ЦРНОГОРСКА АКАДЕМИЈА НАУКА И УМЈЕТНОСТИ
ГЛАСНИК ОДЈЕЉЕЊА ПРИРОДНИХ НАУКА, 26, 2023.

ЧЕРНОГОРСКАЯ АКАДЕМИЯ НАУК И ИСКУССТВ
ГЛАСНИК ОТДЕЛЕНИЯ ЕСТЕСТВЕННЫХ НАУК, 26, 2023

THE MONTENEGRIN ACADEMY OF SCIENCES AND ARTS
PROCEEDINGS OF THE SECTION OF NATURAL SCIENCES, 26, 2023

UDK 514.7:004.8

Vladimir Jaćimović[*], *Nevena Mijajlović*[*], *Zinaid Kapić*[**],
Aladin Crnkić[**], *Milojica Jaćimović*[***]

CONTINUOUS-TIME DYNAMICAL SYSTEMS ON MANIFOLDS FOR GEOMETRIC DEEP LEARNING

Abstract

We propose to use Kuramoto models on Lie groups and spheres in order to develop new framework for a wide class of problems of Geometric Deep Learning. Our approach is illustrated on unsupervised learning on the data set that contains economic indicators of 16 countries for the time period between 1990 and 2020.

Keywords: Geometric Deep Learning; Kuramoto model; unsupervised learning

I. MACHINE LEARNING VIA CONTINUOUS-TIME DYNAMICAL SYSTEMS

In recent decades neural networks have shown a surprising success in problems of Machine Learning and Artificial Intelligence. Explosive growth of neural networks with increasing number of layers have brought new perspectives into the field, leading to a great number of new challenges and applications that are broadly named Deep Learning.

Traditionally, neural networks are trained by backpropagation, which is typically implemented by the stochastic gradient ascent.

[*] Faculty of Natural Sciences and Mathematics, University of Montenegro, Podgorica, Montenegro, vladimir@ucg.ac.me, nevenami@ucg.ac.me

[**] Faculty of Technical Engineering, University of Bihać, Bihać, Bosnia and Herzegovina, zinaid.kapic@unbi.ba, aladin.crnkic@unbi.ba

[***] Montenegrin Academy of Sciences and Arts, Podgorica, Montenegro, milojica@jacimovic.me

Recently, many researchers have explored the idea of using continuous-time dynamical systems for problems of Machine Learning. This idea stems from the observation that neural networks are essentially discrete dynamical systems that map the input data into a certain output. From this point of view, continuous-time dynamical systems can be seen as neural networks with infinitely many (continuum) of layers. In particular, residual neural networks can be interpreted as a discretization (Euler scheme) of certain systems of ordinary differential equations, see [1].

There are several potential advantages of such an approach that are still to be explored. One obvious advantage is that differential equations have been studied for centuries in Mathematics and other sciences, and a great number of sophisticated solvers are available in various software packages for numerical mathematics. Using these solvers for problems of Machine Learning is a tempting idea.

A major step in this direction has been made in the highly influential paper [2], where authors have introduced so-called Neural ODE for problems of Deep Learning. Neural ODE is a neural network with infinitely many layers, in fact, a system of ODE's that is interpreted as a neural network. The main idea in [2] is to implement training of the network by backpropagation, using the classical adjoint method from theory of differential equations. In other words, neural network is in fact a system of ODE's, while the corresponding adjoint system is used for gradual adjustment of parameters (weights). This approach places deep learning into the framework of classical mathematical control theory and Pontryagin's maximum principle.

In parallel, normalizing flows have been introduced in [3, 4] for the problem of density estimation in Artificial Intelligence. The idea is to construct a sequence of invertible transformations in order to transform the data into a certain simple target density, from which one can easily obtain samples. This simple density is typically multivariate Gaussian density. Therefore, we construct a dynamical system, where initial conditions are randomly sampled from the Gaussian distribution and solve it in order to approximate the data chosen from a certain complicated probability distribution. This dynamical system can consist of a finite number of invertible transformations, but can also consist of continuum of infinitesimal transformations. In the later case we obtain a continuous-time dynamical system, where initial conditions are sampled from Gaussian distribution.

