Komatsuzaki Seminar
13:00~, March 9th, 2026.

Toward scientific discovery with AI
Part II

Yoh-ichi Mototake
Hitotsubashi university

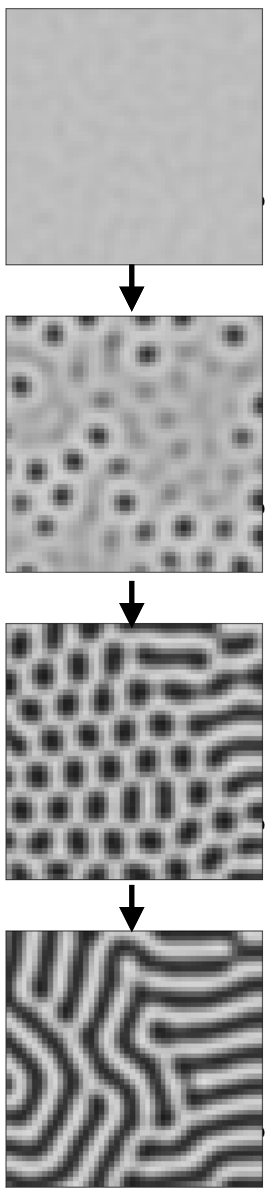
My approaches for Interpretable AI in physics

Our approach to interpretable AI

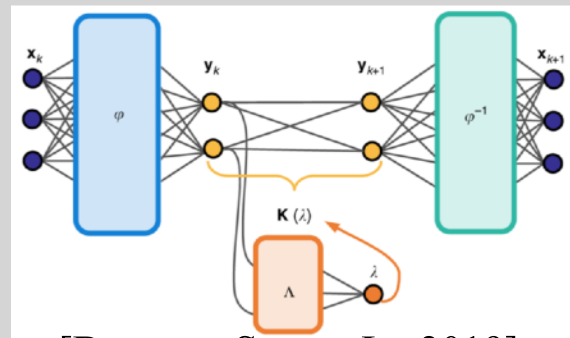
Part I | Part II

Complex phenomenon

dynamics



Deep neural network-type approach



[Brunton, Steven L.. 2019]

Cons: low interpretability

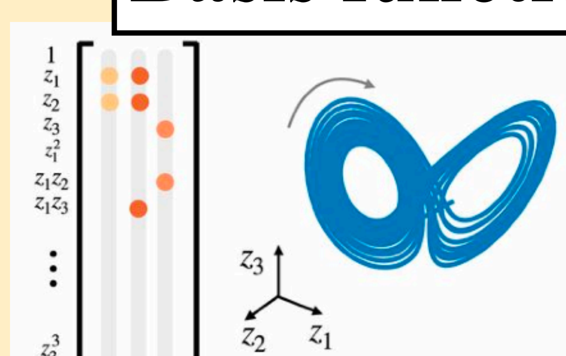
extract **Interpretable** physical information from DNN

Bayesian inference

Interpretable and high expressive model

Part III

Basis function type approach



[Steven L. Brunton et al., PNAS, 2016]

Cons: Poor expressive power

Construction of a reduced model based on **topological data analysis**

Expressive Power

Interpretability

Hitotsubashi University

Interpretable reduced modeling of large-scale pattern dynamics

- from materials science to astrophysics -

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

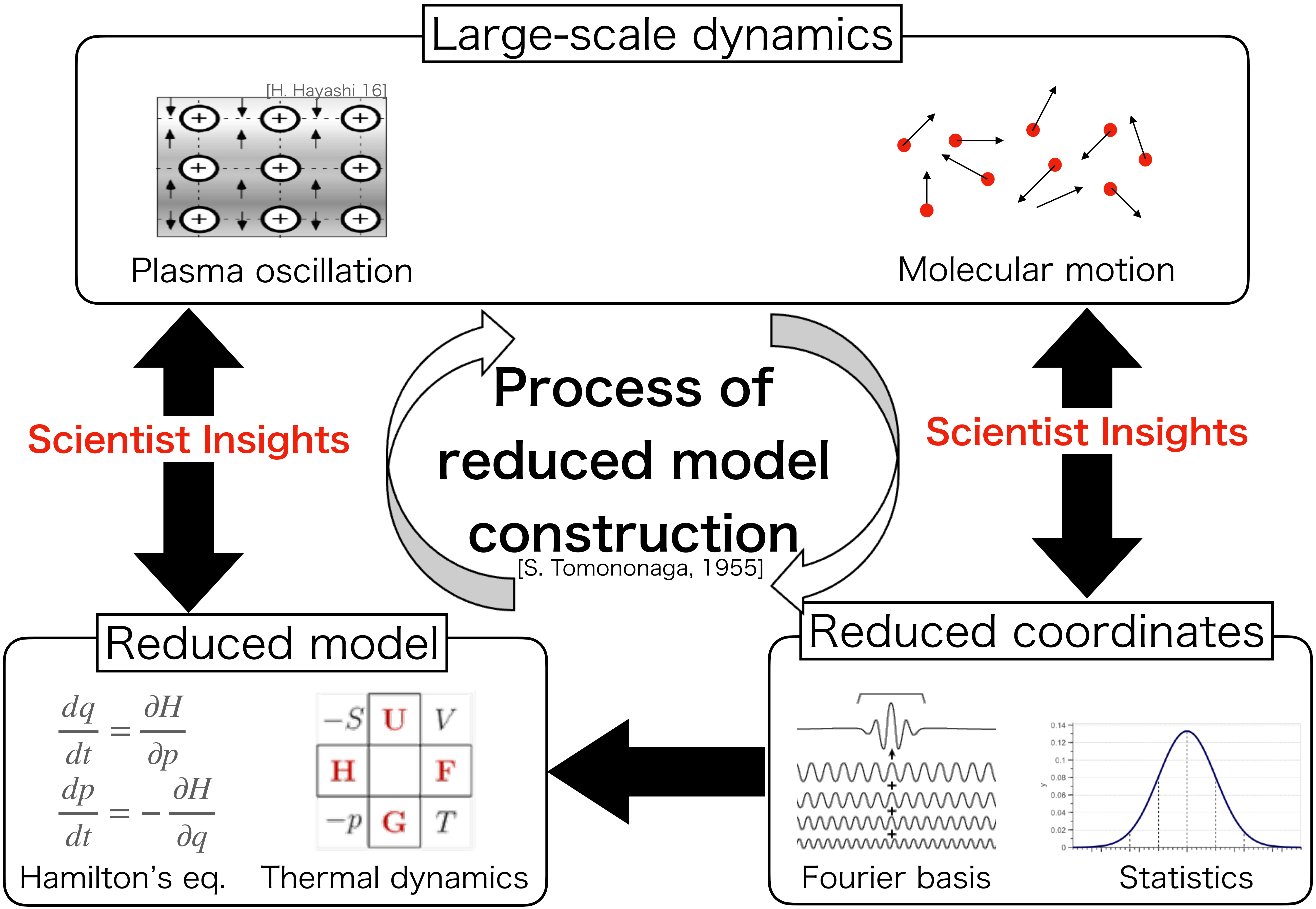
Yoh-ichi Mototake

Hitotsubashi university (Japan)

