A Spectral-Grassmann Wasserstein metric for operator representations of dynamical systems

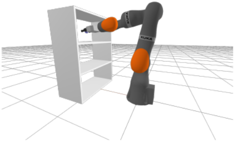
Thibaut Germain, **Rémi Flamary**, Vladimir R. Kostic, Karim Lounici

CMAP, École Polytechnique, Institut Polytechnique de Paris

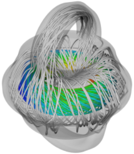
December 17th 2025, École normale supérieure de Lyon

Table of content

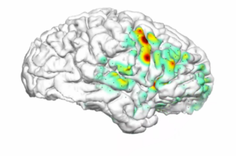
Dynamical system & machine learning



Robotic



Fluid dynamic



Brain dynamic

Dynamical systems

- Dynamical systems are backbone models of temporally evolving phenomena.
- Continuous time: $\frac{dx(t)}{dt} = g(t, x(t))$
- Discrete time: $x_{t+1} = g(t, x_t)$

Machine learning for dynamical systems

- Classical approach: ODE/PDE/SDE design + parameter fittting
- Data-driven approach: learn dynamics from data (with physics-informed constraints)

Learning dynamical systems with transfer operators

Definition: Transfer Operator

Let us assume:

- A stochastic process: $(X_t)_{t \geq 0} \in \mathcal{X}$
- A real-valued functional space: $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$

Under some assumptions for $t \geq 0$, their exists a linear *transfer operator*, also known as *Koopman operator*, $A_t \colon \mathcal{F} \to \mathcal{F}$ that evolves an observable $f : \mathcal{X} \to \mathbb{R}$ for time t via the conditional expectation :

$$[A_t(f)](x) := E[f(X_t) \,|X_0 = x], \; x \in \mathcal{X}, f \in \mathcal{F}. \tag{1}$$

Remarks

- Even if the dynamical system is non-linear, the transfert operator is linear.
- Time-homogeneous systems : $A_{t+s} = A_t A_s$
- Continuous time: $A_t = \exp(tL)$ with L the infinitesimal generator of the semigroup $(A_t)_{t \geq 0}$.
- Discrete time: $A = A_1 = L$ is enough to describe the dynamics with $A_t = A_1^t$.

Spectral decomposition
Assuming that $\mathcal{F}$ is a separable Hilbert space (typically $L^2_\pi(\mathcal{X})$ with π the system's invariante measure) and L is a non defective operator with purely discrete spectrum. Then L can be written as:

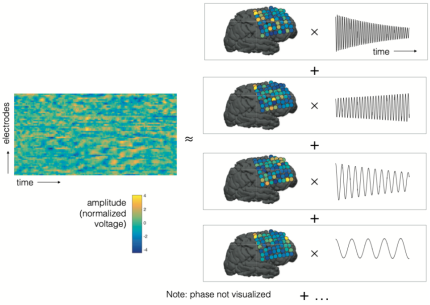
$$L = \sum_{j=1}^{\infty} \lambda_j g_j \otimes f_i, \text{ with } Lf_j{=}\lambda_j f_j,\ L^* g_j{=}\overline{\lambda_j} g_j \text{ , and } \langle f_j, g_j\rangle_{\mathcal{F}}{=}\delta_{i,j}$$

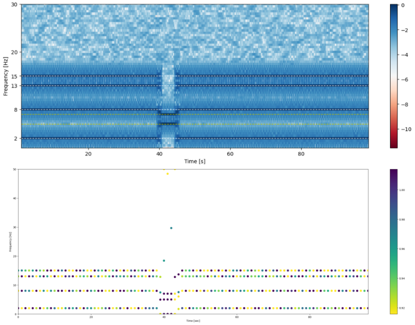
Properties :

- Fast computation : $A_t = \exp(tL) = \sum_{j=1}^{\infty} \exp(t\lambda_j) g_j \otimes f_i$
- Can be used to model evolution of densities of probability distributions.