These approaches can greatly reduce the number of parameters (weights) to be learned, provided clever choice of the dynamical system, depending on the data and concrete problem.

II. GEOMETRIC DEEP LEARNING

In the last decade there has been an explosive interest into machine learning on non-Euclidean spaces. These investigations have merged into a broad field that has been named Geometric Deep Learning. This implied a growing interest of the Machine Learning community into Riemannian geometry and optimization on Riemannian manifolds.

This interest stems from the fact that many data have intrinsic non-Euclidean structure. Such data include angles, rotations, and various group transformations. Moreover, many real-world data are naturally embedded into non-Euclidean spaces. For instance, the data with hierarchical structure (such as trees and complex networks) hide an intrinsic structure that is described by hyperbolic geometry, see for instance [5]. In whole, there are enormous efforts aimed at discovering natural geometries that are hidden behind a certain data sets.

Motivated by these new developments, some researchers have introduced normalizing flows on non-Euclidean spaces, [6].

III. KURAMOTO MODEL AND ITS EXTENSIONS FOR DEEP LEARNING

One of paradigmatic models in Physics of Complex Systems has been introduced by the Japanese physicist Yoshiki Kuramoto in 1975. Kuramoto model describes an ensemble consisting N phase coupled oscillators, see [7]:

$$\dot{\varphi}_i(t) = \omega_i + \frac{K}{N}\sum_{j=1}^{N}\left(\varphi_i - \varphi_j\right), i = 1, \dots, N. (1)$$

In this model, each oscillator is coupled to all other oscillator and K stands for a (global) coupling strength. The notion ω_i stands for an intrinsic frequency of the i-the oscillator. In the absence of coupling ($K = 0$) each oscillator performs simple rotations on the unit circle with the constant angular speed ω_i.

Underline that each oscillator is fully described by its phase φ_i, while amplitudes are neglected. Since the phase is just an angle, each oscillator is described by a point on the unit circle S^1. Hence, the Kuramoto model (1) can be seen as a dynamical system on the unit circle (or on the N-torus). Since the circle is the simplest example of non-Euclidean manifold, the Kuramoto model provides a simple, but highly non-trivial example of a dynamical system on non-Euclidean manifold.

Kuramoto model and its variations have been studied for decades, mostly in context of synchronization and pattern formation in large ensembles of simple individuals.

Subsequently, the Kuramoto model has been extended from the circle to higher-dimensional non-Euclidean manifolds: matrix groups [8-10], homogeneous spaces [11], and spheres [12, 13].

This class of (extended) Kuramoto models provides a new framework with possible applications to problems of Machine Learning in non-Euclidean geometries. The main issue that is still to be explored is the problem of training Kuramoto networks, i.e. learning the couplings based on the data.

In the present paper, we focus on Kuramoto models on spheres. To that end we introduce the model:

$$\dot{x}_i = W_i x_i + \frac{K}{N} \sum_{j=1}^{N} \big(x_j - \langle x_i, x_j \rangle x_i\big) (2)$$

Here, x_i is generalized "phase" of the i-the "oscillator", i.e. x_i is a unit vector in $\mathbb{R}^d$, that is - a point on the $d-1$-dimensional sphere S^{d-1}. The symbol $\langle \cdot,\cdot \rangle$ stands for standard inner product in $\mathbb{R}^d$, while K denotes the global coupling strength. Finally, W_i is a $d \times d$ anti-symmetric matrix, that is interpreted as a "frequency" of generalized "oscillator" with index i.

IV. KURAMOTO MODEL ON SPHERE FOR GEOMETRIC DEEP LEARNING

It is tempting to apply Kuramoto models on manifolds to various problems of supervised and unsupervised Machine Learning. Arguably, the most important problem of unsupervised learning is clustering of data. In the present paper we continue investigations on clustering functional and dynamical data that have been presented in our previous papers [14, 15].

