Received 28 May 2020; revised 28 October 2020; accepted 3 November 2020. Date of publication 16 November 2020; date of current version 11 December 2020. The review of this article was arranged by Associate Editor S. Aviyente.

Digital Object Identifier 10.1109/OJSP.2020.3038369

Subspace Detection and Blind Source Separation of Multivariate Signals by Dynamical Component Analysis (DyCA)

CHRISTIAN UHL [1], MORITZ KERN [1], MONIKA WARMUTH [1], AND BASTIAN SEIFERT [2] (Member, IEEE)

[1]Center for Signal Analysis of Complex Systems, University of Applied Sciences Ansbach, 91522 Ansbach, Germany
[2]Department of Computer Science, ETH Zurich, CH-8092 Zïrich, Switzerland

CORRESPONDING AUTHOR: CHRISTIAN UHL (e-mail: christian.uhl@hs-ansbach.de)

This work was supported in part by by the German Federal Ministry of Education and Research (BMBF, Funding number: 05M20WBA) and in part by the European Regional Development Fund (ERDF, Project: TZM).

ABSTRACT The decomposition of a multivariate signal is an important tool for the analysis of measured or simulated data leading to possible detection of the relevant subspace or the sources of the signal. A new method – dynamical component analysis (DyCA) – is based on modeling the signal by a set of coupled ordinary differential equations. Its derivation and its features are presented in-depth. The corresponding algorithm is nearly as simple as principal component analysis (PCA). The results obtained by DyCA however yield a deeper insight into the underlying dynamics of the data. To illustrate the broad area of possible applications a set of examples of analyzing data by DyCA is presented - involving both measured EEG, motion and ECG data as well as data obtained from stochastic differential equations. Thereby our alternative tool for dimensionality reduction is compared toresults obtained PCA and ICA and demonstrate the gain of this approach.

INDEX TERMS Biomedical data, blind source separation, differential equations, dimensionality reduction, dynamical component analysis, independent component analysis, low dimensional dynamics, motion detection, principal component analysis.

I. INTRODUCTION

Even though many aspects of data science and signal processing are nowadays dominated by the usage of machine learning tools, and neural networks in particular, there is still a great interest in deterministic (and thus interpretable) tools for the processing of signals. Indeed it has been argued recently [1] that trying to explain black box models might maintain bad practices and potentially cause problems in healthcare, criminal justice and other critical domains.

In this paper we present a dimensionality reduction method – dynamical component analysis (DyCA) – introduced in the preliminaries works of Seifert *et al.* [2] and Korn *et al.* [3]. The proposed method is governed by the assumption that a multivariate measurement of a dynamical system can be split into a deterministic part, which can be described by a system of differential equations, and independent noise components. This is for example the case for EEG data of epileptic seizures, which can approximately be described by a system of ordinary differential equations consisting of two linear and one non-linear equations, as shown in [3], [4]. This approach has similarities to the method of PIPs (principal interacting patterns) and POPs (principal oscillating patterns) introduced by Hasselmann [5] and further developed by Kwasniok [6], [7]. Compared to [2], [3] we present the proposed method more in-depth and detailed and present a broad spectrum of possible applications.

A. RELATED WORK

Principal component analysis (PCA) has been widely used in a broad spectrum of possible applications, also under different names like empirical orthogonal function (EOF) analysis, proper orthogonal decomposition or Karhunen-Loéve expansion, maximizing the variance of the principal components [8], [9].

In geophysical applications EOF analysis have been applied to spatio-temporal climate data to obtain data-driven models [10] and have been further developed in different directions like independent subspace analysis [11], linear and nonlinear dynamical mode decomposition [12], [13] to name only a few.

In the field of chemometrics and process controling dynamic PCA [14], [15] and dynamic-inner PCA [16] have been introduced and applied considering the data matrix and augmented time-lagged values of the data. Recursive PCA methods [17], [18] allow an efficient recursive updating of the principal components.

PCA extensions like robust subspace tracking and learning [19]–[22] can be found in video analytics applications (foreground - background separation). These aim at detecting and tracking of low-dimensional subspaces slowly changing in time corrupted by sparse outliers.

Independent component analysis (ICA) [23]–[25] is based on the statistical assumption that the relevant components of the signals are statistically independent non-Gaussian signals. It has been applied in different scientific fields [26], e.g. also in the field of neuroscience to analyze EEG- and fMRI-signals (e.g. [27]–[29]). By detecting independent sources ICA facilitates artifact-removal and reveals interesting insights to understand brain sources.

By introducing DyCA, we provide an alternative method for subspace detection and source separation focusing on sources which are not statistically independent but dynamically coupled by a set of ODE. To achieve this, the time derivative of the signal is considered as well, which in return corresponds to the analysis of the its time-lagged representation. DyCA aims at the detection of low-dimensional subspaces containing the dynamics of the signal corrupted by - what we call - noise components.

