Analytical Koopman approach to recurrent neural networks

Benjamin Claveau[1,2], Vincent Thibeault[1,2], Antoine Allard[1,2], Patrick Desrosiers[1,2,3]

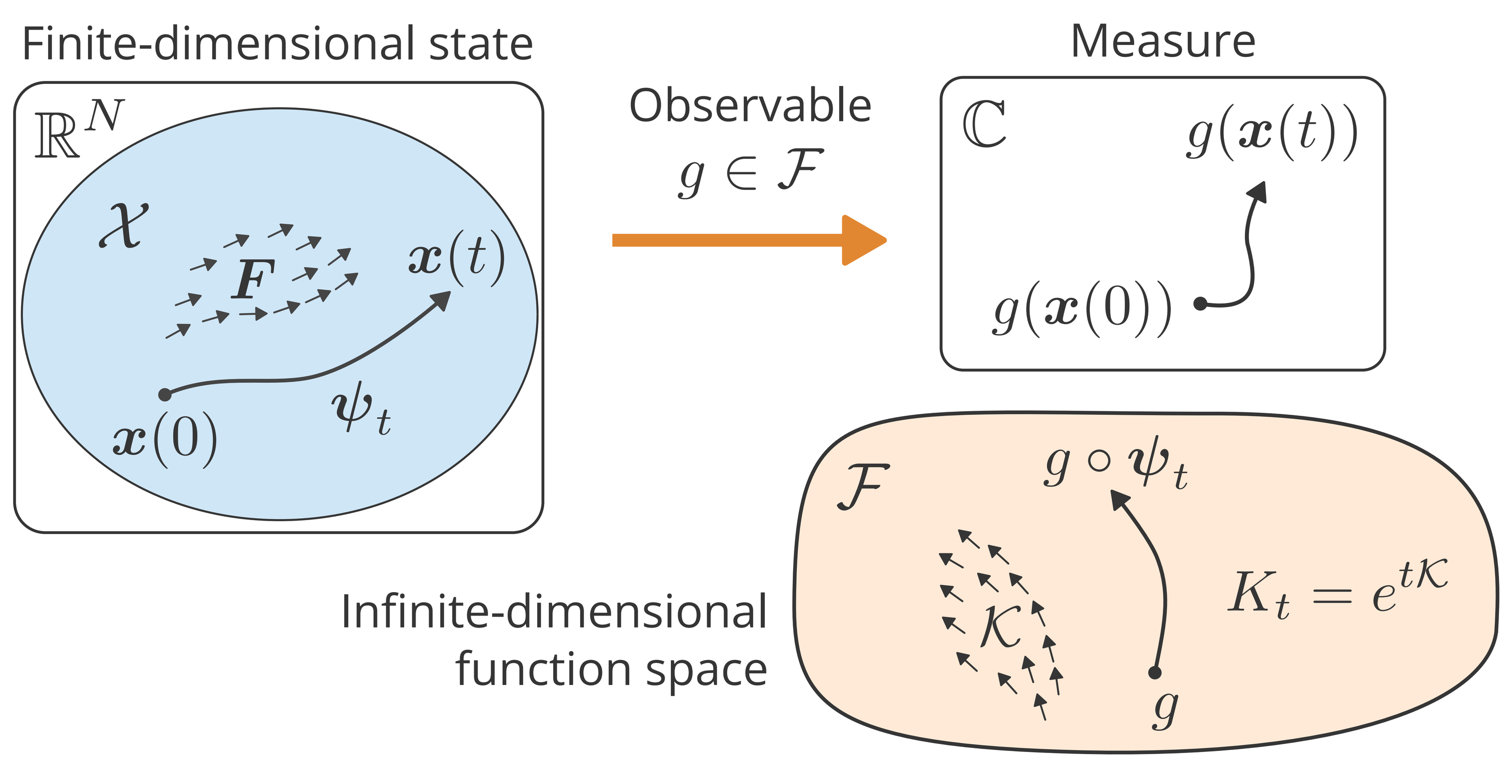
1. Département de physique, génie physique et d'optique, Université Laval, Québec, Canada
2. Centre Interdisciplinaire en Modélisation Mathématique de l'Université Laval, Québec, Canada
3. Centre de recherche CERVO, Québec, Canada