Matrix view for operators

$$L = \underbrace{\begin{pmatrix} | & & | \\ f_1 & \cdots & f_r \\ | & & | \end{pmatrix}}_{F} \underbrace{\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{pmatrix}}_{\Lambda} \underbrace{\begin{pmatrix} - & g_1 & - \\ & \vdots & \\ - & g_r & - \end{pmatrix}}_{H^*} \quad \text{with} \quad H^* F = I_r$$





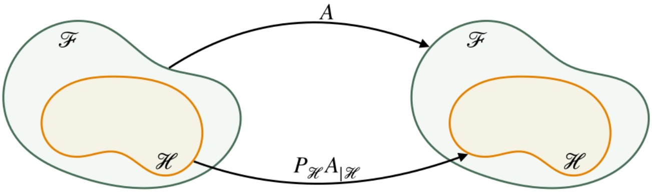
Spectral Decomposition

The spectral decomposition of $A_t = \exp(tL)$ decouples the evolution of any arbitrary observable $f \in \mathcal{F}$ as:

$$[A_t f](x) = E[f(X_t)|X_0 = x] = \sum_{j\in\mathbb{N}} e^{\lambda_j t}\langle f_j, g_j\rangle_{\mathcal{F}} f_j(x) = \sum_{j\in\mathbb{N}} e^{\tau_j t} e^{i2\pi\omega_j t} m_j^f(x),$$

into temporal and static components.

- Temporal: $e^{\tau_j t}$ (decay/growth) and $e^{i2\pi\omega_j t}$ (oscillation)
- Static: $m_j^f(x) = \langle f_j, g_j\rangle_{\mathcal{F}} f_j(x)$



Leraning operators from data

- Only trajectories of dynamical systems are observed. Neither the operator A nor its domain F are known.
- Learn the operator L from data in a RKHS $\mathcal{H} \subseteq \mathcal{F}$ with kernel $k(x, y) = \langle \phi(x), \phi(y) \rangle$ and estimate a projected operator.
- Given data $\{x_i, y_i\}_{i=1}^N$ (typically a trajectory with $y_i = x_{i+1}$) estimate $\hat{L}$ minimizing the empirical risk:

$$\min_{G \in HS(\mathcal{X})} \frac{1}{N} \sum_{i=1}^{N} \|\phi(y_i) - G\phi(x_i)\|_{\mathcal{H}}^2$$

Transfer operators in practice

Classical approaches

- Dynamic Mode Decomposition (DMD) [Kutz et al., 2016, Brunton et al., 2022] (Linear kernel but only for $f = Id$).
- Koopman operators with kernel methods [Williams et al., 2014, Kawahara, 2016].
- Reduced rank operator estimation [Kostic et al., 2022].
- Neural network approaches [Lusch et al., 2018, Kostic et al., 2024].

Open questions

- How to compare transfer operators ?
- Existing approaches:
  - Hilbert-Schmidt and operator norms are too conservative.
  - Martin distance [Martin, 2002]: pseudo-metric on ARMA models
  - Binet-Cauchy kernel [Chaudhry and Vidal, 2013]: Martin distance extension to LDS
  - Optimal Transport on spectrum (SOT)[Redman et al., 2024].
  - Optimal transport on eigenspaces (GOT) [Antonini and Cavalletti, 2021].
- $\rightarrow$ Propose a novel geometry for transfer operators based on optimal transport.

Spectral-Grassmann Wasserstein Metric (SGOT)

SGOT intuition & assumptions

Assumptions

1. **Time homogeneous Markovian dynamical systems** $\{L_k\}_{k\in[N]}$ (stationary).
2. **Low rank operators**: For all k, L_k has rank $r \ll N$.
3. **Common functional space** $\mathcal{H}$ for all operators. There exists an RKHS $\mathcal{H}$ such that for any operator, the estimation of its r-restriction, $Tk = exp(L_{k|r})$, is well defined.

Low rank spectral decomposition

$$L = \sum_{i\in[\ell]} \sum_{j\in[m_i]} \lambda_i g_{i,j} \otimes f_{i,j} = \sum_i \lambda_i P_i, \quad \langle f_{i,j}, g_{i',j'} \rangle_{\mathcal{H}} = \delta_{i,i'}\delta_{j,j'}, \quad \sum_i m_i = r$$

The representatin above is unique up to :

- Permutation of the indexes of the decomposition i.
- Change of basis of each spectral projectors P_i (Grassmann manifold).

$\rightarrow$ We need a metric invariant to these transformations : OT with proper geometry.