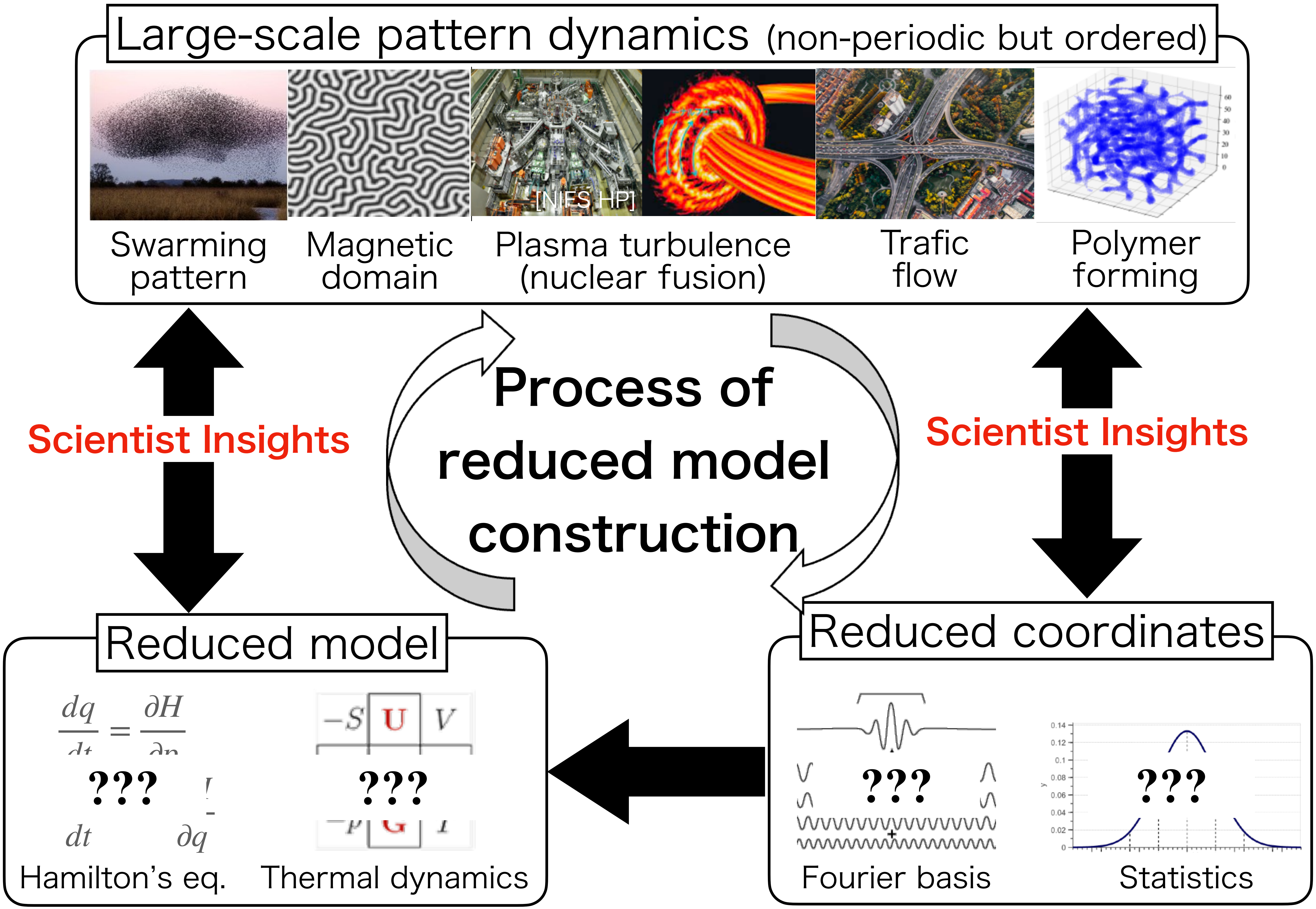
PRESTO

SAKIGAKE

Reduced Modeling of Large System



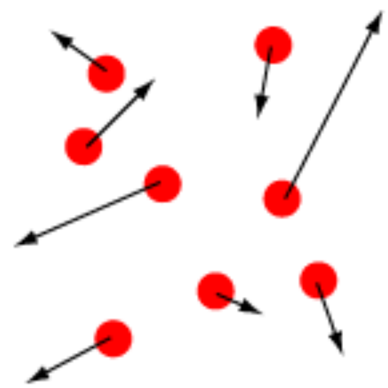
Reduced Modeling of Complex Systems



Reduced Modeling of Complex Systems

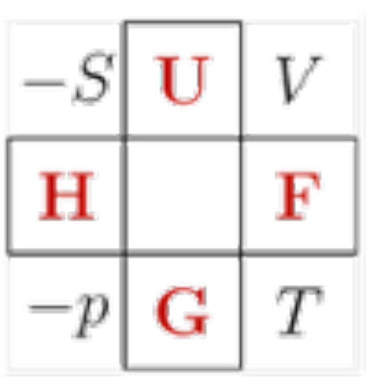
Random ⟵⟶ **Ordered**

Random structure

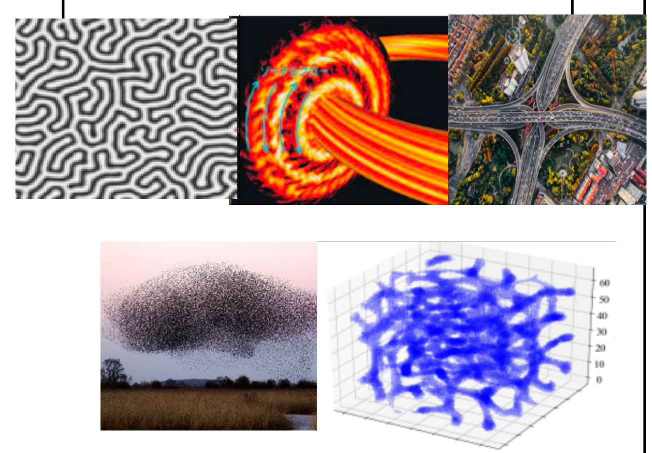


Molecular motion of gas

Statistics

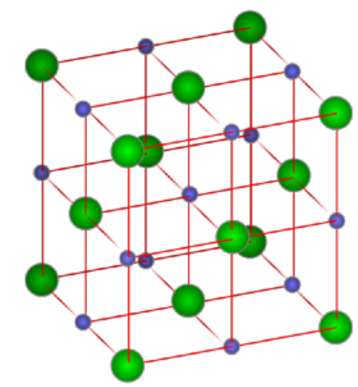


Non-periodic ordered structure



???

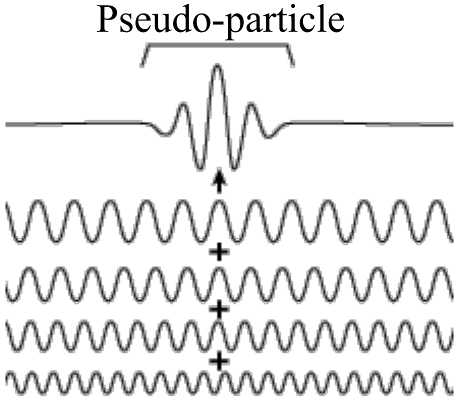
Periodic structure



NaCl crystal (Wikipedia)

Fourier basis

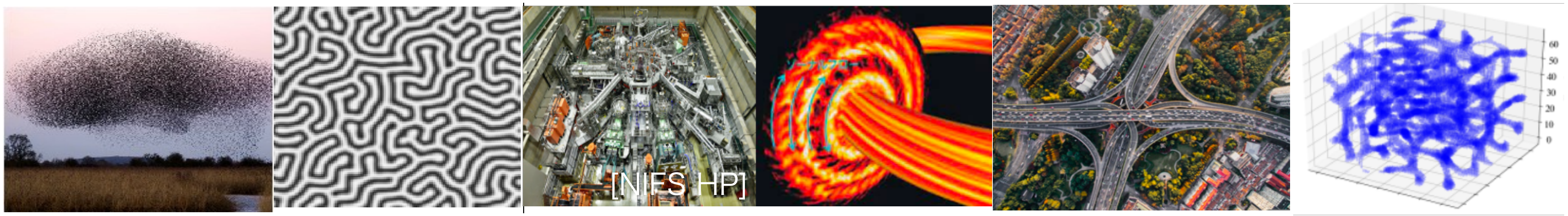
Pseudo-particle



http://k-hiura.cocolog-nifty.com/

ML and Reduced Modeling

Large-scale pattern dynamics



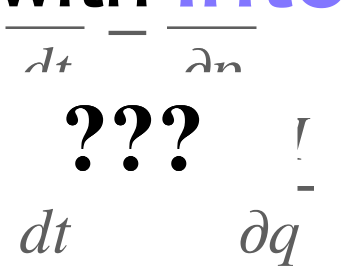
Swarming pattern | Magnetic domain | Plasma turbulence (nuclear fusion) | Trafic flow | Polymer forming

Scientist Insights
+ Machine learning

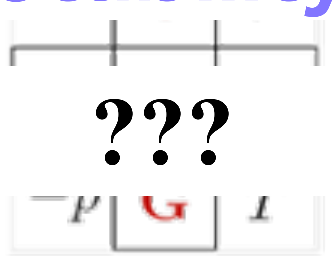
Scientist Insights
+ Machine learning

Reduced model

with Interpretability



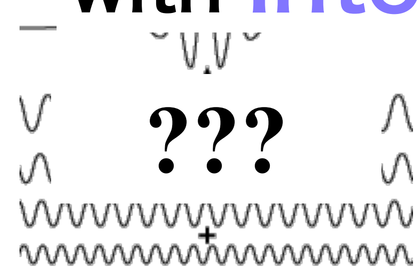
Hamilton's eq.



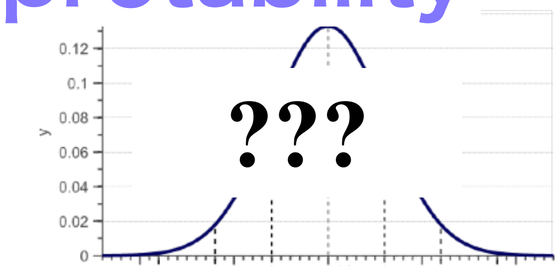
Thermal dynamics

Reduced coordinates

with Interpretability



Fourier basis



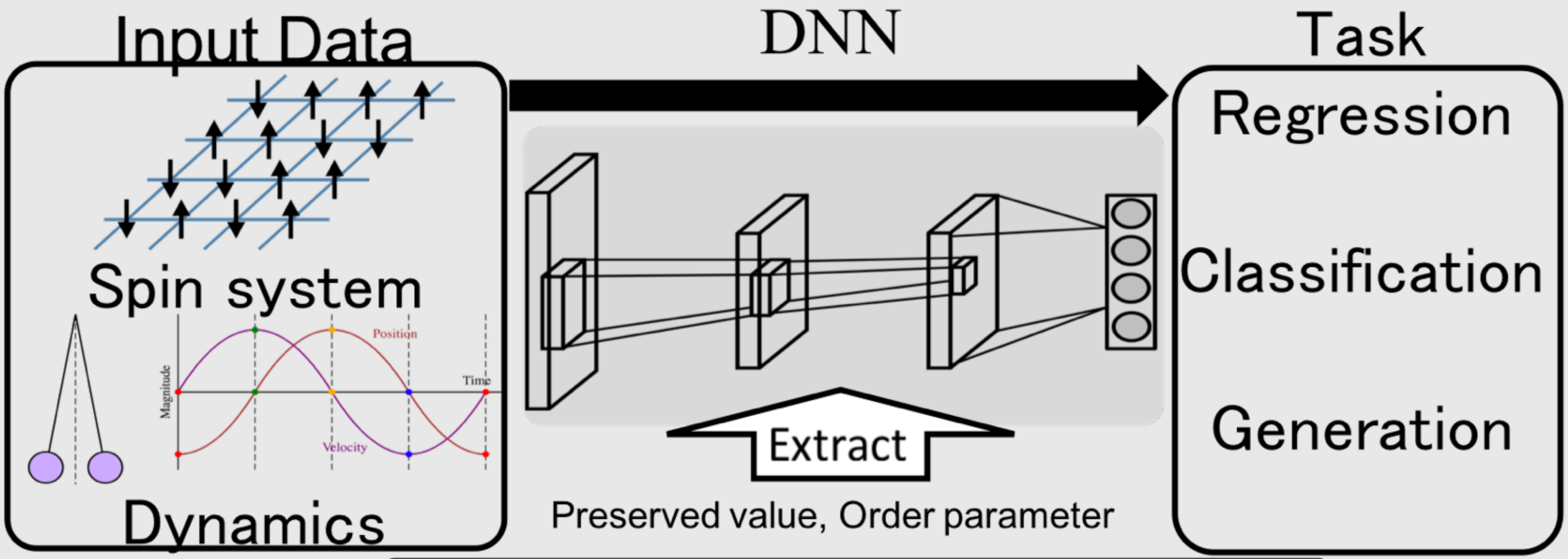
Statistics

Interpretable AI for reduced coordinate search

Reduce coordinates
Timeseries dataset

Conservation laws

Proposed methods



$$\boldsymbol{m} = \frac{1}{N}\sum_{i=1}^{N} \mu_0 \,\boldsymbol{\sigma}_i \qquad \boldsymbol{L} \equiv \boldsymbol{r} \times \boldsymbol{p}$$

Extracting conservation laws from a trained DNN

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Outline of research

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

▶The purpose of the proposed framework is not to analyze physical data with deep learning but to extract interpretable physical information from trained DNNs.

▶With Noether's theorem and by an efficient sampling method, the proposed framework infers conservation laws by extracting the symmetries of dynamics from trained DNNs.

▶The proposed framework is developed by deriving the relationship between a manifold structure of a time-series data set and the necessary conditions for Noether's theorem.

Related researches refering our study:

[Ziming Liu and Max Tegmark, Phys. Rev. Lett. 126 180604 (2021)]
[Seungwoong Ha and Hawoong Jeong, Phys. Rev. Research 3 L042035 (2021)]
[Han Zhang, Huawei Fan, Liang Wang, and Xingang Wang, Phys. Rev. E 104 024205 (2021)]

Outline of research

Theory

1. Time series data manifold and Noether's theorem
2. Deep neural networks and the manifold hypothesis

Method

3. Method to extract symmetries from trained DNN
4. Method to estimate conservation laws from extracted symmetries

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Outline of research

Theory

1. Time series data manifold and Noether's theorem
2. Deep neural networks and the manifold hypothesis

Method

3. Method to extract symmetries from trained DNN
4. Method to estimate conservation laws from extracted symmetries

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Noether's theorem：

▶A theorem linking the continuous symmetry of Hamiltonian systems to conservation laws [Noether 1918].