Other related methods are given by dynamic mode decomposition (DMD) [30] searching for modes with a fixed growth rate and oscillation frequency, forecastable component analysis (ForeCA) [31] decomposing the signal in amplitudes optimized to forecasting, and multivariate empirical mode decomposition (multivariate EMD) [32] aiming at the extraction of common rotational modes. Further approaches in the time-frequency domain are variational mode decompostion (VMD) [33], multivariate variational mode decomposition [34] and a decomposition of multichannel nonstationary multicomponent signals presented in [35]. These methods pursue the same goal as DyCA, which is to extract multivariate multicomponent signals, but they differ in their approach. Our proposed method is not based in the time-frequency domain but rather focusing on the coupling of the components in terms of differential equations.

Two other methods share the name dynamical component(s) analysis. In [36] first a temporal and then kernelized spatial PCA is performed to obtain dynamical components of fMRI data. In [37] the subspace with maximal predictive information is obtained using a Gaussian approximation of the data.

B. ORGANIZATION OF PAPER

The paper is organized as follows. In Section II we deduce DyCA in detail, discuss the properties and present the algorithm. The application of DyCA to different kind of data sets is presented Section III and a comparison to PCA and ICA is shown. Finally we conclude our results in Section IV followed by an Appendix providing the basic theorems required for the derivation of DyCA.

II. DYNAMICAL COMPONENT ANALYSIS (DyCA)

A. ASSUMPTIONS AND GOAL

Starting point is a multivariate time series $q(t) \in \mathbb{R}^N$ with N vector components and t representing the time in a discretized manner: $t = t_1, t_2, \ldots, t_T$ with $T \geq N$. It is assumed that under ideal conditions the signal $q(t)$ can be decomposed into deterministic components

$$q(t) = \sum_{i=1}^{n} x_i(t) w_i. \tag{1}$$

In practice, however, the time series is often contaminated with noise components $\sum_{j=1}^{p} \xi_j(t)\psi_j$, i.e.

$$q(t) = \sum_{i=1}^{n} x_i(t) w_i + \sum_{j=1}^{p} \xi_j(t)\psi_j \quad \text{with } n + p \leq N \tag{2}$$

and $w_i, \psi_j \in \mathbb{R}^N$ being linearly independent for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. This assumption of linear independent components leads to the above mentioned relation $n + p \leq N$.

The deterministic amplitudes $x_i(t)$ are assumed to obey a set of ordinary differential equations (ODE), of which m are linear

$$\begin{aligned} \dot{x}_1(t) &= \sum_{k=1}^{n} a_{1,k} x_k(t) \\ &\vdots \\ \dot{x}_m(t) &= \sum_{k=1}^{n} a_{m,k} x_k(t), \end{aligned} \tag{3}$$

with $m \geq \frac{n}{2}$. The corresponding coefficient matrix,

$$A := \begin{bmatrix} \underbrace{\begin{matrix} a_{1,1} & \cdots & a_{1,m} \\ a_{2,1} & \cdots & a_{2,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{matrix}}_{=:A_1 \in \mathbb{R}^{m\times m}} & \underbrace{\begin{matrix} a_{1,m+1} & \cdots & a_{1,n} \\ a_{2,m+1} & \cdots & a_{2,n} \\ \vdots & \ddots & \vdots \\ a_{m,m+1} & \cdots & a_{m,n} \end{matrix}}_{=:A_2 \in \mathbb{R}^{m\times (n-m)}} \end{bmatrix}, \tag{4}$$

can be written as indicated by two submatrices $A = [A_1, A_2] \in \mathbb{R}^{m\times n}$. The remaining $n - m$ ODEs are non-linear, i.e.

$$\begin{aligned} \dot{x}_{m+1}(t) &= f_{m+1}(x_1(t), x_2(t), \ldots, x_n(t)) \\ &\vdots \\ \dot{x}_n(t) &= f_n(x_1(t), x_2(t), \ldots, x_n(t)), \end{aligned} \tag{5}$$

whereupon $f_{m+1}, \ldots, f_n$ are unknown, non-linear, smooth functions. These functions are mentioned in this context to fully represent the assumed dynamics. However, the proposed algorithm is able to extract not only the amplitudes $x_1, \ldots, x_m$ but also the amplitudes $x_{m+1}, \ldots, x_n$ although the non-linear functions $f_{m+1}, \ldots, f_n$ remain unknown in our context. This is achieved by considering only the linear differential equations (3).

The amplitudes $\xi_j(t)$ are considered to be of stochastic character. Based on these assumptions we are dealing with stationary signals, non-stationarities can only be considered if they occur in the stochastic components.