Low rank spectral decomposition

$$L = \sum_{i \in [\ell]} \sum_{j \in [m_i]} \lambda_i g_{i,j} \otimes f_{i,j} = \sum_i \lambda_i P_i, \quad \langle f_{i,j}, g_{i',j'} \rangle_{\mathcal{H}} = \delta_{i,i'} \delta_{j,j'}, \quad \sum_i m_i = r$$

- m_i is the multiplicity of eigenvalue λ_i.
- $P_i = \sum_{j \in [m_i]} g_{i,j} \otimes f_{i,j}$ is the spectral projector associated to λ_i.
- $\mathcal{V}_i = \text{span}\{g_{i,j} \otimes f_{i,j}\}_{j \in [m_i]}$ is the subspace of $\mathcal{H}$ associated to λ_i.

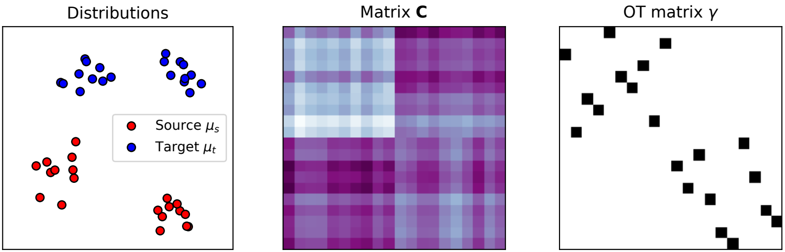
Property : distributional embedding

The operator L can be represented as a probability distribution over the product space of eigenvalues and projectors:

$$\mu(L) = \sum_{i \in [\ell]} \frac{m_i}{r} \delta_{(\lambda_i, \mathcal{V}_i)}$$

The embedding above is injective. For fixed rank r those distributions can be compared with discrete Optimal Transport

Optimal transport with discrete distributions



OT Linear Program and Wasserstein distance

When $\mu_s = \sum_{i=1}^{n} a_i \delta_{\mathbf{x}_i^s}$ and $\mu_t = \sum_{i=1}^{n} b_i \delta_{\mathbf{x}_i^t}$

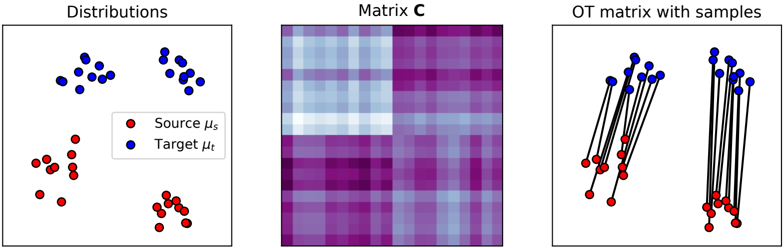
$$W_{d,p}^p(\mu_s, \mu_t) = \min_{\mathbf{T} \in \Pi(\mu_s, \mu_t)} \left\{ \langle \mathbf{T}, \mathbf{D} \rangle_F = \sum_{i,j} T_{i,j} c_{i,j}^p \right\}$$

where $\mathbf{D}$ is a distance matrix with $d_{i,j} = d(\mathbf{x}_i^s, \mathbf{x}_j^t)$ and the marginals constraints are

$$\Pi(\mu_s, \mu_t) = \left\{ \mathbf{T} \in (\mathbb{R}^+)^{n_s \times n_t} | \, \mathbf{T}\mathbf{1}_{n_t} = \mathbf{a}, \mathbf{T}^T \mathbf{1}_{n_s} = \mathbf{b} \right\}$$

Linear program with $n_s n_t$ variables and $n_s + n_t$ constraints, can be solved with complexity $O(r^3 \log r)$ if $n_s = n_t = r$.