If the Hamiltonian $H(\boldsymbol{q}, \boldsymbol{p})$ and the canonical equation of motion $\dfrac{\partial H(\boldsymbol{q},\boldsymbol{p})}{\partial q_i} = -\dot{p}_i$, $\dfrac{\partial H(\boldsymbol{q},\boldsymbol{p})}{\partial p_i} = \dot{q}_i$ are invariant with respect to an infinitesimal transformation $(q_i', p_i') = (q_i + \delta q_{ij}, p_i + \delta p_{ij})$, then the following relation holds with the conserved value G_j.

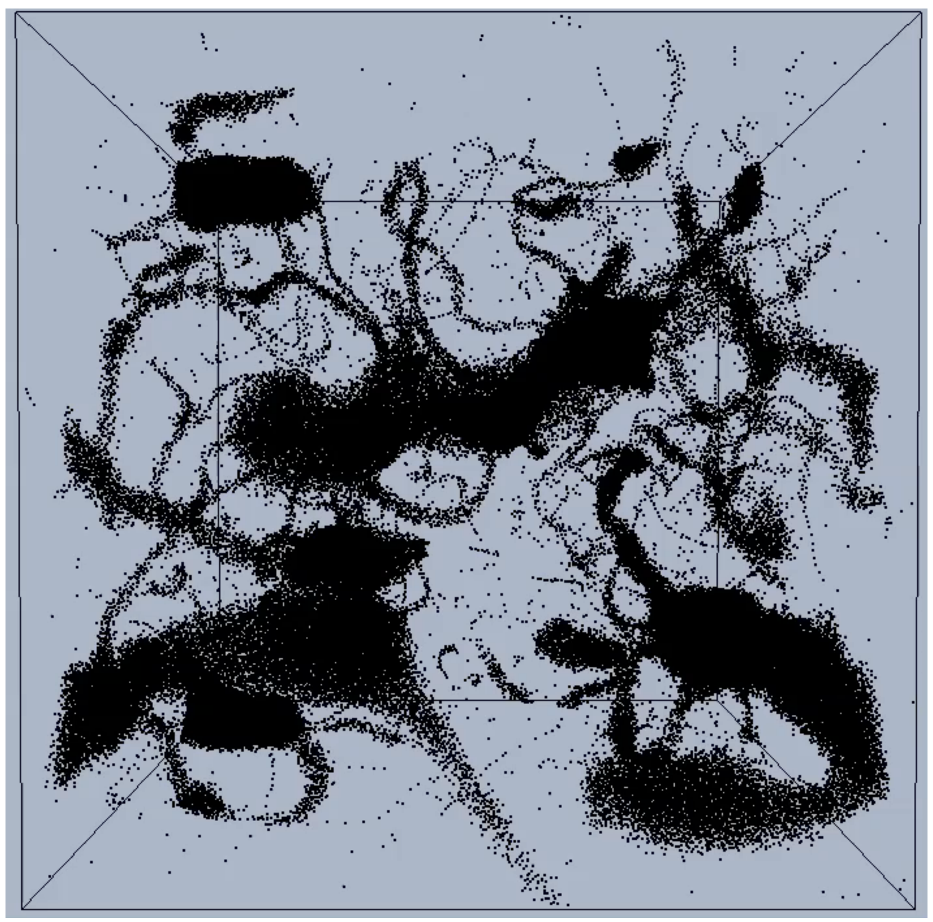
$$(\delta q_{ij}, \delta p_{ij}) = \left(\frac{\partial G_j}{\partial p_i}, -\frac{\partial G_j}{\partial q_i}\right)$$

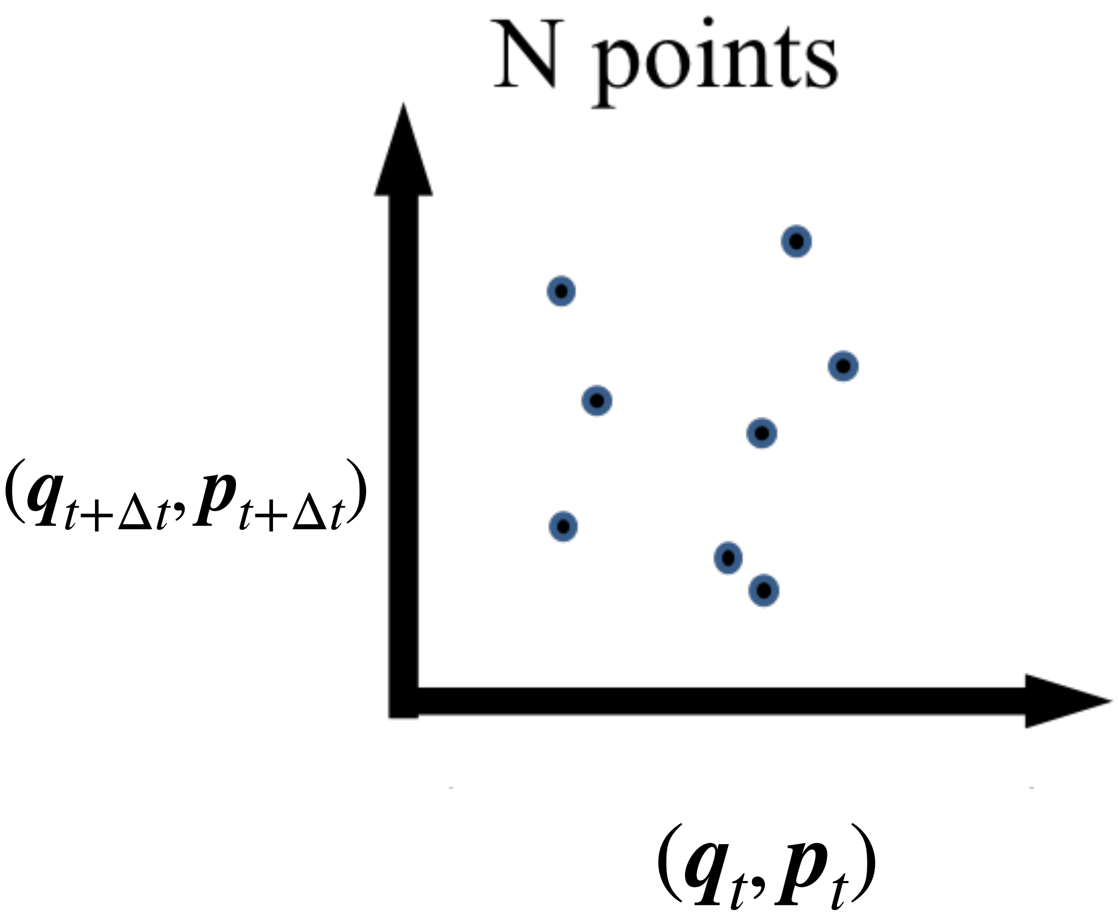
[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Time series data of a dynamical system in this study

=The set of possible states of the system in phase space and the state after Δtseconds.

$$D := \left\{ \boldsymbol{q}^i_{t_i}, \boldsymbol{p}^i_{t_i}, \boldsymbol{q}^i_{t_i+\Delta t}, \boldsymbol{p}^i_{t_i+\Delta t} \right\}_{i=1}^{N}$$





[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Symmetry of Hamiltonian system

Invariance of the Hamiltonian : $H'(\boldsymbol{q},\boldsymbol{p}) \equiv H(\boldsymbol{q},\boldsymbol{p}),\ H'(\boldsymbol{Q},\boldsymbol{P}) := H\left(\boldsymbol{q}(\boldsymbol{Q},\boldsymbol{P}),\boldsymbol{p}(\boldsymbol{Q},\boldsymbol{P})\right)$

Invariance of the canonical equation of motion : $\begin{cases} \boldsymbol{q}_{T+\Delta T} = \dfrac{\partial H(\boldsymbol{q}_T,\boldsymbol{p}_T)}{\partial \boldsymbol{p}_T}\Delta T + \boldsymbol{q}_T \\ \boldsymbol{p}_{T+\Delta T} = -\dfrac{\partial H(\boldsymbol{q}_T,\boldsymbol{p}_T)}{\partial \boldsymbol{q}_T}\Delta T + \boldsymbol{p}_T \end{cases}$

The transformation of the $(2d+2)$-dimensional space $(\mathbf{q},\mathbf{p})$ is defined as

$$\mathfrak{c} : \Gamma \times \mathbb{R} \times \mathbb{R} \longrightarrow \Gamma \times \mathbb{R} \times \mathbb{R}, \quad (5)$$

$$(\mathbf{q},\mathbf{p}) \longmapsto (\mathbf{Q},\mathbf{P}) := (\mathbf{Q}(\mathbf{q},\mathbf{p}),\mathbf{P}(\mathbf{q},\mathbf{p})), \quad (6)$$

Symmetry of time series data manifolds

A **time series data manifold** that takes energy E at time t

$$\forall E,\ \left\{ \boldsymbol{q}_{t+\Delta t},\boldsymbol{p}_{t+\Delta t},\boldsymbol{q}_t,\boldsymbol{p}_t \;\middle|\; H(\boldsymbol{q}_t,\boldsymbol{p}_t) = E, \boldsymbol{p}_{t+\Delta t} = \boldsymbol{p}_t - \frac{\partial H(\boldsymbol{q}_t,\boldsymbol{p}_t)}{\partial \boldsymbol{q}_t}, \boldsymbol{q}_{t+\Delta t} = \boldsymbol{q}_t + \frac{\partial H(\boldsymbol{q}_t,\boldsymbol{p}_t)}{\partial \boldsymbol{p}_t} \right\}$$

$$= \left\{ \boldsymbol{Q}_{T+\Delta T},\boldsymbol{P}_{T+\Delta T},\boldsymbol{Q}_T,\boldsymbol{P}_T \;\middle|\; H(\boldsymbol{q}_t,\boldsymbol{p}_t) = E, \boldsymbol{p}_{t+\Delta t} = \boldsymbol{p}_t - \frac{\partial H(\boldsymbol{q}_t,\boldsymbol{p}_t)}{\partial \boldsymbol{q}_t}, \boldsymbol{q}_{t+\Delta t} = \boldsymbol{q}_t + \frac{\partial H(\boldsymbol{q}_t,\boldsymbol{p}_t)}{\partial \boldsymbol{p}_t} \right\}$$

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Facilitating the problem: relaxing the requirement condition $\forall E$

$$\forall E, \left\{ \boldsymbol{q}_{t+\Delta t}, \boldsymbol{p}_{t+\Delta t}, \boldsymbol{q}_t, \boldsymbol{p}_t \;\middle|\; H(\boldsymbol{q}_t, \boldsymbol{p}_t) = E, \boldsymbol{p}_{t+\Delta t} = \boldsymbol{p}_t - \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{q}_t}, \boldsymbol{q}_{t+\Delta t} = \boldsymbol{q}_t + \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{p}_t} \right\}$$

$$= \left\{ \boldsymbol{Q}_{T+\Delta T}, \boldsymbol{P}_{T+\Delta T}, \boldsymbol{Q}_T, \boldsymbol{P}_T \;\middle|\; H(\boldsymbol{q}_t, \boldsymbol{p}_t) = E, \boldsymbol{p}_{t+\Delta t} = \boldsymbol{p}_t - \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{q}_t}, \boldsymbol{q}_{t+\Delta t} = \boldsymbol{q}_t + \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{p}_t} \right\}$$