Dynamica · Université Laval

benjamin.claveau.1@ulaval.ca

KOOPMAN OPERATOR THEORY

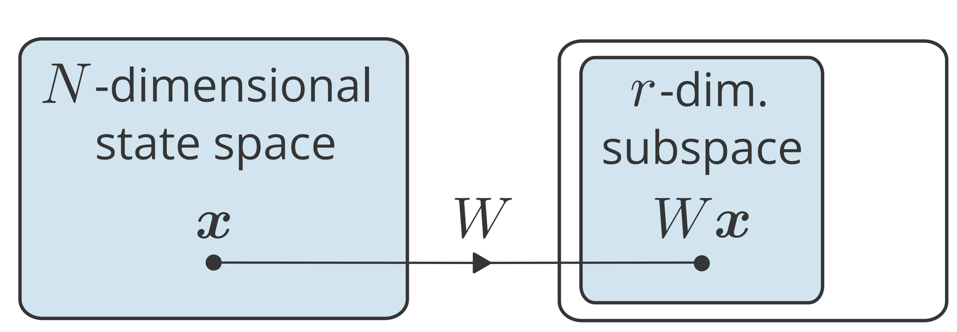
- Inspired by quantum mechanics, Koopman theory provides a mathematical framework that describes the behaviour of **observables of dynamical systems** [1].
- For linear and nonlinear systems, the **linear time-evolution operator** of the observables is the Koopman operator K_t. Its generator $\mathcal{K}$ is known from the vector field $\boldsymbol{F}$ as
$$\mathcal{K} = \sum_{i=1}^N F_i \frac{\partial}{\partial x_i}\,.$$
- An **eigenfunction** ϕ of the Koopman generator of eigenvalue λ is a particular observable with an exponential behaviour, i.e.
$$K_t[\phi](\boldsymbol{x}(t)) = e^{\lambda t}\phi(\boldsymbol{x}(0))\,.$$
- Koopman eigenfunctions are commonly approximated through data-driven methods [2], but **analytical approaches** can lead to **exact eigenfunctions and symmetries** [3, 4].



RESEARCH QUESTION

- We are interested in dynamics of complex networks with **weight matrix** W.
- Structural property of interest:

The **rank** of W is the dimension of its image.
$\mathrm{rank}(W) = r \leq N$



- Previous works relate the rank of W to the **dimension of the dynamics** of complex networks [4, 5], including recurrent neural networks (RNNs) [6].
- The exact effect of a low-rank weight matrix is still unclear in many cases.

Can Koopman eigenfunctions characterize the impact of the rank on the dynamics?

MAIN RESULT

We found **two families** for which rank deficiencies of W imply Koopman eigenfunctions:

1. $$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{1}{\zeta_i'(x_i)}\left[-\zeta_i(x_i) + \sum_{j=1}^N W_{ij} h_j(\boldsymbol{x})\right], \qquad \begin{aligned}\phi(x) &= \boldsymbol{u}^\top \boldsymbol{\zeta}(x)\\ \lambda &= -1\end{aligned}$$

2. $$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{1}{\zeta_i'(x_i)}\left[-c_i + \sum_{j=1}^N W_{ij} h_j(\boldsymbol{x})\right], \qquad \begin{aligned}\phi(x) &= \exp\left(\boldsymbol{u}^\top \boldsymbol{\zeta}(x)\right)\\ \lambda &= -\boldsymbol{u}^\top \boldsymbol{c}\end{aligned}$$

for $i \in \{1, \dots, N\}$ with $W^\top \boldsymbol{u} = \boldsymbol{0}$, x_i the activity of the i-th element, arbitrary functions ζ_i, h_j and arbitrary constants c_i.