In terms of matrix notation we can rewrite (2) as

$$Q = WX + \Psi\Xi, \tag{6}$$

with

$$Q = \begin{bmatrix} | & | & & | \\ q(t_1) & q(t_2) & \cdots & q(t_T) \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{N\times T},$$

$$W = \begin{bmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_n \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{N\times n},$$

$$X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_T) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_T) \end{bmatrix} = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \in \mathbb{R}^{n\times T}, \tag{7}$$

as well as

$$\Psi = \begin{bmatrix} | & | & & | \\ \psi_1 & \psi_2 & \cdots & \psi_p \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{N\times p}, \quad \text{and}$$

$$\Xi = \begin{bmatrix} \xi_1(t_1) & \xi_1(t_2) & \cdots & \xi_1(t_T) \\ \vdots & \vdots & \ddots & \vdots \\ \xi_p(t_1) & \xi_p(t_2) & \cdots & \xi_p(t_T) \end{bmatrix} = \begin{bmatrix} \xi_1(t) \\ \vdots \\ \xi_p(t) \end{bmatrix} \in \mathbb{R}^{p\times T}.$$

As an additional condition we must assume that the data matrix Q and its time derivative $\dot{Q}$ are of full rank N. The goal of DyCA is to perform a dimensionality reduction of the signal $q(t)$ such that the underlying dynamics of the ODE system are captured in the best possible way and hence the deterministic part of the signal WX can be separated from the stochastic part $\Psi\Xi$. The method strongly utilizes the fact that the amplitudes $x_i(t)$ representing the matrix X are governed by a set of differential equations (3) and (5).

We would like to emphasize at this point that one does not need to know neither the exact parameters $a_{i,k}$ of the system, nor the dimensions n and m. Rather, after applying DyCA we obtain not only estimates $\widetilde{x}_i(t)$ and $\widetilde{w}_i$ for the actual amplitudes $x_i(t)$ and DyCA components w_i, but also estimates for the possibly unknown parameters $a_{i,k}$, n and m. However, to estimate the number n of differential equations we need the additional condition that the submatrix A_2 of the coefficient matrix of the ODE (4) has full rank $n - m$.

To achieve those goals, we are seeking a generalized left inverse $W^- \in \mathbb{R}^{n\times N}$ of W, i.e. $W^-W = I_n$, such that

$$X = W^- Q. \tag{8}$$

Due to the assumption of linearly independent vectors (modes) w_i we can consider the rows in W^- to be a set of linearly independent projecting vectors $\{u_1^\top, \ldots, u_n^\top\}$ (see lemma IV.3). Hence, the amplitudes $x_i(t)$ can be calculated by the scalar product

$$x_i(t) = u_i^\top q(t) = q(t)^\top u_i. \tag{9}$$

The time derivative of (9) is given by

$$\dot{x}_i(t) = \dot{q}(t)^\top u_i. \tag{10}$$

DyCA aims at estimating these projecting vectors u_i first and approximating the corresponding modes w_i in a second step.

B. DERIVATION OF DyCA

The basic idea of DyCA can be described as follows: Using conditions (9) and (10), we fit the data to the linear part of the ODE (3) in the Euclidean norm. This leads via a least squares minimization problem to a generalized eigenvalue problem. By solving this eigenvalue problem and introducing a suitable threshold, we obtain an estimate for the number of u_i for which this linear approximation is well suited, i.e. for which the error of the fitting is small, as well as an estimation for those first u_i themselves. The condition to the rank of A_2 then yields the missing u_i. We will now derive that procedure in detail.

By considering the linear differential equations (3), i.e. $\dot{x}_i(t) = \sum_{k=1}^n a_{i,k}x_k(t)$ for $i = 1, \ldots, m$, and inserting projections (9) and their time derivatives (10) we obtain

$$\dot{q}(t)^\top u_i = \sum_{k=1}^{n} a_{i,k} q(t)^\top u_k = q(t)^\top v_i \tag{11}$$

with $v_i := \sum_{k=1}^n a_{i,k}u_k$ for $i = 1, \ldots, m$.

For the estimation of the projecting vectors u_i and v_i we use a least squares approach that is generally well-suited for data with additional noise. We first define the error function

$$D(u_i, v_i) = \frac{\langle \|\dot{q}(t)^\top u_i - q(t)^\top v_i\|^2\rangle}{\langle \|\dot{q}(t)^\top u_i\|^2\rangle} \tag{12}$$

for all $i = 1, \ldots, m$ with $\|\cdot\|$ denoting the Euclidean norm and compute the approximation vectors $\widetilde{u}_i$, $\widetilde{v}_i$ as a solution of the least squares problem

$$\begin{matrix}\widetilde{u}_1,\ldots,\widetilde{u}_m,\\ \widetilde{v}_1,\ldots,\widetilde{v}_m\end{matrix} = \underset{\substack{u_1,\ldots,u_m\in\mathbb{R}^N\\ v_1,\ldots,v_m\in\mathbb{R}^N}}{\arg\min} \sum_{i=1}^{m} \frac{\langle\|\dot{q}(t)^\top u_i - q(t)^\top v_i\|^2\rangle}{\langle\|\dot{q}(t)^\top u_i\|^2\rangle} \tag{13}$$