Discretization of energy

$$\{\mathbf{Q}_i(\mathbf{q}, \mathbf{p}), \mathbf{P}_i(\mathbf{q}, \mathbf{p}) \mid \forall i \in \Lambda_E, \text{satisfy Eq. (32)}\} = \bigcap_{i \in \Lambda_E} \{\mathbf{Q}_i(\mathbf{q}, \mathbf{p}), \mathbf{P}_i(\mathbf{q}, \mathbf{p}) \mid \text{satisfy Eq. (32)}\}.$$

$$\left\{ \mathbf{q}_{t+\Delta t}, \mathbf{p}_{t+\Delta t}, \mathbf{q}_t, \mathbf{p}_t \;\middle|\; H(\mathbf{q}_t, \mathbf{p}_t) = E_i, \mathbf{p}_{t+\Delta t} = \mathbf{p}_t - \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{q}_t}, \mathbf{q}_{t+\Delta t} = \mathbf{q}_t + \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{p}_t} \right\}$$
$$= \left\{ \mathbf{Q}_{T+\Delta T}, \mathbf{P}_{T+\Delta T}, \mathbf{Q}_T, \mathbf{P}_T \;\middle|\; H(\mathbf{q}_t, \mathbf{p}_t) = E_i, \mathbf{p}_{t+\Delta t} = \mathbf{p}_t - \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{q}_t}, \mathbf{q}_{t+\Delta t} = \mathbf{q}_t + \frac{\partial H(\mathbf{q}_t, \mathbf{p}_t)}{\partial \mathbf{p}_t} \right\}. \quad (32)$$

Necessary conditions

$$S_i := \left\{ \boldsymbol{q}_{t+\Delta t}, \boldsymbol{p}_{t+\Delta t}, \boldsymbol{q}_t, \boldsymbol{p}_t \;\middle|\; H(\boldsymbol{q}_t, \boldsymbol{p}_t) = E_i, \boldsymbol{p}_{t+\Delta t} = \boldsymbol{p}_t - \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{q}_t}, \boldsymbol{q}_{t+\Delta t} = \boldsymbol{q}_t + \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{p}_t} \right\}$$

➡The goal is now to investigate the symmetry of the time series data manifold S_i at one specific energy E_i ! **[Step1]**

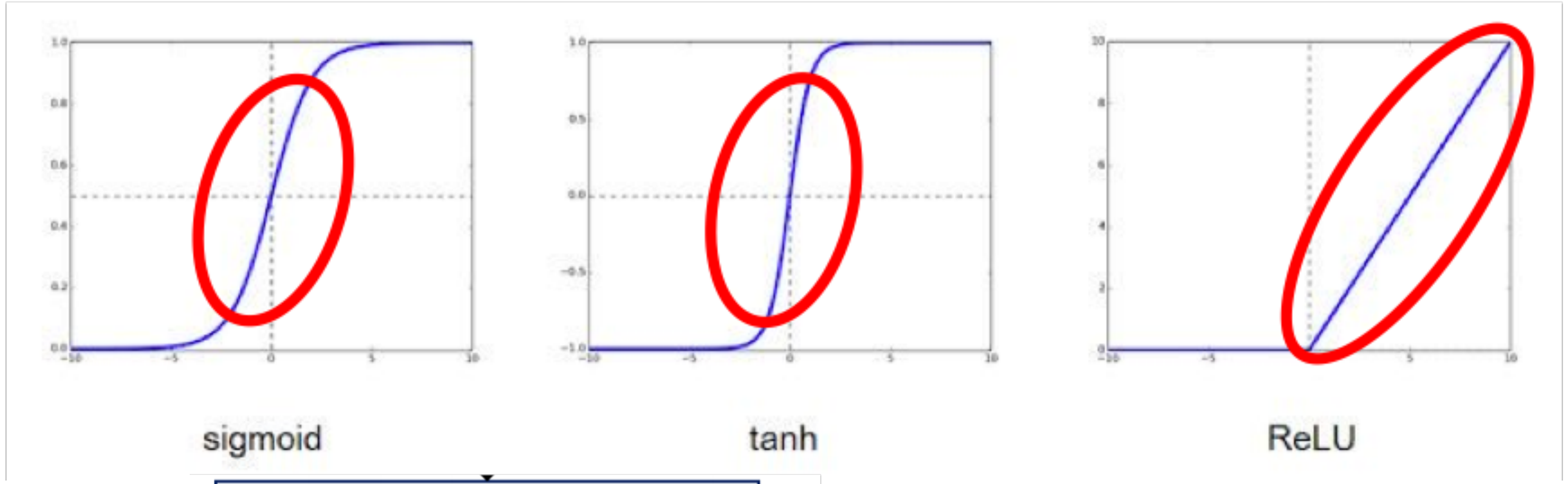
[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

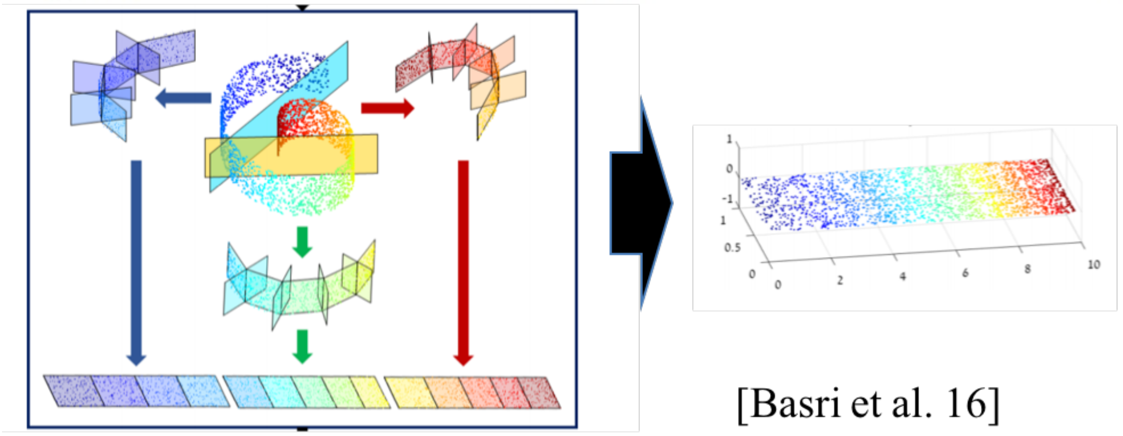
Neural networks and data manifolds

$$\boldsymbol{h} = (h_1, h_2, \cdots, h_{d_h})$$

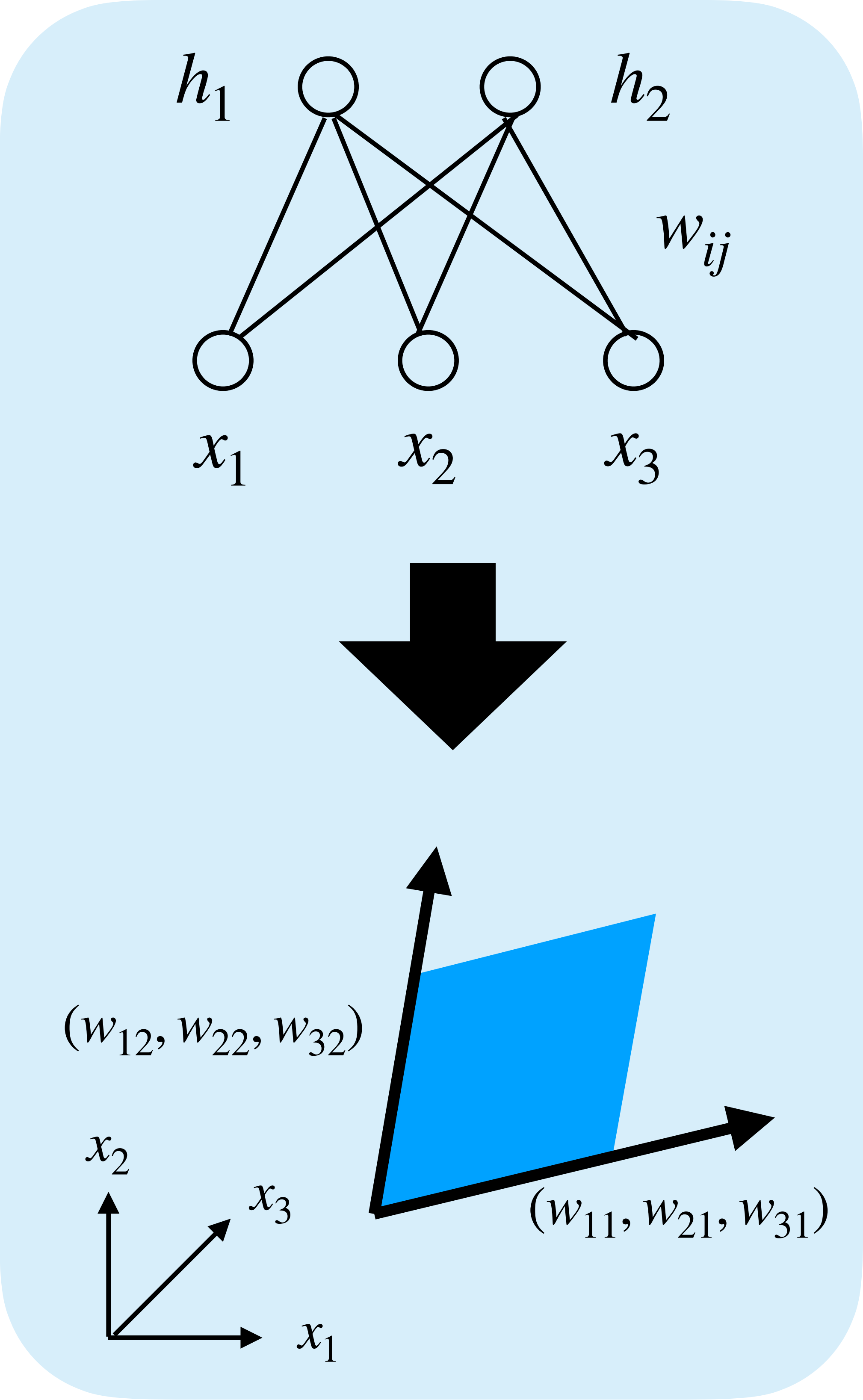
$$\boldsymbol{\varphi}(\boldsymbol{w}^{\text{in}}\boldsymbol{x}) = (\varphi_1, \varphi_2, \cdots, \varphi_{d_h}), \varphi_j = \varphi\left[\sum_i^{d_{\text{in}}}\left(w_{ij}^{\text{in}} x_i\right)\right]$$

Activation function : $\varphi(x)$





[Basri et al. 16]



Neural networks can model manifolds by pasting segmental hyperplanes.