Optimal transport with discrete distributions



OT Linear Program and Wasserstein distance

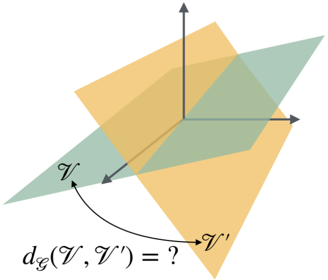
When $\mu_s = \sum_{i=1}^{n} a_i \delta_{\mathbf{x}_i^s}$ and $\mu_t = \sum_{i=1}^{n} b_i \delta_{\mathbf{x}_i^t}$

$$W_{d,p}^p(\mu_s, \mu_t) = \min_{\mathbf{T} \in \Pi(\mu_s, \mu_t)} \left\{ \langle \mathbf{T}, \mathbf{D} \rangle_F = \sum_{i,j} T_{i,j} c_{i,j}^p \right\}$$

where $\mathbf{D}$ is a distance matrix with $d_{i,j} = d(\mathbf{x}_i^s, \mathbf{x}_j^t)$ and the marginals constraints are

$$\Pi(\mu_s, \mu_t) = \left\{ \mathbf{T} \in (\mathbb{R}^+)^{n_s \times n_t} | \, \mathbf{T}\mathbf{1}_{n_t} = \mathbf{a}, \mathbf{T}^T \mathbf{1}_{n_s} = \mathbf{b} \right\}$$

Linear program with $n_s n_t$ variables and $n_s + n_t$ constraints, can be solved with complexity $O(r^3 \log r)$ if $n_s = n_t = r$.



Proposition : Extension to infinite dimension space

Let $\mathcal{H}$ be a separable Hilbert space, and $\mathcal{G}_r(\mathcal{H})$ the set of vector subspaces of with dimension at most r. With the application:

$$d_{\mathcal{G}}(\mathcal{V}, \mathcal{V}') = \|P_{\mathcal{V}} - P_{\mathcal{V}'}\|_{\mathcal{H}}, \quad \forall \mathcal{V}, \mathcal{V}' \in \mathcal{G}_r(\mathcal{H})^2$$

where $P_{\mathcal{V}}$ is the orthogonal projector onto $\mathcal{V}$, then $(\mathcal{G}_r(\mathcal{H}), d_{\mathcal{G}})$ is a metric space.

Example for 1D subspaces:

$d_{\mathcal{G}}(\mathcal{V}, \mathcal{V}') = \sqrt{2 - 2\langle f, f' \rangle \langle g, g' \rangle}$ with f, f', g, g' normalized.

SGOT definition

Spectral-Grassmann Optimal Transport (SGOT)

Let H be a separable $\mathbb{C}$-Hilbert space and $\mathcal{S}_r(\mathcal{H})$ the set of non-defective operator on $\mathcal{H}$ with rank at most r. For $p \geq 1$ and $\eta \in [0,1]$, we define the Spectral-Grassmann Wasserstein metric distance between two operators $L, L' \in \mathcal{S}_r(\mathcal{H})$ as:

$$d^p_{SGOT,p}(L, L') = W^p_{c_\eta,p}(\mu(L'), \mu(L'))$$

where the cost matrix is defined as:

$$c_\eta((\lambda, \mathcal{V}), (\lambda', \mathcal{V}')) = \eta|\lambda^s_i - \lambda^t_j| + (1-\eta)d_{\mathcal{G}}(\mathcal{V}, \mathcal{V}')$$

Then the space $(\mathcal{S}_r(\mathcal{H}), d_{SGOT,p})$ is a metric space.

Computation

- Pre-compute the cost matrix $\mathbf{C}$ with complexity $O(n^2r^2)$.
- Solve the OT problem with complexity $O(r^3 \log r)$ with network simplex.
- Overall complexity: $O(n^2r^2 + r^3 log(r))$.

Existing bounds for operator estimation

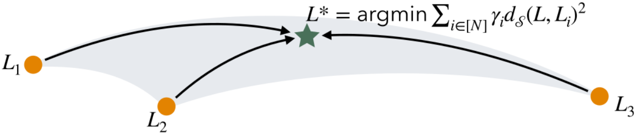
Under kernel universality assumption, estimation of Koopman operators come with spectral estimation guarantees [Kostic et al., 2023], specifically for the Reduced Rank Regression (RRR) method [Kostic et al., 2022].

Theorem (simplified)

Suppose two dynamical systems with low rank operator projection $L_1, L_2 \in \mathcal{S}_r(\mathcal{H})$ and their estimations $\hat{L}_1, \hat{L}_2 \in \mathcal{S}_r(\mathcal{H})$ from n samples with the RRR method [Kostic et al., 2022]. Suppose $\alpha \in (1, 2)$, and $\beta \in [0, 1]$ bounding the empirical covariance. In the i.i.d setting, for any $\delta \in (0, 1)$, with probability $1 - \delta$ it holds:

$$|d_{SGOT,p}(\hat{L}_1, \hat{L}_2) - d_{SGOT,p}(L_1, L_2)| \lesssim n^{-\frac{\alpha-1}{2(\alpha+\beta)}} \ln(2\delta^{-1})$$

Proof sketch: Use convergence of individual spectral elements [Kostic et al., 2023] and use the identity OT plan for upper bound.



