these types of equations. Furthermore, the Sylvester matrix equations arise when the finite-difference discretization is performed on a rectangular grid. The relations (2) and (7) might not hold for finite-difference on, for example, spherical coordinates.

## 3 Future work

**Applications.** Spatiotemporal PDEs are prominent techniques for modeling real-world physical systems. One such system arises in space physics, where solar flares and coronal mass ejections are associated with rapid changes in filed connectivity and are powered by partial dissipation of electrical currents in the solar atmosphere [18]. The nonlinear force-free filed model is often used to describe the solar coronal magnetic field [4, 29] and can be derived from the convection-diffusion
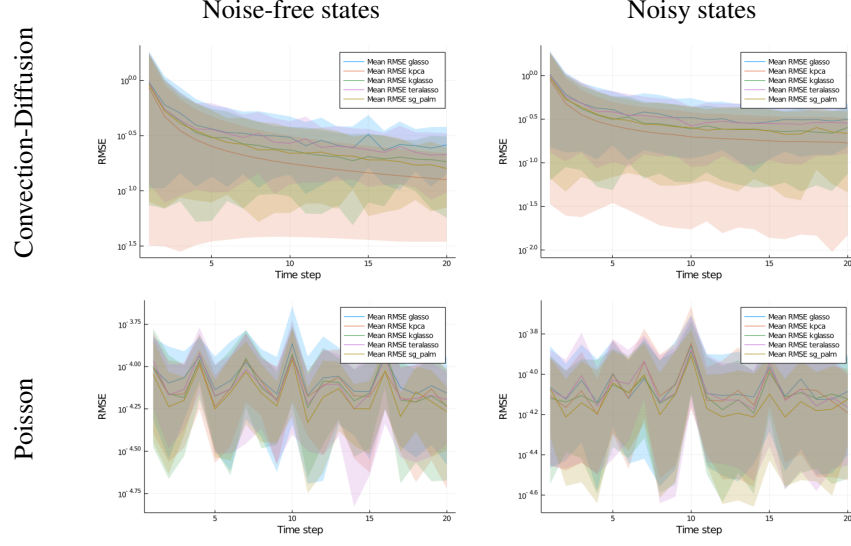
Figure 2: RMSEs of the estimated states via EnKF over 20 time steps using different (inverse) covariance estimators. RMSEs over all ensemble members are shown here with the mean highlighted using solid lines. Here, each state is of dimension $64 \times 64$ and is generated via either a convection-diffusion (top row) or Poisson equation (bottom row). The best performers in terms of mean RMSE over all ensemble members are KPCA for convection-diffusion and SG-PALM for Poisson.
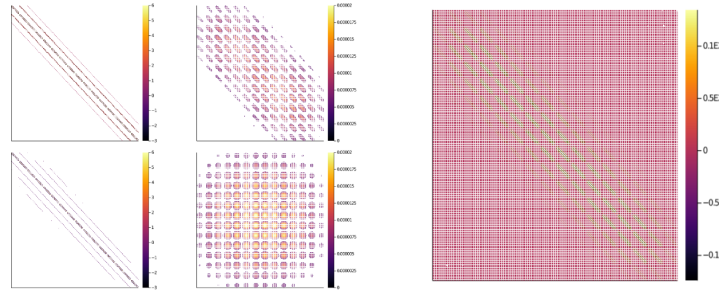


Figure 3: Covariance/precision structures for Poisson and Convection-Diffusion dynamics and their estimates. Here, white/blank entries indicate zeros in the (inverse) covariance matrix. For Poisson dynamics the Sylvester graphical model matches the true structure of the precision matrix. For convection-diffusion dynamics the covariance instead of the precision matrix is structured and sparse.

process described in this work. Additionally, global maps of the solar photospheric magnetic flux are fundamental drivers for simulations of the corona and solar wind. However, observations of the solar photosphere are only made intermittently over approximately half of the solar surface. Hickmann et al. [10] introduced the Air Force Data Assimilative Photospheric Flux Transport model that uses localized ensemble Kalman filtering to adjust a set of photospheric simulations to agree with the available observations. In future work, we plan to incorporate our proposed multiway EnKF framework for tracking these solar physical systems.

**Interpretability of factorization-based multiway model.** While the Kronecker products expansion used in KPCA captures dense structures in the covariance matrix of data generated from more complex spatio-temporal physical processes, it lacks physical interpretability. In contrast to the case of Sylvester graphical model and Poisson processes, it is not obvious whether the sum of Kronecker products structure corresponds to any true physical models. Recent work in quantum informatics [3] has demonstrated a link between estimation of the density matrix for entangled quantum states and the structured tensor approximation via $\sum_{i=1}^{r} \mathbf{A}_i \otimes \mathbf{B}_i$. Further characterizing these connections and extending them to study its connections with certain classes of discretized PDEs would be an interesting future direction.

## Acknowledgments and Disclosure of Funding

## References

[1] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble kalman filter. *Monthly weather review*, 126(6):1719–1724, 1998.

[2] Subrahmanyan Chandrasekhar. Stochastic problems in physics and astronomy. *Reviews of modern physics*, 15(1):1, 1943.

[3] Moody T Chu and Matthew M Lin. Nonlinear power-like and svd-like iterative schemes with applications to entangled bipartite rank-1 approximation. *SIAM Journal on Scientific Computing*, (0):S448–S474, 2021.