subject to the u_i being pairwise linearly independent.[1] We would like to highlight that for minimizing (13), we have to minimize each of the m summands, i.e. $D(u_1, v_1), \ldots, D(u_m, v_m)$, while obeying the condition that the u_i are linearly independent. Since these summands are each of the same nature, we neglect the index i for reasons of simplicity and only examine a general error function $D(u, v) = \frac{\langle\|\dot{q}(t)^\top u - q(t)^\top v\|^2\rangle}{\langle\|\dot{q}(t)^\top u\|^2\rangle}$ for minima. Introducing the correlation matrices

$$\begin{aligned} C_0 &= \langle q(t)q(t)^\top\rangle = \frac{1}{T}QQ^\top \\ C_1 &= \langle \dot{q}(t)q(t)^\top\rangle = \frac{1}{T}\dot{Q}Q^\top \\ C_2 &= \langle \dot{q}(t)\dot{q}(t)^\top\rangle = \frac{1}{T}\dot{Q}\dot{Q}^\top \end{aligned} \tag{14}$$

the error functions (12) can be rewritten as

$$\begin{aligned} D(u, v) &= \frac{\langle(\dot{q}(t)^\top u - q(t)^\top v)^\top(\dot{q}(t)^\top u - q(t)^\top v)\rangle}{\langle(\dot{q}(t)^\top u)^\top(\dot{q}(t)^\top u)\rangle} \\ D(u, v) &= \frac{u^\top C_2\, u - 2\, u^\top C_1 v + v^\top C_0 v}{u^\top C_2\, u} \\ D(u, v) &= 1 - \frac{2\, u^\top C_1 v - v^\top C_0 v}{u^\top C_2\, u}. \end{aligned} \tag{15}$$

Variation of (15) with respect to v and setting zero yields

$$\frac{\partial D}{\partial v} = 0 \quad \Rightarrow \quad 2\, u^\top C_1 = v^\top(C_0 + C_0^\top)$$

$$C_1^\top u = C_0 v, \tag{16}$$

since the correlation matrix C_0 is symmetric. Variation with respect to u we obtain

$$2(u^\top C_2\, u)v^\top C_1^\top = (2\, u^\top C_1 v - v^\top C_0 v)u^\top(C_2 + C_2^\top)$$

$$C_1 v = \lambda C_2\, u \tag{17}$$

since $C_2 = C_2^\top$ with

$$\lambda = \frac{2\, u^\top C_1 v - v^\top C_0 v}{u^\top C_2\, u}. \tag{18}$$

As we assumed the data matrix Q and its time derivative $\dot{Q}$ to be of full rank N, the correlations matrices C_0 and C_2 are positive definite and hence, regular with inverses C_0^{-1} and C_2^{-1}. Inverting (15) to

$$v = C_0^{-1} C_1^\top u \tag{19}$$

and inserting (18) into (16) leads to a generalized eigenvalue problem for the projecting vectors u

$$C_1 C_0^{-1} C_1^\top u = \lambda C_2\, u \tag{20}$$

In case of singular matrices C_0 and C_2, i.e. if the data matrix Q and its time derivative $\dot{Q}$ are not of full rank N, the original signal $q(t)$ has to be preprocessed by projecting into an non-redundant subspace by PCA projection neglecting the components with minimal contribution to the signal.

C. SOLVING THE LEAST SQUARES PROBLEM

By solving the generalized eigenvalue problem (19) we receive N eigenvalues λ_i, $i = 1, \ldots, N$ as well as N associated eigenvectors $\widetilde{u}_i$. Due to the symmetry of $C_1 C_0^{-1} C_1^\top$ and the positive definiteness of C_2, all eigenvalues λ_i and thus also the eigenvectors $\widetilde{u}_i$ are real. In addition, the $\widetilde{u}_i$ are linearly independent and pairwise orthogonal in the scalar product $(\cdot, \cdot)_{C_2}$ and therefore form a basis of $\mathbb{R}^N$.[2] Furthermore, we immediately obtain N corresponding vectors $\widetilde{v}_i$ from equation (19). Our task now is to find from these N candidates $(\widetilde{u}_i, \widetilde{v}_i)$ that one which actually minimizes the error function $D(u, v)$. In order to do so, we first sort the eigenvalues λ_i in descending order

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \tag{21}$$

and sort the $\widetilde{u}_i$ and $\widetilde{v}_i$ accordingly. Considering again the definition of λ in (18) one can easily notice that

$$D(u, v) = 1 - \lambda \tag{22}$$

which is obviously minimal for the largest eigenvalue λ_1. All in all, it holds that

$$\min_{u,v\in\mathbb{R}^N} D(u, v) = 1 - \lambda_1 \tag{23}$$

$$\underset{u,v\in\mathbb{R}^N}{\arg\min}\, D(u, v) = (\widetilde{u}_1, \widetilde{v}_1). \tag{24}$$