We extract some knowledge from the data set that contains economic and societal indicators of 16 developed countries. In such a way we investigate economic trends and the impact of recent crisis on economies of developed countries. The data set has been taken from [16].

In order to draw some conclusions, we solve the system (2) on the 3-sphere, where the data are encoded into time-dependent frequency matrix $W_i(t)$. Since W_i is a 3×3 anti-symmetric matrix, we can encode 6-dimensional data into W_i. For the data of higher dimension, we propose to use Kuramoto model on spheres of higher dimensions.

Hence, the data is encoded into generalizes frequencies of "oscillators", while initial conditions for the system (2) are randomly sampled from the uniform distribution on S^3.

Our simulation results are presented in the next Section. As one can see, our method successfully performs clustering of data and recognizes trends and changes of the situation.

V. SIMULATION: ECONOMIC INDICATORS OF 16 DEVELOPED COUNTRIES IN 1990-2020.

In this section, we present simulation results of our method on a data set containing economic indicators of 16 developed countries in the period from 1990 to 2020.

The method identifies two big clusters and four separate clusters as follows (see Figure 1):

I cluster: Belgium, Denmark, France, Netherlands, Sweden, Switzerland, United Kingdom, Finland,

II cluster: Canada, New Zealand, United States, Japan,

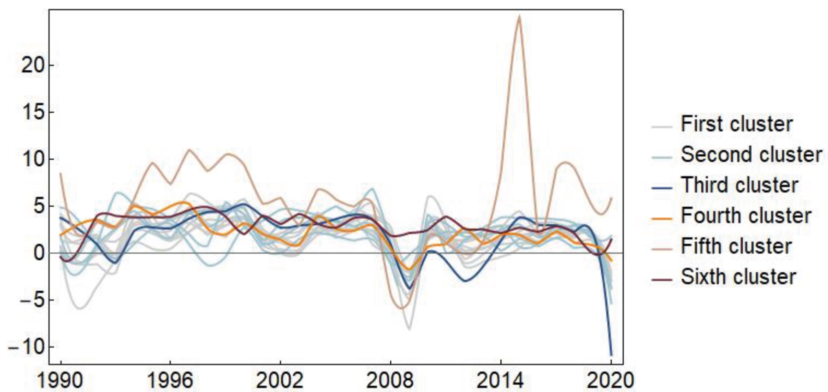
III cluster: Spain,

IV cluster: Norway,

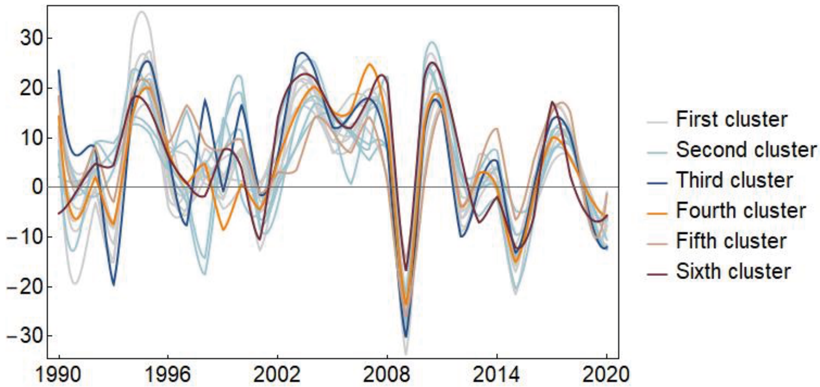
V cluster: Ireland,

VI cluster: Australia.