[4] ML DeRosa, MS Wheatland, KD Leka, Graham Barnes, Tahar Amari, Aurélien Canou, SA Gilchrist, JK Thalmann, Gherardo Valori, Thomas Wiegelmann, et al. The influence of spatial resolution on nonlinear force-free modeling. *The Astrophysical Journal*, 811(2):107, 2015.

[5] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99 (C5):10143–10162, 1994.

[6] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[7] Lars Grasedyck. Existence and computation of low kronecker-rank approximations for large linear systems of tensor product structure. *Computing*, 72(3-4):247–265, 2004.

[8] Kristjan Greenewald and Alfred O Hero. Robust kronecker product pca for spatio-temporal covariance estimation. *IEEE Transactions on Signal Processing*, 63(23):6368–6378, 2015.

[9] Kristjan Greenewald, Shuheng Zhou, and Alfred Hero III. Tensor graphical lasso (teralasso). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(5):901–931, 2019.

[10] Kyle S Hickmann, Humberto C Godinez, Carl J Henney, and C Nick Arge. Data assimilation in the adapt photospheric flux transport model. *Solar Physics*, 290(4):1105–1118, 2015.

[11] Elizabeth Hou, Earl Lawrence, and Alfred O Hero. Penalized ensemble kalman filters for high dimensional non-linear systems. *PloS one*, 16(3):e0248046, 2021.

[12] Matthias Katzfuss, Jonathan R Stroud, and Christopher K Wikle. Understanding the ensemble kalman filter. *The American Statistician*, 70(4):350–357, 2016.

[13] Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

[14] Zichao Long, Yiping Lu, Xianzhong Ma, and Bin Dong. Pde-net: Learning pdes from data. In *International Conference on Machine Learning*, pages 3208–3216. PMLR, 2018.

[15] Xiang Lyu, Will Wei Sun, Zhaoran Wang, Han Liu, Jian Yang, and Guang Cheng. Tensor graphical model: Non-convex optimization and statistical inference. *IEEE transactions on pattern analysis and machine intelligence*, 42(8):2024–2037, 2019.

[16] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, et al. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.

[17] B Roberts. Slow magnetohydrodynamic waves in the solar atmosphere. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 364(1839): 447–460, 2006.

[18] CJ Schrijver, ML DeRosa, Thomas Metcalf, Graham Barnes, Bruce Lites, T Tarbell, James McTiernan, Gherardo Valori, Thomas Wiegelmann, MS Wheatland, et al. Nonlinear force-free field modeling of a solar active region around the time of a major flare and coronal mass ejection. *The Astrophysical Journal*, 675(2):1637, 2008.

[19] Thomas Stocker. *Introduction to climate modelling*. Springer Science & Business Media, 2011.

[20] Naoya Takeishi, Yoshinobu Kawahara, and Takehisa Yairi. Learning koopman invariant subspaces for dynamic mode decomposition. *arXiv preprint arXiv:1710.04340*, 2017.

[21] James William Thomas. *Numerical partial differential equations: finite difference methods*, volume 22. Springer Science & Business Media, 2013.

[22] Theodoros Tsiligkaridis and Alfred O Hero. Covariance estimation in high dimensions via kronecker product expansions. *IEEE Transactions on Signal Processing*, 61(21):5347–5360, 2013.

[23] Theodoros Tsiligkaridis, Alfred O Hero III, and Shuheng Zhou. On convergence of kronecker graphical lasso algorithms. *IEEE transactions on signal processing*, 61(7):1743–1755, 2013.

[24] Pantelis R Vlachas, Wonmin Byeon, Zhong Y Wan, Themistoklis P Sapsis, and Petros Koumoutsakos. Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2213):20170844, 2018.

[25] Rui Wang, Karthik Kashinath, Mustafa Mustafa, Adrian Albert, and Rose Yu. Towards physics-informed deep learning for turbulent flow prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1457–1466, 2020.

[26] Yu Wang and Alfred Hero. Sg-palm: a fast physically interpretable tensor graphical model. *International Conference on Machine Learning (ICML), arXiv preprint arXiv:2105.12271*, 2021.

[27] Yu Wang, Byoungwook Jang, and Alfred Hero. The sylvester graphical lasso (syglasso). In *International Conference on Artificial Intelligence and Statistics*, pages 1943–1953. PMLR, 2020.

[28] E Weinan, Jiequn Han, and Linfeng Zhang. Integrating machine learning with physics-based modeling. *arXiv*, 2020.

[29] MS Wheatland and SA Gilchrist. The state of nonlinear force-free magnetic field extrapolation. In *Journal of Physics: Conference Series*, volume 440, page 012037. IOP Publishing, 2013.

[30] Linfeng Zhang, Jiequn Han, Han Wang, Wissam Saidi, Roberto Car, and Weinan E. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/e2ad76f2326fbc6b56a45a56c59fafdb-Paper.pdf.

## Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Both at the end of the Section 2 and the future work discussed in Section 3.

    (c) Did you discuss any potential negative societal impacts of your work? [N/A]

    (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...

    (a) Did you state the full set of assumptions of all theoretical results? [N/A]

    (b) Did you include complete proofs of all theoretical results? [N/A]

3. If you ran experiments...

    (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] The code is included as part of a software package, available at: `https://github.com/ywa136/TensorGraphicalModels.jl`

    (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]

    (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]

    (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] All numerical experiments were run on a laptop equipped with `Intel Core i5 Quad-Core 1.4 GHz` with a `16 GB RAM`.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

    (a) If your work uses existing assets, did you cite the creators? [N/A]

    (b) Did you mention the license of the assets? [N/A]

    (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]

    (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]

    (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

    (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

    (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

    (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]