Through this insight, we now directly obtain the first m eigenvectors $\widetilde{u}_i$ of (20) with corresponding $\widetilde{v}_i$ as solution of the original least squares problem (13). It holds that

$$\underset{\substack{u_1,\ldots,u_m\in\mathbb{R}^N\\ v_1,\ldots,v_m\in\mathbb{R}^N}}{\arg\min} \sum_{i=1}^{m} D(u_i, v_i) = (\widetilde{u}_1, \ldots, \widetilde{u}_m, \widetilde{v}_1, \ldots, \widetilde{v}_m) \tag{25}$$

$$\min_{\substack{u_1,\ldots,u_m\in\mathbb{R}^N\\ v_1,\ldots,v_m\in\mathbb{R}^N}} \sum_{i=1}^{m} D(u_i, v_i) = \sum_{i=1}^{m}(1 - \lambda_i). \tag{26}$$

D. PROPERTIES OF DyCA

1) EIGENVALUES λ_i AND ESTIMATION OF m

Considering (26) one can easily see that each single error function $D(u_i, v_i)$ measures the quality of the fit with the help

[1] Note that $\langle c(t)\rangle := \frac{1}{T}\sum_{j=1}^{T} c(t_j)$ defines the time average of some vector $c(t) \in \mathbb{R}^N$ over all time points $t = t_1, \ldots, t_T$.

[2] Note that each symmetric positive definite matrix $B \in \mathbb{R}^{N\times N}$ defines a scalar product by $(x, y)_B = x^\top By$ for some vectors $x, y \in \mathbb{R}^N$.

Algorithm 1: Dynamical Component Analysis.

function DyCAQ, α
$\dot{Q} \leftarrow$ time-derivative of input signal
$C_0 \leftarrow \frac{1}{T} Q Q^\top$ ▷Compute auto-correlation of the signal
$C_1 \leftarrow \frac{1}{T} \dot{Q} Q^\top$ ▷Compute cross-correlation of the signal and its time-derivative
$C_2 \leftarrow \frac{1}{T} \dot{Q} \dot{Q}^\top$ ▷Compute auto-correlation of the time-derivative
$\lambda, \widetilde{u} \leftarrow$ solutions of $C_1 C_0^{-1} C_1^\top u = \lambda C_2\, u$ ▷solve generalized eigenvalue problem
Sort s.t. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$
$\widetilde{m} \leftarrow |\{\lambda_i \,|\, \lambda_i \geq \alpha\}|$ ▷Estimate significant subspace for linear equations (3)
$\widetilde{v}_i \leftarrow C_0^{-1} C_1^\top \widetilde{u}_i$ ▷Calculate projection vectors for non-linear equations (5)
$\widetilde{n} \leftarrow \dim(\operatorname{span}\{\widetilde{u}_1, \ldots, \widetilde{u}_{\widetilde{m}}, \widetilde{v}_1, \ldots, \widetilde{v}_{\widetilde{m}}\})$ ▷Estimate dimensionality of significant subspace
$\widetilde{W}^- \leftarrow [\widetilde{u}_1, \ldots, \widetilde{u}_{\widetilde{m}}, \widetilde{v}_{k_1}, \ldots, \widetilde{v}_{k_{\widetilde{n}-\widetilde{m}}}]^\top$ ▷Choose minimal spanning set of significant subspace $\mathbb{R}^{\widetilde{n}}$
$\widetilde{X} \leftarrow \widetilde{W}^- Q$ ▷Compute projected time-series
$C_{\widetilde{X}} \leftarrow \frac{1}{T} \widetilde{X} \widetilde{X}^\top$ ▷Compute auto-correlation of projected signal
$\widetilde{W} \leftarrow \frac{1}{T} Q \widetilde{X}^\top C_{\widetilde{X}}^{-1}$ ▷Estimate pseudoinverse of $\widetilde{W}^-$
$\widetilde{Q} \leftarrow \widetilde{W} \widetilde{X}$ ▷Compute reconstructed time-series
$\widetilde{U} \leftarrow [\widetilde{u}_1, \ldots, \widetilde{u}_{\widetilde{m}}]^\top$
$\widetilde{A} \leftarrow \frac{1}{T} \widetilde{U} \dot{Q} \widetilde{X}^\top C_{\widetilde{X}}^{-1}$ ▷Estimate coefficient matrix (4) of the ordinary differential-equation
return $\widetilde{X}, \widetilde{W}, \widetilde{A}$
end Function

of the corresponding eigenvalue λ_i:

$$\min_{u_i, v_i \in \mathbb{R}^N} D(u_i, v_i) = D(\widetilde{u}_i, \widetilde{v}_i) = 1 - \lambda_i, \tag{27}$$

i.e. for eigenvalues $\lambda_i \approx 1$ the linear approximation in terms of (3) is well-suited. These eigenvalues correspond to amplitudes $\widetilde{x}_i(t) = q(t)^\top \widetilde{u}_i$ with its time derivative $\dot{\widetilde{x}}_i(t)$ approximating the left-hand side of one of the assumed linear differential equations (3). The right-hand side is approximated by $q(t)^\top \widetilde{v}_i$ due to equation (11) calculated by the eigenvectors $\widetilde{u}_i$ and equation (19). The number $\widetilde{m}$ of eigenvalues λ_i close to one is an estimate for the actual number m of linear differential equations occurring in the dynamic model. For a threshold value $\alpha > 0$ it holds:

$$\widetilde{m} = \big|\{\lambda_i \,\big|\, \lambda_i \geq \alpha\}\big| \tag{28}$$