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

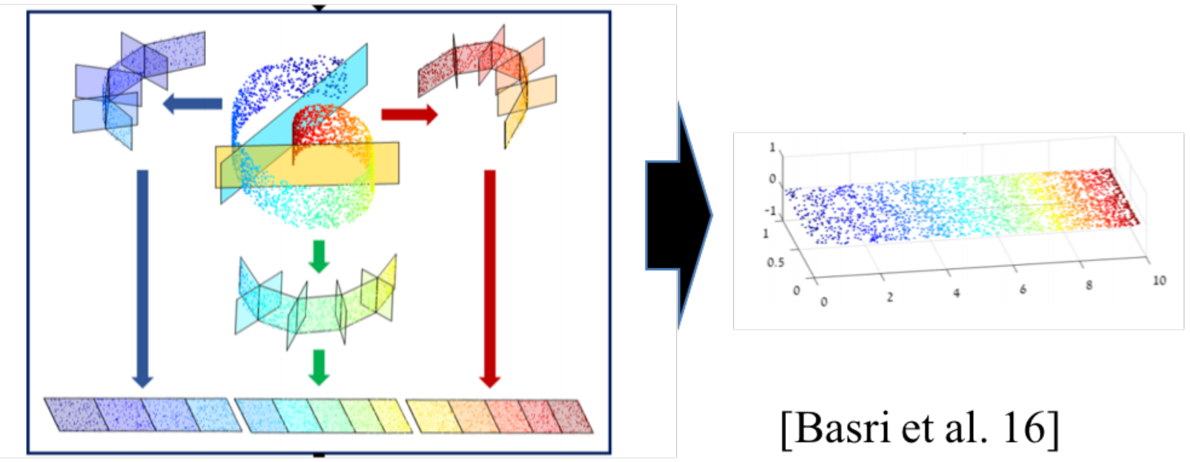
Summary of theory part

- If we know the symmetry of the time series data manifold S_i with respect to the coordinate transformation, we can estimate the conservation law.

$$S_i := \left\{ \boldsymbol{q}_{t+\Delta t}, \boldsymbol{p}_{t+\Delta t}, \boldsymbol{q}_t, \boldsymbol{p}_t \;\middle|\; H(\boldsymbol{q}_t, \boldsymbol{p}_t) = E_i, \boldsymbol{p}_{t+\Delta t} = \boldsymbol{p}_t - \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{q}_t}, \boldsymbol{q}_{t+\Delta t} = \boldsymbol{q}_t + \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{p}_t} \right\}$$

- DNNs can represent manifolds as a kind of hyperplane pasting.

➡A trained DNN would model S_i.



[Basri et al. 16]

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Outline of research

Theory

1. Time series data manifold and Noether's theorem
2. Deep neural networks and the manifold hypothesis

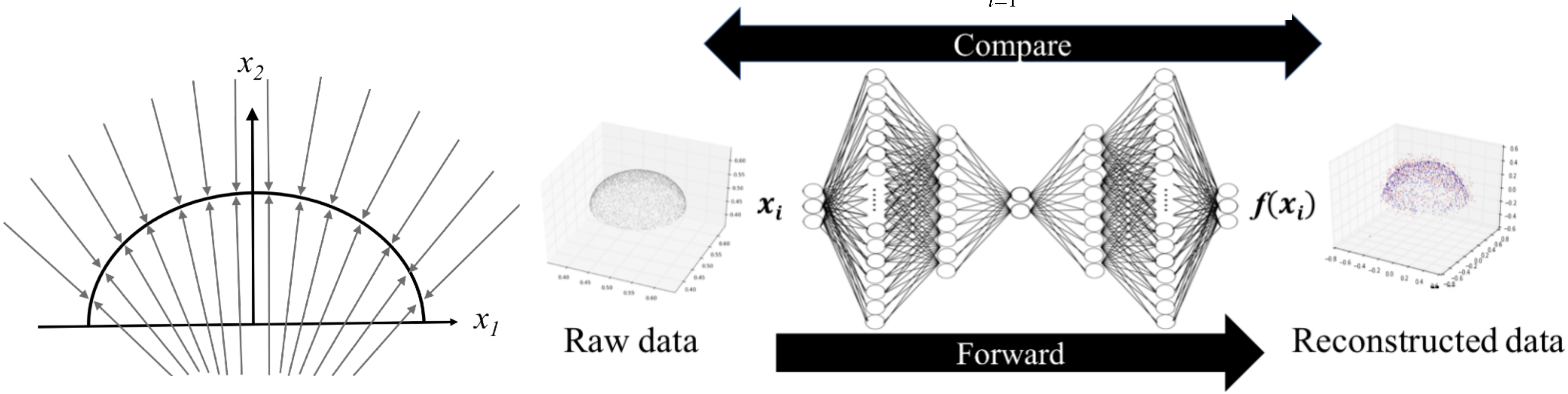
Method

3. Method 1 to extract symmetries from trained DNN
4. Method 2 to estimate conservation laws from extracted symmetries

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Method 1 to extract symmetries from trained DNN

$$E_{\mathrm{samp}}[X(\cdot)] = \frac{1}{N}\sum_{i=1}^{N}\left\{X(x_i) - f_{\mathrm{DNN}}[X(x_i)]\right\}^2$$



The set of symmetric transformations is obtained as the set of transformations

$$\boldsymbol{Q}(\cdot,\cdot), \boldsymbol{P}(\cdot,\cdot) \text{ which satisfy: } \left\{ \boldsymbol{Q}(\cdot,\cdot), \boldsymbol{P}(\cdot,\cdot) \;\middle|\; \underset{\boldsymbol{Q}(\cdot,\cdot),\boldsymbol{P}(\cdot,\cdot)}{\arg\min}\, E_{\mathrm{samp}}\left[\boldsymbol{Q}(\cdot,\cdot), \boldsymbol{P}(\cdot,\cdot)\right] \right\}$$

Solve as sampling problem

$$\mathrm{P}\left(\boldsymbol{Q}(\cdot,\cdot), \boldsymbol{P}(\cdot,\cdot)\right) \sim \frac{1}{Z}\exp\left\{-\frac{N}{2\sigma^2}E_{\mathrm{samp}}\left[\boldsymbol{Q}(\cdot,\cdot), \boldsymbol{P}(\cdot,\cdot)\right]\right\}$$

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Implementing physics constraints or prior to Method 1 knowledge20

● **Introducing physical constraints using a Bayesian inferential framework**

$$\mathrm{P}(a_{01}, a_{02}, a_{11}, \cdots, a_{2d\,2d}) = \frac{1}{Z}\exp\left[-\frac{N}{2\sigma^2}E_{\mathrm{samp}}(a_{01}, a_{02}, a_{11}, \cdots, a_{2d\,2d})\right]$$

$$A\boldsymbol{x} + A_0 = \begin{pmatrix} a_{11} & \dots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{d1} & \dots & a_{dd} \end{pmatrix}\boldsymbol{x} + \begin{pmatrix} a_{01} \\ \vdots \\ a_{0d} \end{pmatrix}$$

↓ **physics constraints or prior knowledge** ↓

$$\mathrm{P}(a_{01}, a_{02}, a_{11}, \cdots, a_{2d\,2d}) = \frac{1}{Z}\exp\left[-\frac{N}{2\sigma^2}E_{\mathrm{samp}}(a_{01}, a_{02}, a_{11}, \cdots, a_{2d\,2d})\right] \times q(a_{01}, \cdots, a_{2d\,2d})$$

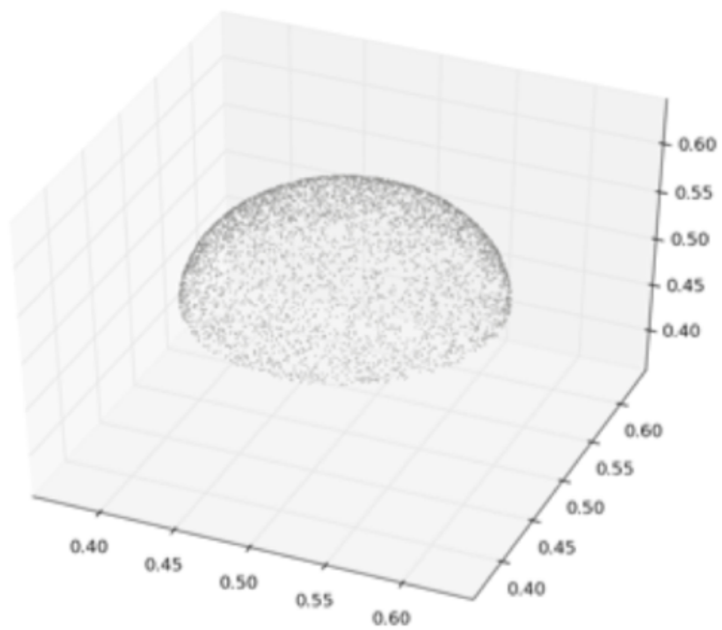
$$q(a_{01}, \cdots, a_{2d\,2d}) = \begin{cases} \text{const.} & \text{for} \quad \det A = 1 \\ 0 & \text{for} \quad \det A \neq 1 \end{cases}$$

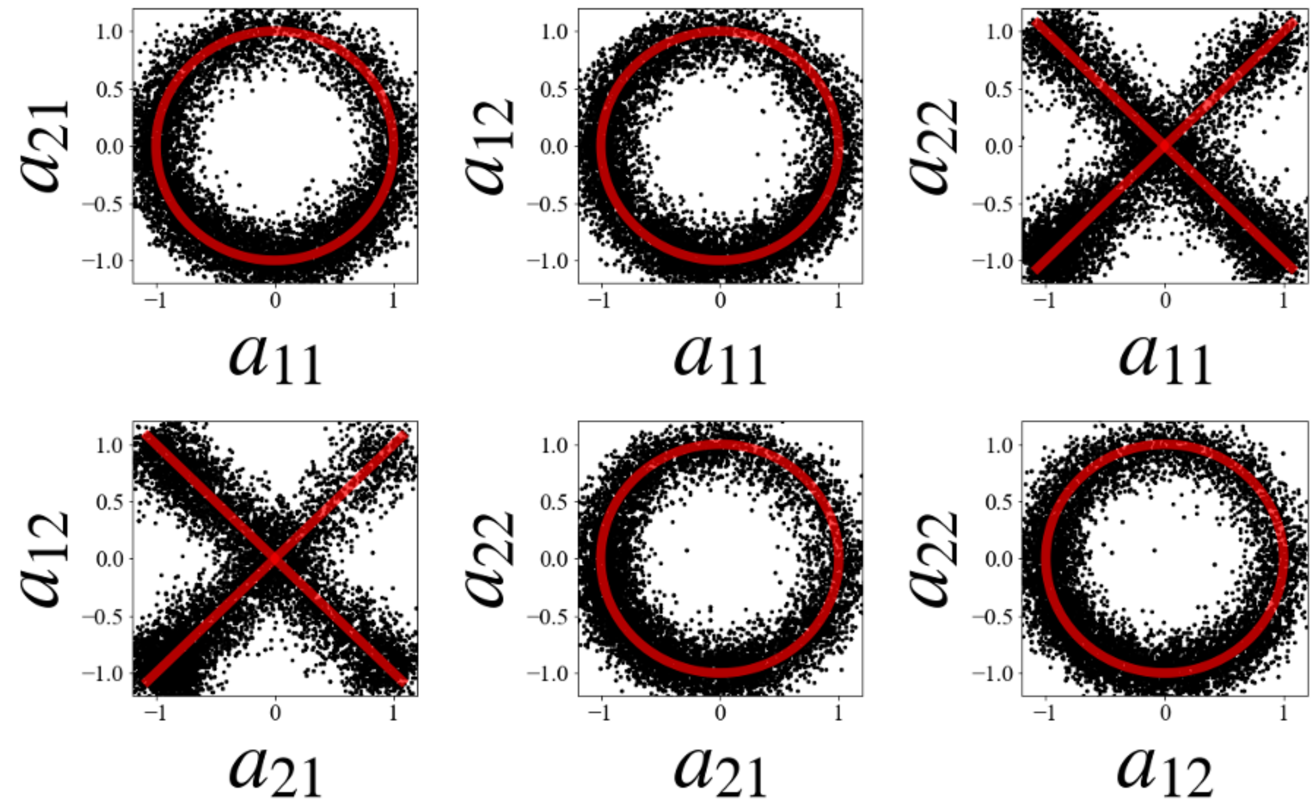
▶ Liouville's theorem (law of conservation of volume)