SGOT Frechet mean

$$\underset{L\in\mathcal{S}_r(\mathcal{H})}{\text{argmin}} \sum_{k\in[N]} \gamma_i d_{SGOT}(L, L_k)^2, \tag{2}$$

where $\gamma \in \Sigma_N$ is a weight vector.

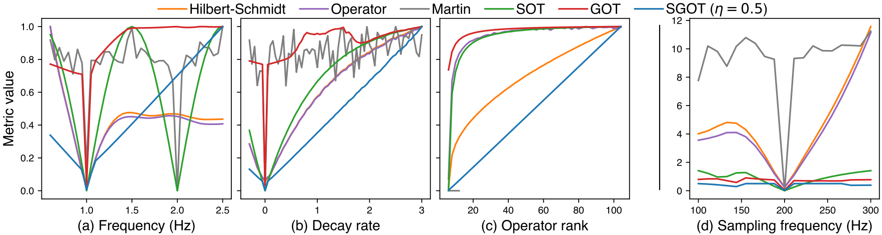
Numerical optimization of SGOT barycenter
We consider in the kernel case the operator $L_{\boldsymbol{\theta}}$ representation parametrized by $\boldsymbol{\theta} \triangleq (\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{x})$:

$$L_{\boldsymbol{\theta}} : h \in \mathcal{H} \mapsto \sum_{i\in[r]} \lambda_i \langle \kappa_{\mathbf{x}} \boldsymbol{\alpha}_i, h\rangle_{\mathcal{H}} \kappa_{\mathbf{x}} \boldsymbol{\beta}_i \in \mathcal{H}$$

The barycenter is optimized by optimizing:

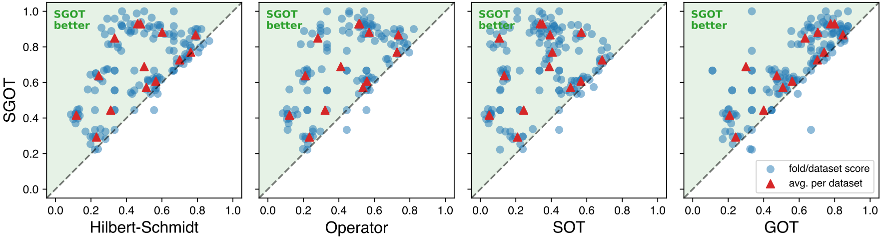
$$\underset{\boldsymbol{\theta},\mathbf{P}}{\text{argmin}} \sum_{i\in[N]} \gamma_i \langle \mathbf{C}_i(\boldsymbol{\theta}), \mathbf{P}_i\rangle_F \text{ s.t. } \begin{cases} \boldsymbol{\alpha}^* \mathbf{K} \boldsymbol{\beta} = \mathbf{I} & \mathbf{K} = \{\kappa(x_i, x_j)\}_{(i,j)\in[n]^2} \\ \boldsymbol{\beta}_j^* \mathbf{K} \boldsymbol{\beta}_j = 1, \ \forall j \in [r] & \mathbf{P}_i \in \Pi(\mu(L_{\boldsymbol{\theta}}), \mu(L_i)), \ \forall i \end{cases}$$

Numerical experiments



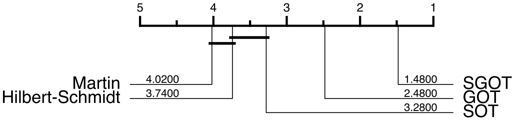
Numerical setup

- Compare SGOT with operator, Hislbert-Schmidt Martin, SOT and GOT distances.
- Simple 1D oscillatory dynamical systems with varying frequency, damping, rank and sampling frequency.
- SGOT has a unique minimum at the true parameters and capture well the variations of the systems.
- SGOT is robust (invariant) to sampling frequency changes.