2) ESTIMATION OF THE PARAMETER *n*

Furthermore we now also get an estimate for the parameter n, which we call $\widetilde{n}$. In order to explain the idea behind this estimation procedure, however, we first return to the original u_i and v_i from (9) and (11). In consequence of the linear independence of $\{u_1, \ldots, u_n\}$ this set forms a basis of an n-dimensional subspace of $\mathbb{R}^N$. Due to the definition of the $v_i = \sum_{k=1}^{n} a_{i,k} u_k$, $i = 1, \ldots, m$ as a linear combination of the $\{u_1, \ldots, u_n\}$ it is obvious that

$$\{v_1, \ldots, v_m\} \subseteq \operatorname{span}\{u_1, \ldots, u_n\}. \tag{29}$$

As a result of the assumption that the matrix A_2 of the ODE coefficient matrix $A = [A_1, A_2]$ has full rank $n - m$, it follows that $\operatorname{rank}(A) \geq n - m$. This guarantees the linear independence of at least $n - m$ vectors of the $\{v_1, \ldots, v_m\}$ (see lemma IV.4). Furthermore, the condition to the rank of A_2 indicates that $n - m$ elements of the set of vectors v_i cannot be represented by a linear combination of only the first m vectors u_i but also require contributions from $u_{m+1}, \ldots, u_n$. As a result of lemma IV.5 and the Steinitz exchange lemma IV.6 as well as the condition $m \geq \frac{n}{2}$ it is possible to replace $u_{m+1}, \ldots, u_n$ with suitable, linearly independent v_i. After renumbering the v_i appropriately the following holds:

$$\operatorname{span}\{u_1, \ldots, u_n\} = \operatorname{span}\{u_1, \ldots, u_m, v_{m+1}, \ldots, v_n\} \cong \mathbb{R}^n. \tag{30}$$

For the projection vectors $\widetilde{u}_i, \widetilde{v}_i$ with $i = 1, \ldots, \widetilde{m}$ obtained from DyCA, we also pursue the idea of considering the $\widetilde{v}_i$ as a replacement for the missing $\widetilde{n} - \widetilde{m}$ vectors $\widetilde{u}_i$ and thus obtain a basis of an $\widetilde{n}$-dimensional subspace of $\mathbb{R}^N$ consisting of the $\widetilde{m}$ pairs of projection vectors $\widetilde{u}_i$ and suitable $\widetilde{v}_i$. Therefore we consider the linear hull of all $\widetilde{u}_i, \widetilde{v}_i$

$$\operatorname{span}\{\widetilde{u}_1, \ldots, \widetilde{u}_{\widetilde{m}}, \widetilde{v}_1, \ldots, \widetilde{v}_{\widetilde{m}}\}, \tag{31}$$

that spans an $\widetilde{n}$-dimensional subspace of $\mathbb{R}^N$ and define

$$\widetilde{n} := \dim(\operatorname{span}\{\widetilde{u}_1, \ldots, \widetilde{u}_{\widetilde{m}}, \widetilde{v}_1, \ldots, \widetilde{v}_{\widetilde{m}}\}) \tag{32}$$

as an estimate of n.

E. ESTIMATION OF THE DyCA COMPONENTS w_i AND RECONSTRUCTION OF THE SIGNAL $q(t)$

Reconsidering (31) and (32) we obtain a basis for the relevant subspace by choosing a minimal subset of vectors $\widetilde{v}_i$ linearly independent to all $\widetilde{u}_i$, spanning $\mathbb{R}^{\widetilde{n}}$, and renaming the selected $\widetilde{v}_i$ as $\widetilde{u}_{\widetilde{m}+1}, \ldots, \widetilde{u}_{\widetilde{n}}$, i.e.

$$\operatorname{span}\{\widetilde{u}_1, \ldots, \widetilde{u}_{\widetilde{m}}, \widetilde{u}_{\widetilde{m}+1}, \ldots, \widetilde{u}_{\widetilde{n}}\} \cong \mathbb{R}^{\widetilde{n}} \tag{33}$$

In the following we transpose the vectors $\widetilde{u}_1, \ldots, \widetilde{u}_{\widetilde{n}}$ and define them as row vectors of a matrix $\widetilde{W}^- \in \mathbb{R}^{\widetilde{n} \times N}$.