[Step2]

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

A set of symmetric transformations $D_a := \{(a_{01}, a_{02}\cdots, a_{0\,2d}, a_{11}\cdots, a_{2d\,2d})_{n_a}\}_{n_a=1}^{N_a}$





Manifolds of Lie groups

$$(\delta q_{ij}, \delta p_{ij}) = \left(\varepsilon \left. \frac{\partial Q_i(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{\theta})}{\partial \theta_j} \right|_{\boldsymbol{\theta}=\vec{0}}, \varepsilon \left. \frac{\partial P_i(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{\theta})}{\partial \theta_j} \right|_{\boldsymbol{\theta}=\vec{0}} \right)$$

Infinitesimal transformation

→

$$M_{\text{invariant}} \sim \left\{ A(\boldsymbol{\theta}) \begin{pmatrix} \boldsymbol{q} \\ \boldsymbol{p} \end{pmatrix} + A_0(\boldsymbol{\theta}) \,\middle|\, \boldsymbol{\theta} \in \mathbb{R}^{d_\theta} \right\}$$

Tangent spaces of manifolds of Lie groups $T_I M_{\text{invariant}}$

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Set of symmetric transformations

$$D_a := \{(a_{01}, a_{02}\cdots, a_{0\,2d}, a_{11}\cdots, a_{2d\,2d})_{n_a}\}_{n_a=1}^{N_a}$$

Model the manifold of a Lie group in implicit representation.

$$\begin{cases} f_1(a'_1, \cdots, a'_{d'}) = 0 \\ \quad\vdots \\ f_{d'-d_\theta}(a'_1, \cdots, a'_{d'}) = 0 \end{cases}$$

$$A'(\boldsymbol{\theta}) = (a'_1(\boldsymbol{\theta}), \cdots, a'_{d'}(\boldsymbol{\theta})) := (a_{01}(\boldsymbol{\theta}), \cdots, a_{0d}(\boldsymbol{\theta}), a_{11}(\boldsymbol{\theta}), \cdots, a_{1d}(\boldsymbol{\theta}), \cdots, a_{d1}(\boldsymbol{\theta}), \cdots, a_{2d\,2d}(\boldsymbol{\theta}))$$

Estimating infinitesimal transformations from implicit function models

$$\begin{pmatrix} \delta \boldsymbol{q}_l \\ \delta \boldsymbol{p}_l \end{pmatrix} = \varepsilon \left.\frac{A(b_l)}{\partial b_l}\right|_{A=\boldsymbol{I}} \begin{pmatrix} \boldsymbol{q} \\ \boldsymbol{p} \end{pmatrix} + \varepsilon \left.\frac{A_0(b_l)}{\partial b_l}\right|_{A_0=\boldsymbol{0}} = \varepsilon \begin{pmatrix} \frac{\partial a_{1\,1}}{\partial b_l}\Big|_{A=\boldsymbol{I}} & \cdots & \frac{\partial a_{d\,1}}{\partial b_l}\Big|_{A=\boldsymbol{I}} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_{1\,d}}{\partial b_l}\Big|_{A=\boldsymbol{I}} & \cdots & \frac{\partial a_{d\,d}}{\partial b_l}\Big|_{A=\boldsymbol{I}} \end{pmatrix} \begin{pmatrix} \boldsymbol{q} \\ \boldsymbol{p} \end{pmatrix} + \varepsilon \begin{pmatrix} \frac{\partial a_{0\,1}}{\partial b_l}\Big|_{A=\boldsymbol{I}} \\ \vdots \\ \frac{\partial a_{0\,d}}{\partial b_l}\Big|_{A=\boldsymbol{I}} \end{pmatrix}$$

$$(b_1, b_2, \cdots, b_{d_\theta}) \subset A', \qquad \{c_k\}_{k=1}^{d'-d_\theta} := A' \backslash \{b_l\}_{l=1}^{d_\theta}, \qquad c_k = g_i(b_1, \cdots, b_{d_\theta})$$

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Extract knowledge by exploring the noise parameter of MCMC

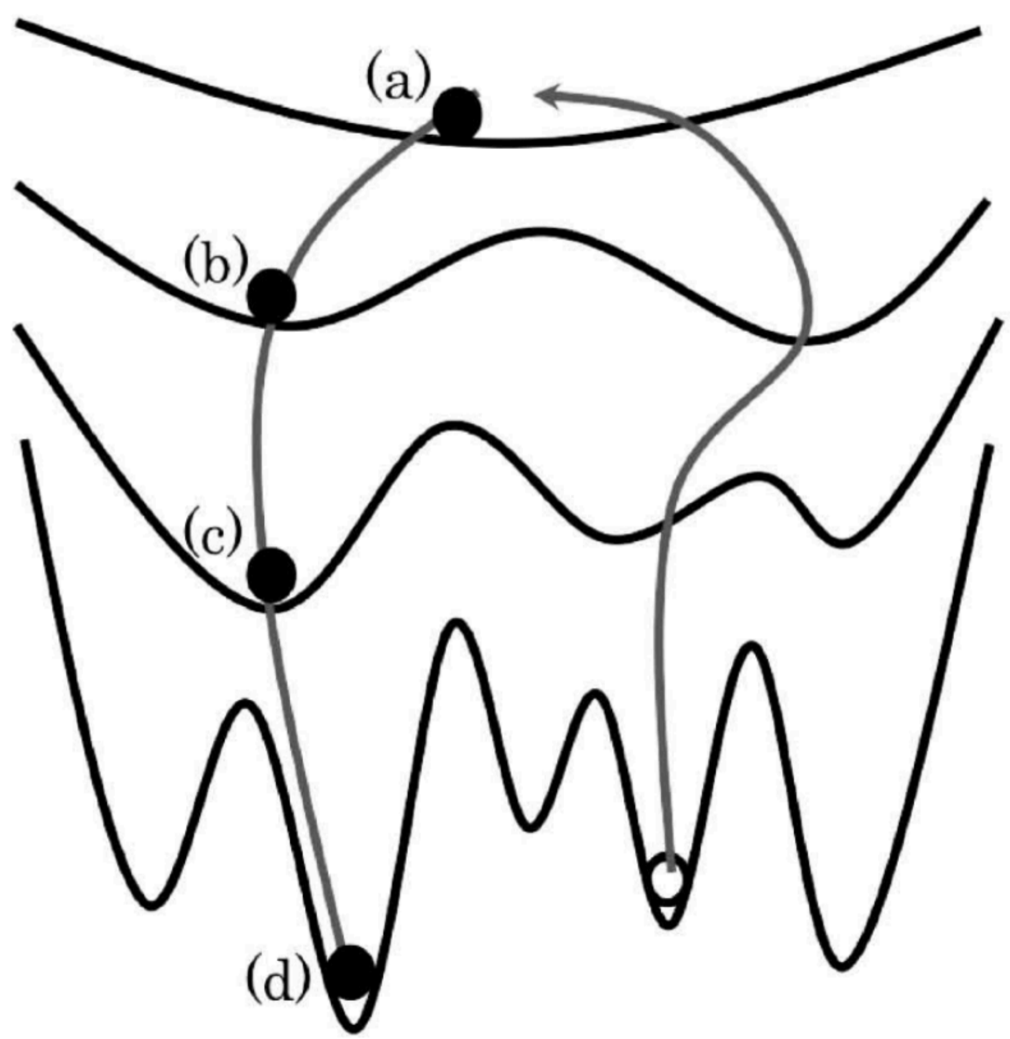
$$\mathrm{P}(a_{11}, a_{12}, a_{21}, \cdots, a_{2d\,2d}) = \frac{1}{Z}\exp\left[-\frac{N}{2\sigma^2}E_{\mathrm{samp}}(a_{11}, a_{12}, a_{21}, \cdots, a_{2d\,2d})\right]$$

- Replica exchange Monte Carlo method [Hukushima and Nemoto 1995]

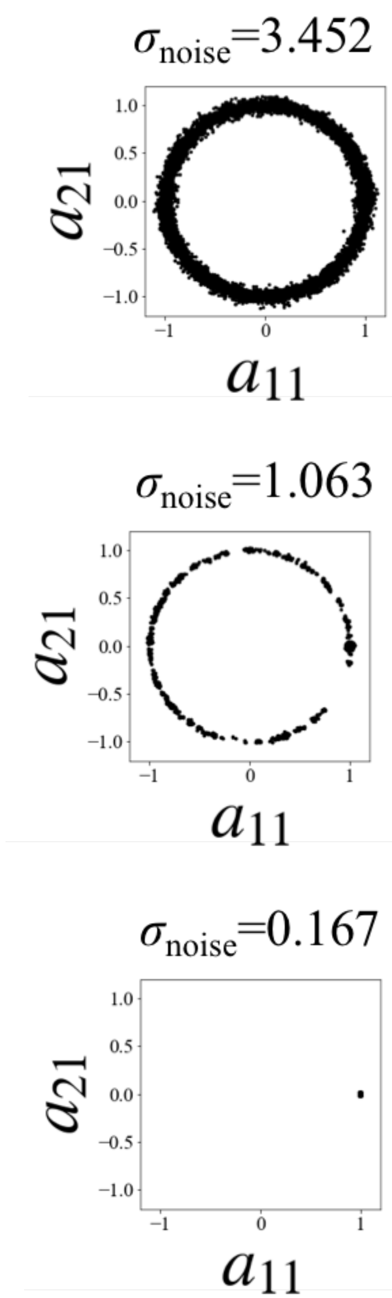
$$\mathrm{P}(A^l) = \frac{1}{Z}\exp\left[-\beta\frac{N}{2\sigma^2}E_{\mathrm{samp}}(A^l)\right]$$

MCMC sampling
from the following simultaneous distributions

$$\mathrm{P}(A^1, A^2 \cdots A^L) = \prod_{l-1}^{L}\mathrm{P}(A^l)$$



[Nagata and Okada 15]