Numerical setup

- 14 multivariate time series datasets from the UEA database [Ruiz et al., 2021].
- Each time series is modeled as a dynamical system (linear, kernel and MLP) and its transfer operator is estimated with RRR [Kostic et al., 2022].
- 1-NN classification with SGOT, GOT, SOT, Martin and Hilbert-Schmidt.
- SGOT outperforms other operator distances on most datasets.
- SGOT computation time is comparable to best existing methods and much faster than Hilbert-Schmidt and Operator norms.

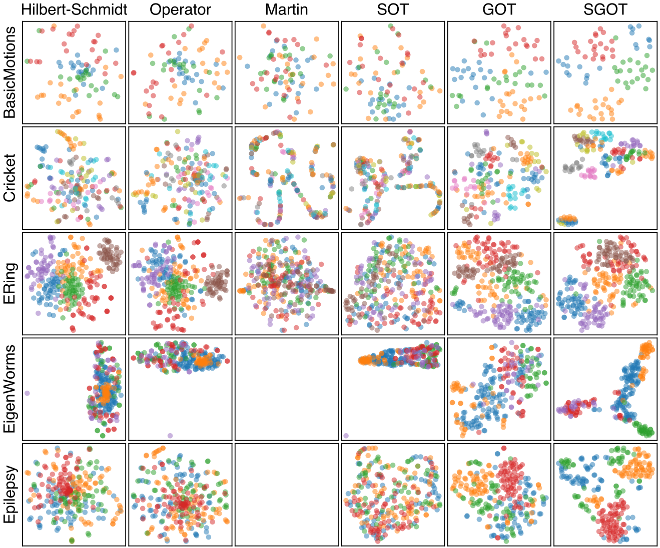


Hilbert-Schmidt	Operator	Martin	SOT	GOT	SGOT
4.96ms	13.04ms	0.02ms	0.03ms	0.14ms	0.12ms

Numerical setup

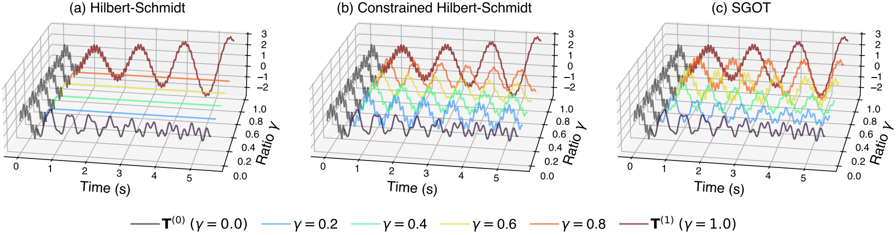
- 14 multivariate time series datasets from the UEA database [Ruiz et al., 2021].
- Each time series is modeled as a dynamical system (linear, kernel and MLP) and its transfer operator is estimated with RRR [Kostic et al., 2022].
- 1-NN classification with SGOT, GOT, SOT, Martin and Hilbert-Schmidt.
- SGOT outperforms other operator distances on most datasets.
- SGOT computation time is comparable to best existing methods and much faster than Hilbert-Schmidt and Operator norms.

TSNE Embeddings of time series



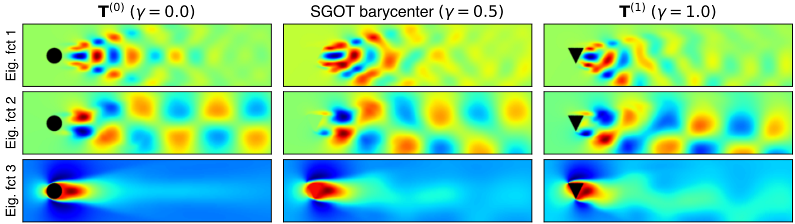
TNSE embeddings of time series using SGOT and other metrics.

Interpolation for 1D system



Numerical setup

- Consider two 1D oscillatory dynamical systems with different frequencies and dampings.
- Estimate their transfer operators and compute their SGOT, HS and constrained (low rank on manifold) HS barycenters for varying weights.
- Simulate the barycenter dynamical system, starting form the same initialization.
- SGOT barycenters interpolate well between the two systems while HS barycenters fail to capture the change in dynamics.