The estimates $\widetilde{x}_i(t)$ of the amplitudes $x_i(t)$ of (2) are given by

$$\widetilde{x}_i(t) = q(t)^\top \widetilde{u}_i \tag{34}$$

or in matrix notation

$$\widetilde{X} = \widetilde{W}^- Q \in \mathbb{R}^{\widetilde{n}\times T}. \tag{35}$$

To estimate the DyCA components w_i of (2), we compute a right inverse $\widetilde{W} \in \mathbb{R}^{N\times\widetilde{n}}$ of $\widetilde{W}^-$ such that the data matrix Q (or resp. the signal $q(t)$) is represented best in the Euclidean norm. This leads to solving again a least squares problem

$$\underset{\widetilde{w}_1,\ldots,\widetilde{w}_{\widetilde{n}}\in\mathbb{R}^N}{\arg\min} \ \langle \| q(t) - \sum_{i=1}^{\widetilde{n}} \widetilde{x}_i(t)\widetilde{w}_i \|^2 \rangle \tag{36}$$

that reads

$$\underset{\widetilde{W}\in\mathbb{R}^{N\times\widetilde{n}}}{\arg\min} \|Q - \widetilde{W}\widetilde{X}\|_F^2 \tag{37}$$

in matrix notation with $\|\cdot\|_F$ denoting the Frobenius norm. Defining $C_{\widetilde{X}} := \frac{1}{T}\widetilde{X}\widetilde{X}^\top \in \mathbb{R}^{\widetilde{n}\times\widetilde{n}}$,

$$\widetilde{W} = \frac{1}{T} Q\widetilde{X}^\top C_{\widetilde{X}}^{-1} \tag{38}$$

is a solution of (37) that reads

$$\widetilde{w}_i = \sum_{j=1}^{\widetilde{n}} \left(C_{\widetilde{X}}^{-1}\right)_{ij} \langle \widetilde{x}_j(t) q(t)\rangle \tag{39}$$

in vector notation (see theorem IV.7). Please note that the inverse $C_{\widetilde{X}}^{-1}$ exists since the choice of the minimal subset of $\widetilde{v}_i$ leads to a regular correlation matrix $C_{\widetilde{X}}$. The column vectors $\widetilde{w}_i$ of $\widetilde{W}$ are called *DyCA components*. By

$$\begin{aligned}
\widetilde{W}^-\widetilde{W} &= \frac{1}{T}\widetilde{W}^- Q\widetilde{X}^\top C_{\widetilde{X}}^{-1} \\
&= \frac{1}{T}\widetilde{X}\widetilde{X}^\top C_{\widetilde{X}}^{-1} \\
&= C_{\widetilde{X}} C_{\widetilde{X}}^{-1} \\
&= I_{\widetilde{n}}
\end{aligned}$$

one can easily see that $\widetilde{W}^-$ indeed is a left inverse of $\widetilde{W}$.

The signal $q(t)$ can then be reconstructed by

$$\widetilde{q}(t) = \sum_{i=1}^{\widetilde{n}} \widetilde{x}_i(t)\widetilde{w}_i \tag{40}$$

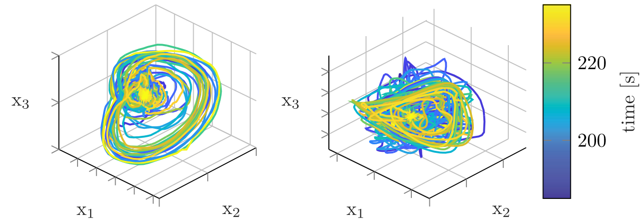
or $\widetilde{Q} = \widetilde{W}\widetilde{X}$ in matrix notation.

In addition, by solving a third least squares problem

$$\underset{\widetilde{A}\in\mathbb{R}^{\widetilde{m}\times\widetilde{n}}}{\arg\min} \|\widetilde{U}\dot{Q} - \widetilde{A}\widetilde{X}\|_F^2 \tag{41}$$

with $\widetilde{U}$ consisting of the first $\widetilde{m}$ rows of $\widetilde{W}^-$, i.e. of the vectors $\widetilde{u}_1^\top, \ldots, \widetilde{u}_{\widetilde{m}}^\top$, we obtain an estimation of the ODE parameters $a_{i,k}$ of (3) by the solution

$$\widetilde{A} = \frac{1}{T}\widetilde{U}\dot{Q}\widetilde{X}^\top C_{\widetilde{X}}^{-1}. \tag{42}$$



(a) DyCA (b) PCA

FIGURE 1. The results of DyCA and PCA on icEEG data recorded during a seizure, 3 dimensions of every subspace are drawn, color indicates time evolution.