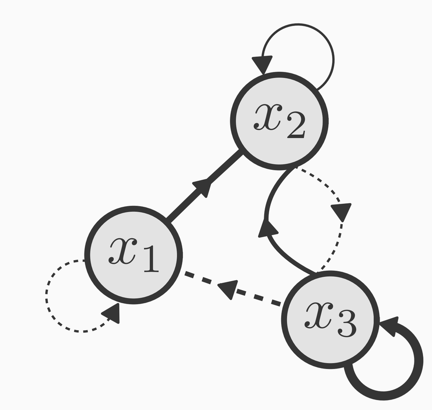
RECURRENT NEURAL NETWORKS

- Data-driven Koopman methods can be used to train RNNs without gradient descent [7] and improve performance in some neural network applications [8].
- In our case, for $\boldsymbol{\zeta}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\theta}$, **the first family of systems yields the RNN dynamics**
$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_i + \sum_{j=1}^N W_{ij}\sigma(x_j) + \theta_i\,, \qquad i \in \{1, \dots, N\}\,.$$
- Thus, RNNs with low-rank matrices have **affine Koopman eigenfunctions**
$$\phi(x) = \boldsymbol{u}^\top(\boldsymbol{x} - \boldsymbol{\theta})\,, \qquad W^\top \boldsymbol{u} = \boldsymbol{0}\,.$$
- Since the associated eigenvalues are negative, the dynamics of low-rank RNNs converge to **low-dimensional affine spaces.**

EXAMPLE

- 3 neuronal populations
- Rank 2 weight matrix
- $\boldsymbol{\theta} = \boldsymbol{0}$

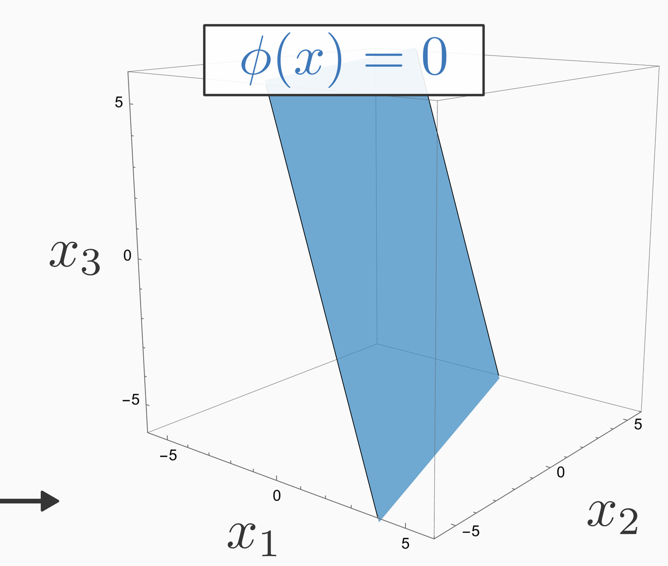
$$W = \begin{bmatrix} -1 & 0 & -2 \\ 3 & 1 & 2 \\ 0 & -1 & 4 \end{bmatrix}$$