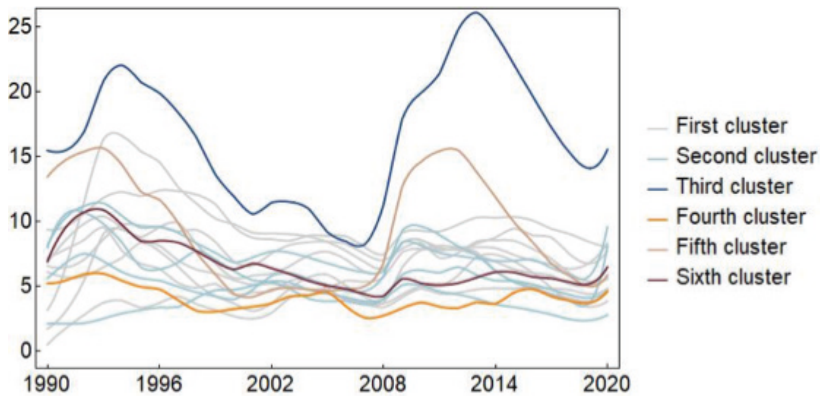
Spain, Norway, Ireland and Australia are set into separate clusters. There are various reasons that led to this. Spain was classified out in a separate cluster mainly due to the high unemployment rate and high interest rates in the early 1990s. As for Norway, the main reasons lie in the low unemployment rate, high exports in 2000 and low exports in 2015 and 2020. Ireland did not fall into big clusters because it had a high rate of GDP and industrial production from 1995 to 2000 and its rapid growth in 2015. Also, one of the reasons is its high interest rate in 2011. When it comes to Australia, the reason for being classified into a separate cluster can be found in high export rates in 2008 and 2010.



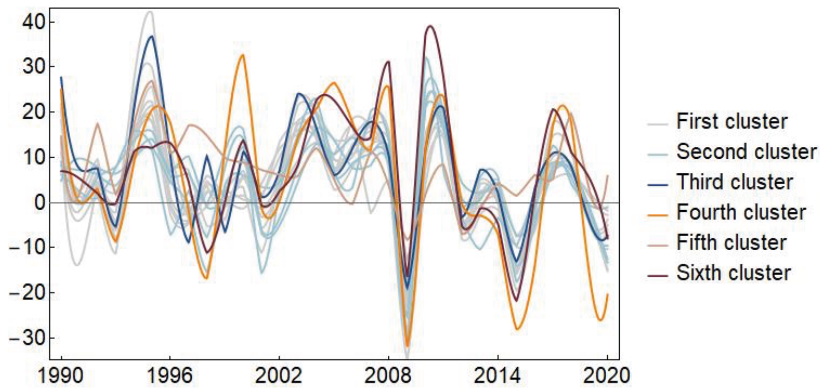
a)


30
20
10
0
−10
−20
−30
1990
1996
2002
2008
2014
2020
First cluster
Second cluster
Third cluster
Fourth cluster
Fifth cluster
Sixth cluster

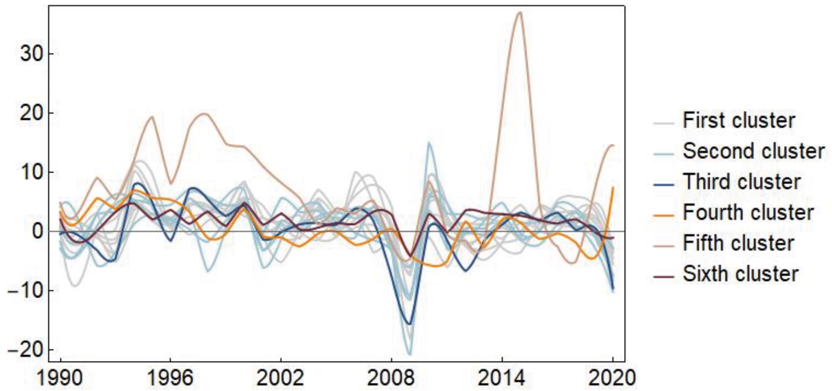
b)


25
20
15
10
5
0
1990
1996
2002
2008
2014
2020
First cluster
Second cluster
Third cluster
Fourth cluster
Fifth cluster
Sixth cluster