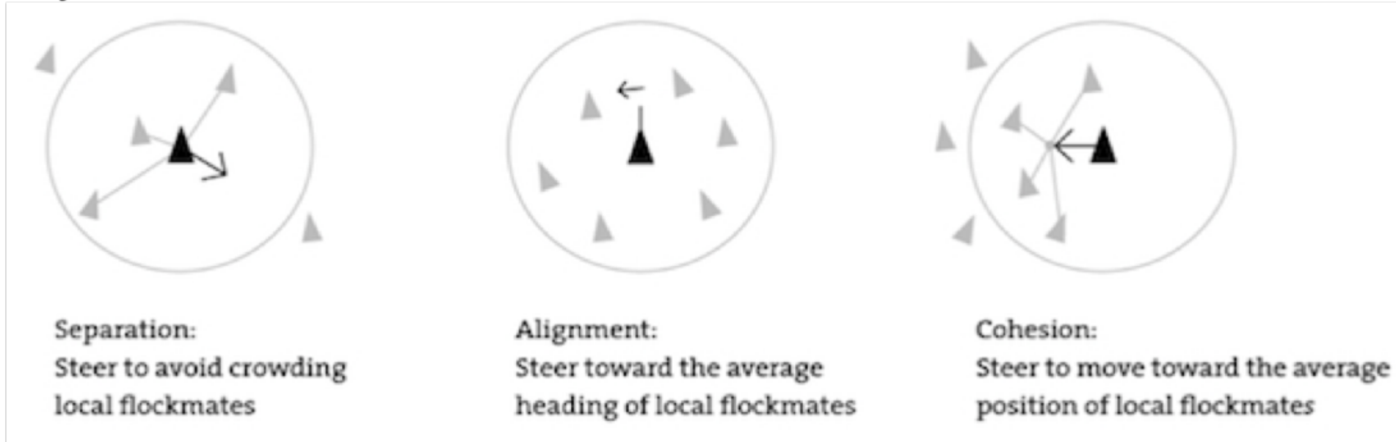
[Step3]

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

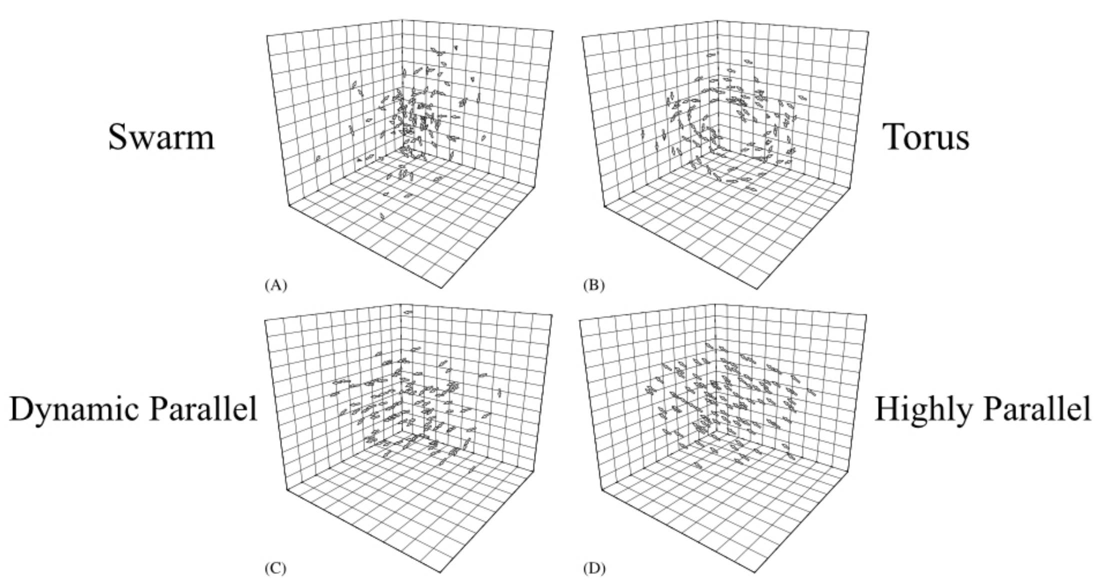
Result ：Collective motion

Model

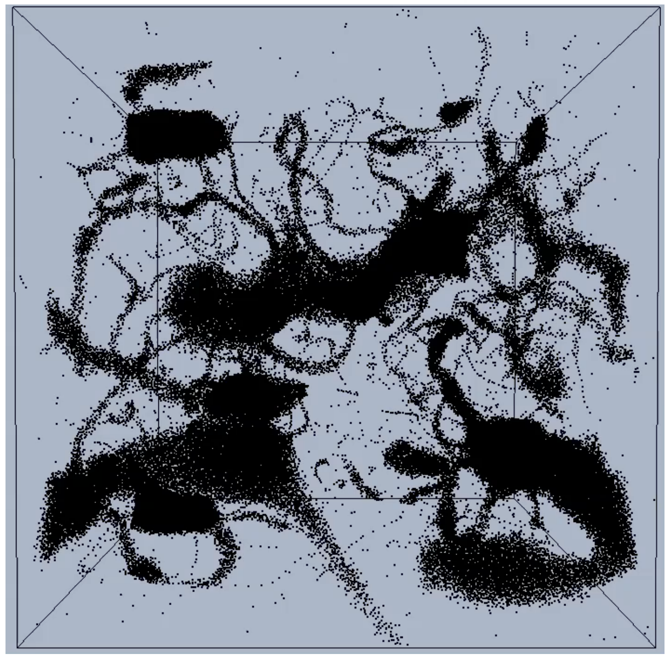
Reynolds model [Reynolds et al. 86]



$$\frac{dv}{dt} = \vec{F}_{separation} + \vec{F}_{alignment} + \vec{F}_{cohesion}$$



[Couzin et.al 2002]



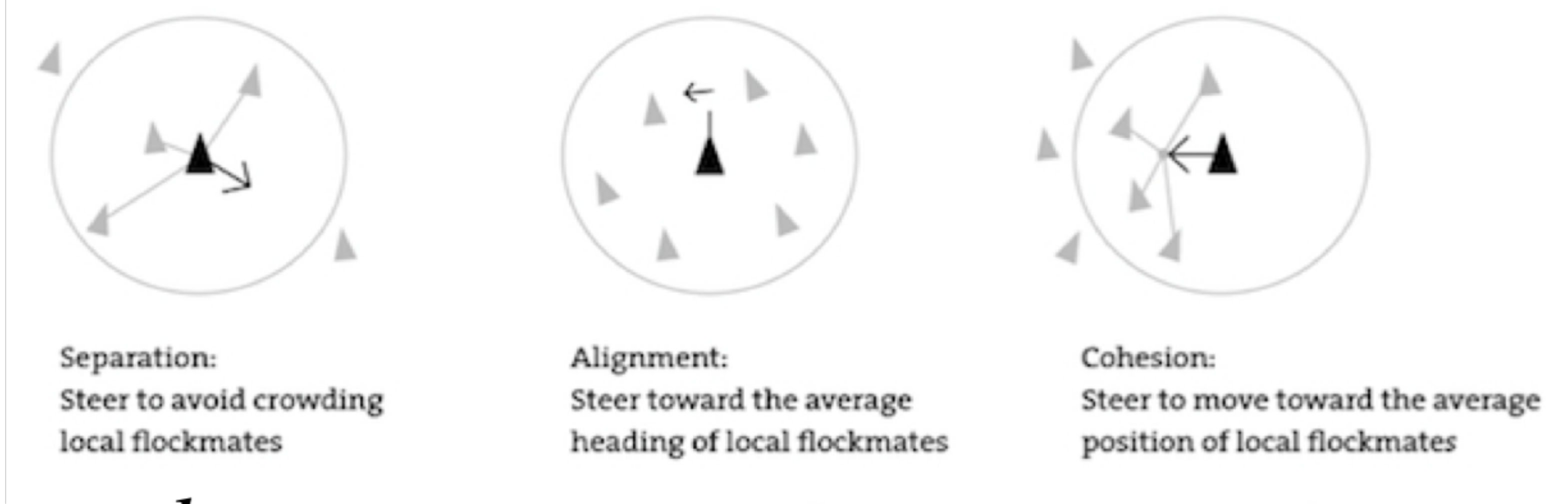
[Mototake et.al 2015]

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Result：Collective motion

Model

Reynolds model [Reynolds et al. 86]



$$\frac{dv}{dt} = \vec{F}_{separation} + \vec{F}_{alignment} + \vec{F}_{cohesion}$$

Coordinate

$$D = \{\tilde{\boldsymbol{q}}(t_i)_i, \tilde{\boldsymbol{p}}(t_i)_i, \tilde{\boldsymbol{q}}(t_i + \Delta t)_i, \tilde{\boldsymbol{p}}(t_i + \Delta t)_i\}_{i=1}^{N_R T}$$

$$(\tilde{\boldsymbol{q}}, \tilde{\boldsymbol{p}}) = (q_1 - \bar{q}_1, q_2 - \bar{q}_2, p_1 - \bar{p}_1, p_2 - \bar{p}_2)$$

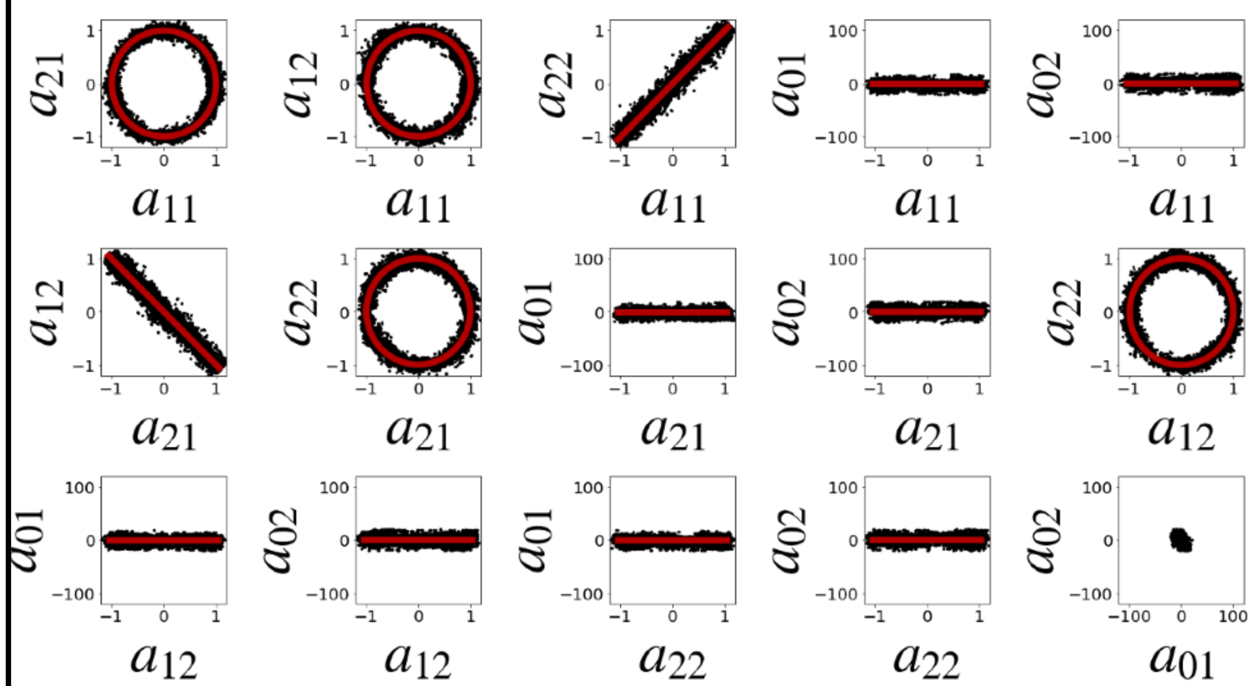
Candidate transformation

$$\begin{pmatrix} a_{11} & 1.0\,a_{21} & 0 & 0 \\ 1.0\,a_{12} & a_{22} & 0 & 0 \\ 0 & 0 & a_{22} & -1.0\,a_{12} \\ 0 & 0 & -1.0\,a_{21} & a_{11} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{pmatrix} + \begin{pmatrix} a_{01} \\ a_{02} \\ 0.167\,a_{02} \\ -0.161\,a_{01} \end{pmatrix}$$