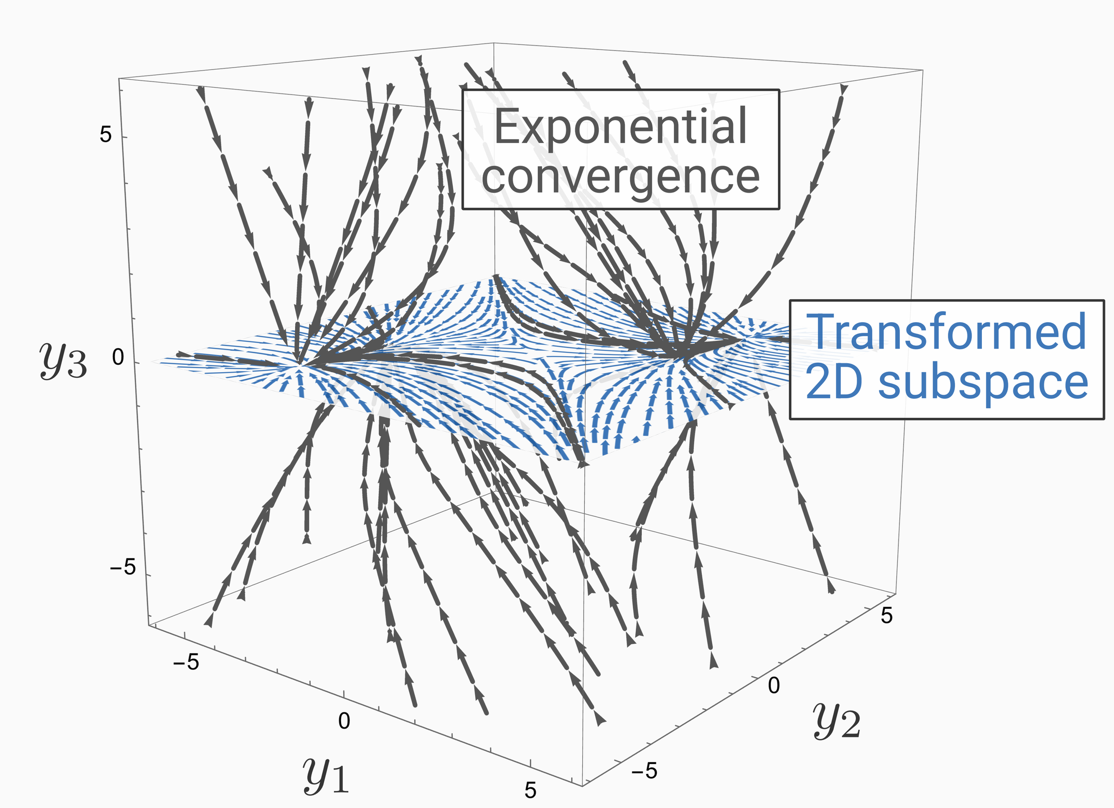
From the singular value decomposition $W = U\Sigma V^\top$, we compute a **left singular vector** $\boldsymbol{u}_3$ of null singular value. This vector is such that $W^\top \boldsymbol{u}_3 = \boldsymbol{0}$.

We thus obtain the **linear eigenfunction**
$$\phi(\boldsymbol{x}) = \boldsymbol{u}_3^\top \boldsymbol{x} = 3x_1 + x_2 + x_3\,, \qquad \lambda = -1\,.$$

The kernel of the Koopman eigenfunction defines a **globally attractive invariant subspace.** →



There is a useful **linear change of variables** $\boldsymbol{y} = U^\top \boldsymbol{x}$, where y_3 is the eigenfunction.



After the change of variables :

- Invariant subspace is now at $y_3 = 0$
- Exponential decrease of y_3 magnitude
- Long-term behaviour described by y_1, y_2

TAKEAWAYS AND FUTURE WORK

- We found two families of dynamics of complex systems for which rank deficiencies of the weight matrix imply Koopman eigenfunctions.
- In recurrent neural networks, these eigenfunctions describe the convergence of the activity towards invariant affine subspaces.
- This approach can be extended by identifying general families of dynamical systems which admit Koopman eigenfunctions of specified forms. By choosing a universal approximator as a Koopman eigenfunction, this framework yields dynamics with arbitrary approximate eigenfunctions.

[1] Budišić et al., Chaos (2012)
[2] Brunton et al., SIAM Rev. (2022)
[3] Salova et al., Chaos (2019)
[4] Thibeault et al., arXiv preprint (2025)
[5] Thibeault et al., Nature Physics (2024)
[6] Mastrogiuseppe and Ostojic, Neuron (2018)
[7] Bolager et al., arXiv preprint (2024)
[8] Xiao et al., IEEE Trans. Intell. Veh. (2023)

CIMMUL · UNIQUE CENTRE · Fonds de recherche Québec · CRSNG NSERC