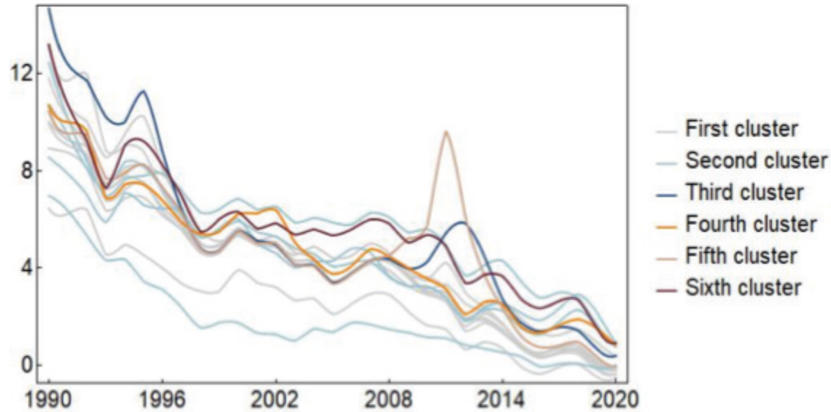
c)


40
30
20
10
0
−10
−20
−30
1990
1996
2002
2008
2014
2020
First cluster
Second cluster
Third cluster
Fourth cluster
Fifth cluster
Sixth cluster

d)



e)



f)

Figure 1. Simulation results: (a) Growth of GDP, (b) Import, (c) Unemployment, (d) Export, (e) Industrial production, and (f) Interest rates.

REFERENCES

[1] E. Weinan, "A Proposal on Machine Learning via Dynamical Systems," Communications in Mathematics and Statistics, vol. 5, pp. 1-11, 2017.

[2] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," arXiv preprint arXiv:1806.07366, 2018.

[3] G. Papamakarios, E. T. Nalisnick, D. Jimenez Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing Flows for Probabilistic Modeling and Inference," Journal of Machine Learning Research, vol. 22(57), pp. 1-64, 2021.

[4] D. Rezende, and M. Shakir, "Variational inference with normalizing flows," International conference on machine learning, PMLR, 2015.

[5] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, "Hyperbolic geometry of complex networks," Physical Review E, 82(3), 2010.

[6] D. Rezende, et al. “Normalizing flows on tori and spheres,” International Conference on Machine Learning, PMLR, 2020.
[7] Y. Kuramoto, “Self-entrainment of a population of coupled nonlinear oscillators,” Proceedings of international symposium on mathematical problems in theoretical physics, vol. 39, pp. 420-422, 1975.
[8] M. A. Lohe, “Non-Abelian Kuramoto models and synchronization,” JPhys A Math Theor, vol. 42, 2009.
[9] M. A. Lohe, “Systems of matrix Riccati equations, linear fractional transformations, partial integrability and synchronization,” Journal of Mathematical Physics, vol. 60(7), p.072701, 2019.
[10] V. Jaćimović, and A. Crnkić, “Low-dimensional dynamics in non-Abelian Kuramoto model on the 3-sphere,” Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 28(8), p.083105, 2018.
[11] A. Sarlette, and R. Sepulchre, “Consensus Optimization on Manifolds,” SIAM Journal on Control and Optimization, vol. 48(1), pp.56-76, 2009.
[12] J. Markdahl, J. Thunberg and J. Gonçalves, "Almost Global Consensus on the n - Sphere," in IEEE Transactions on Automatic Control, vol. 63(6), pp. 1664-1675, 2018.
[13] S. Ha, and H. Park, “From the Lohe Tensor Model to the Lohe Hermitian Sphere Model and Emergent Dynamics,” SIAM Journal on Applied Dynamical Systems, vol. 19(2), pp.1312-1342, 2020.
[14] A. Crnkić, and V. Jaćimović, “Data clustering based on quantum synchronization,” Natural Computing, vol. 18(4), pp.907-911, 2018.
[15] A. Crnkić, I. Ivanović, V. Jaćimović, and N. Mijajlović, “Swarms on the 3-sphere for online clustering of multivariate time series and data streams,” Future Generation Computer Systems, vol. 112, pp. 11-17, 2020.
[16] Organisation for Economic Co-operation and Development (2022), Macronomic indicators of OECD countries, [online] Available at: https://stats.oecd.org/Index.aspx