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

Result：Collective motion

Extracted symmetries



Implicit functions

$$
\begin{cases}
1.01\ a_{11}^2 + 1.01\ a_{21}^2 - 0.01\ a_{11} + 0.00\ a_{21} + 0.00\ a_{11}a_{21} = 1.00 \\
1.01\ a_{11}^2 + 1.00\ a_{12}^2 - 0.01\ a_{11} = 0.99 \\
1.00\ a_{11} - 1.00\ a_{22} + 0.00\ a_{22} = 0 \\
1.00\ a_{21} - 1.00\ a_{12} = 0 \\
1.00\ a_{21}^2 + 1.00\ a_{22}^2 - 0.01\ a_{22} = 0.99 \\
1.02\ a_{12}^2 + 1.00\ a_{22}^2 - 0.01\ a_{12}a_{22} = 1.00
\end{cases}
$$

Infinitesimal transformation

$$
\delta \boldsymbol{q} = \varepsilon \begin{pmatrix} \frac{\partial a_{11}}{\partial a_{21}} & \frac{\partial a_{21}}{\partial a_{21}} \\ \frac{\partial a_{12}}{\partial a_{21}} & \frac{\partial a_{21}}{\partial a_{21}} \end{pmatrix} \boldsymbol{q} + \varepsilon \begin{pmatrix} \frac{\partial a_{01}}{\partial a_{21}} \\ \frac{\partial a_{02}}{\partial a_{21}} \end{pmatrix} = \varepsilon \begin{pmatrix} \left.\frac{-1.01 \times 2a_{21}}{1.01 \times 2a_{11} - 0.01}\right|_{A'=e_{\mathrm{I}}} & 1 \\ \left.\frac{-1.00}{1.00}\right|_{A'=e_{\mathrm{I}}} & \left.\frac{-1.00 \times 2a_{21}}{1.00 \times 2a_{22} + 0.01}\right|_{A'=e_{\mathrm{I}}} \end{pmatrix} \boldsymbol{q} = \begin{pmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{pmatrix} \boldsymbol{q}
$$

$$
\delta \boldsymbol{p} = \varepsilon \begin{pmatrix} \frac{\partial a_{11}}{\partial a_{21}} & \frac{\partial a_{21}}{\partial a_{21}} \\ \frac{\partial a_{12}}{\partial a_{21}} & \frac{\partial a_{21}}{\partial a_{21}} \end{pmatrix} \boldsymbol{p} + \varepsilon \begin{pmatrix} \frac{\partial a_{01}}{\partial a_{21}} \\ \frac{\partial a_{02}}{\partial a_{21}} \end{pmatrix} = \begin{pmatrix} 0 & \varepsilon \\ -\varepsilon & 0 \end{pmatrix} \boldsymbol{p}
$$

Consistent with results suggested in a previous study ➡

[Couzin et.al 2002]

Conservation law

$$G_\delta = \varepsilon(x_1 p_2 - x_2 p_1)$$

[Y. Mototake, Phys. Rev. E, 103, 033303, 2021]

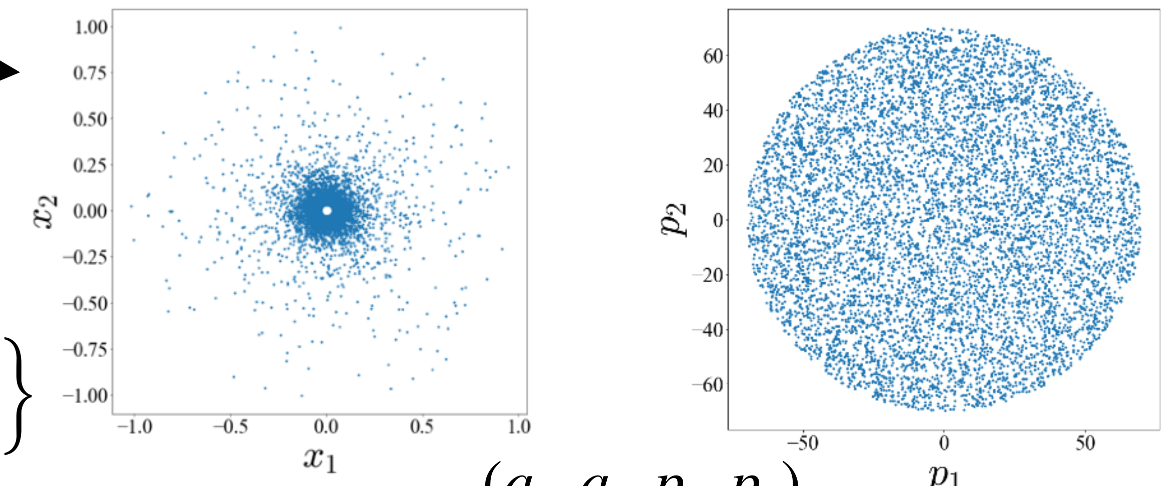
[Y. Mototake, 2023]

The proposed method can be applied to **nonlinear transformations.**

$$H = \frac{\mathbf{p}^2}{2m} + \frac{k}{|\mathbf{r}|}, k : \text{const}.$$

$$\mathbf{q} = (q_1, q_2), \mathbf{p} = (p_1, p_2)$$

$$S_i := \left\{ \boldsymbol{q}_{t+\Delta t}, \boldsymbol{p}_{t+\Delta t}, \boldsymbol{q}_t, \boldsymbol{p}_t \;\middle|\; H(\boldsymbol{q}_t, \boldsymbol{p}_t) = E_i, \boldsymbol{p}_{t+\Delta t} = \boldsymbol{p}_t - \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{q}_t}, \boldsymbol{q}_{t+\Delta t} = \boldsymbol{q}_t + \frac{\partial H(\boldsymbol{q}_t, \boldsymbol{p}_t)}{\partial \boldsymbol{p}_t} \right\}$$



(q_1, q_2, p_1, p_2)

$$(q_1, q_2, q_3, p_1, p_2, p_3) \rightarrow (\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{q}_4, \tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4)$$

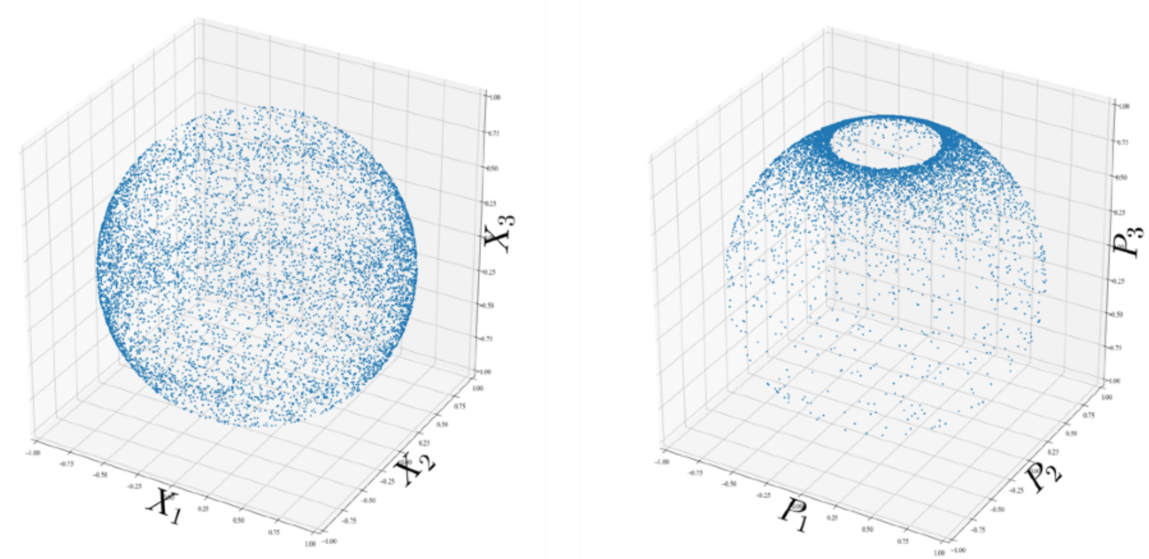
$$\tilde{\boldsymbol{q}} = \tilde{\boldsymbol{q}}(\boldsymbol{q}, \boldsymbol{p}) := \frac{\boldsymbol{q}}{\|\boldsymbol{q}\|_2} - \frac{\boldsymbol{q} \cdot \boldsymbol{p}}{mG}\boldsymbol{p}, \quad \tilde{q}_4 = \tilde{q}_4(\boldsymbol{q}, \boldsymbol{p}) := \frac{p_0}{mG}\boldsymbol{q} \cdot \boldsymbol{p},$$

$$\tilde{\boldsymbol{p}} = \tilde{\boldsymbol{p}}(\boldsymbol{q}, \boldsymbol{p}) := \frac{2p_0\boldsymbol{p}}{p_0^2 + p^2}, \quad \tilde{p}_4 = \tilde{p}_4(\boldsymbol{q}, \boldsymbol{p}) := \frac{p^2 - p_0^2}{p_0^2 + p^2},$$

[H. H. Rogers, J. Math. Phys. 14, 1125 (1973)]

Transform

$(\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{p}_1, \tilde{p}_2, \tilde{p}_3)$



Runge-Lenz vector

$$\vec{A} = \boldsymbol{p} \times L - mG\frac{\boldsymbol{q}}{\|\boldsymbol{q}\|_2},$$

$$L = \boldsymbol{q} \times \boldsymbol{p},$$

▶There is a possibility that the proposed method can find **hidden** symmetry or conservation laws.

Summary and discussion

- We showed that candidates for the symmetry of Hamiltonian systems can be estimated from time series data manifolds.
- It is suggested that the symmetry of the time series data manifold can be estimated from the trained DNN.
- It was suggested that conservation laws can be estimated from the set of symmetry transformations obtained.
- As the number of conservation laws increases, the possibility that the Jacobi matrix is no longer regular increases, and the estimation becomes more difficult.
- As the dimensionality of the phase space increases, difficulties will arise in extracting symmetry using the sampling method.
- In principle, the method can be applied to nonlinear transformations.