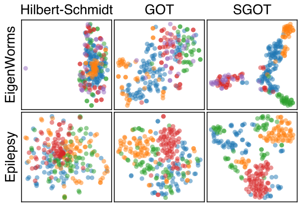


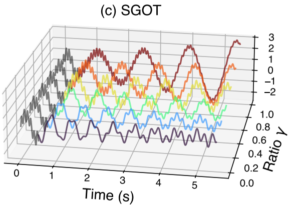
Numerical setup

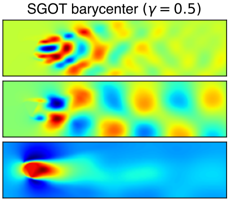
- Consider two fluid dynamics simulations with different shapes: circular with symmetry and triangular without symmetry.
- Estimate their transfer operators and compute their SGOT barycenter.
- Recovered eigen functions and dynamics interpolate well between the two systems.

Conclusion

Conclusion







Contributions

- A novel geometry for transfer operators based on optimal transport on the joint spectral and Grassmann manifolds.
- Statistical guarantees for the SGOT metric estimation from data.
- Numerical experiments showing the interest of SGOT for time series classification and barycenter computation.

Future works

- Comparison of baryenter with physics-based interpolation methods.
- Application on simulated nuclear fusion data (Tokam2D simulator).

Thank you

Doc : https://pythonot.github.io/

Code : https://github.com/PythonOT/POT

- OT LP solver, Sinkhorn (stabilized, GPU)
- Sliced OT, OT on sphere, Gaussian and Gaussian Mixture OT.
- Gromov-Wasserstein, Unbalanced.
- Barycenters, Wasserstein unmixing.
- Differentiable solvers for Numpy/Pytorch/tensorflow/Cupy

Course on OT for ML:
https://tinyurl.com/otml-course

Papers available on my website:
https://remi.flamary.com/

Antonini, P. and Cavalletti, F. (2021).
Geometry of grassmannians and optimal transport of quantum states.
arXiv preprint arXiv:2104.02616.

Brunton, S. L., Budisić, M., Kaiser, E., and Kutz, J. N. (2022).
Modern Koopman theory for dynamical systems.
SIAM Review, 64(2):229–340.

Chaudhry, R. and Vidal, R. (2013).
Initial-state invariant binet-cauchy kernels for the comparison of linear dynamical systems.
In *52nd IEEE Conference on Decision and Control*, pages 5377–5384. IEEE.

Kawahara, Y. (2016).
Dynamic mode decomposition with reproducing kernels for koopman spectral analysis.
Advances in neural information processing systems, 29.

Kostic, V., Lounici, K., Novelli, P., and Pontil, M. (2023).
Sharp spectral rates for koopman operator learning.
Advances in Neural Information Processing Systems, 36:32328–32339.

Kostic, V., Novelli, P., Maurer, A., Ciliberto, C., Rosasco, L., and Pontil, M. (2022).
Learning dynamical systems via koopman operator regression in reproducing kernel hilbert spaces.
Advances in Neural Information Processing Systems, 35:4017–4031.

Kostic, V. R., Novelli, P., Grazzi, R., Lounici, K., and Pontil, M. (2024).
Learning invariant representations of time-homogeneous stochastic dynamical systems.
In *International Conference on Learning Representations*.

Kutz, J. N., Brunton, S. L., Brunton, B. W., and Proctor, J. L. (2016).
Dynamic mode decomposition: data-driven modeling of complex systems.
SIAM.

Lusch, B., Kutz, J. N., and Brunton, S. L. (2018).
Deep learning for universal linear embeddings of nonlinear dynamics.
Nature communications, 9(1):4950.

Martin, R. J. (2002).
A metric for arma processes.
IEEE transactions on Signal Processing, 48(4):1164–1170.

Redman, W., Bello-Rivas, J., Fonoberova, M., Mohr, R., Kevrekidis, Y., and Mezic, I. (2024).
Identifying equivalent training dynamics.
Advances in Neural Information Processing Systems, 37:23603–23629.

Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M., and Bagnall, A. (2021).
The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances.
Data mining and knowledge discovery, 35(2):401–449.

Williams, M. O., Rowley, C. W., and Kevrekidis, I. G. (2014).
A kernel-based approach to data-driven koopman spectral analysis.
arXiv preprint arXiv:1411.2260.