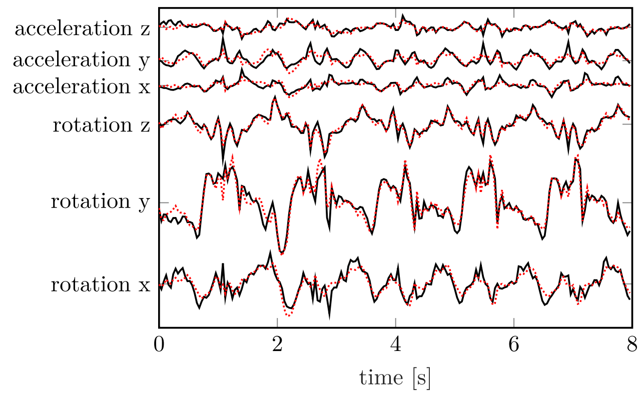


FIGURE 2. Eight second sample of motion data recorded during jogging, subject: male, weight: 102 kg, height: 1,88 m, age 46 years, reconstructed signal using DyCA (red) and raw signal (black) for comparison.

F. *DyCA ALGORITHM*

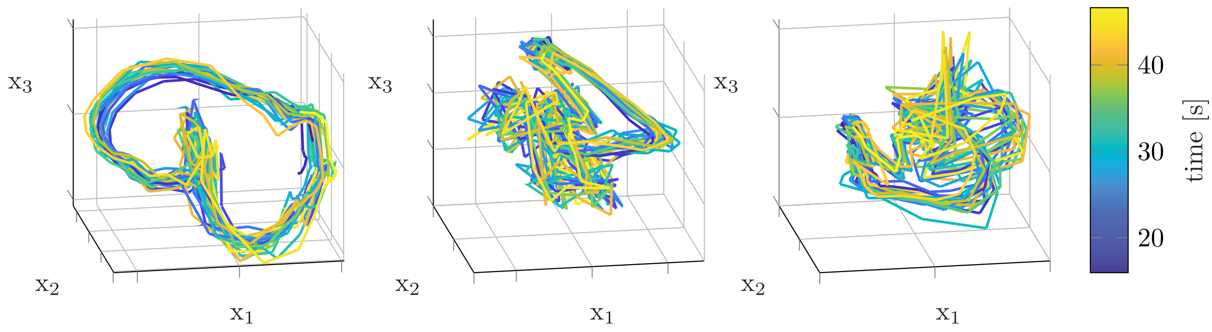
Overall we have deduced and analyzed the Algorithm 1.

III. EXAMPLE APPLICATIONS

In this section data sets from various origins are investigated using DyCA. These examples came from different scientific fields and are intended to illustrate the usefulness of DyCA as a signal processing tool. We show how the properties of DyCA described in Section II-D can be used to gain further insights into the data sets studied.

A. *icEEG-DATA*

Four intracranial EEG (icEEG) recordings of focal epileptic seizures recorded with 512 or 1024 samples per second by 111 to 165 sensors are investigated with DyCA. In all recordings we obtained two eigenvalues close to 1 ($\lambda_1 \approx 0.88$ and $\lambda_2 \approx 0.86$) and a clear structure of the trajectories during seizure. To construct a projection-matrix $\widetilde{W}^-$, the vectors $\widetilde{u}_1, \widetilde{u}_2$ and $\widetilde{v}_1$ were chosen ((31), (32) and (35)). The use of this projection allows an enormous reduction of the dimensionality from 165 to only 3 dimensions. Fig. 1 shows the trajectories for one typical dataset comparing DyCA and PCA. The structure of the signal is obviously better observable by DyCA than by PCA. We did not include ICA trajectories, since we could not find any clear structured trajectories out of the numerous possible combinations of ICA-amplitudes.



(a) DyCA (b) ICA (c) PCA

FIGURE 3. The results of different dimensionality reduction methods on motion sense data recorded during jogging, 3 dimensions of every subspace are drawn, color indicates time evolution.

The analysis of EEG-data recorded with 256 samples per second by 25 sensors during an epileptic seizure was the motivating application to introduce DyCA in the first place, see [2]. We demonstrated the application of DyCA as a suitable preprocessing tool for dimension reduction, similar to the above mentioned analysis of icEEG data.

In [3] the eigenvalues (Property II-D1) were used as a feature to classify windows of the time series as seizure or non-seizure events. The rationality behind this was motivated by [38] and [39] suggesting that epileptic seizures exhibit Shilnikov chaos. A system giving rise to Shilnikov chaos can be described by a set of ordinary differential equations (ODE), as formulated in Section II with (2) and (3) with two linear and one non-linear ODE. This assumed form of the ODEs was confirmed by the resulting eigenvalues with the two largest generalized DyCA eigenvalues during seizure were found to be close to 1.

Another example of the application of DyCA on EEG data is given in [40]. Different approaches are investigated to test multivariate EEG data for determinism by the Kaplan-Glass determinism test [41]. The study demonstrated that DyCA is an efficient way to preprocess the data for the determinism test providing evidence for deterministic chaos in certain types of epileptic seizures.

B. MOTION SENSE DATASET

In this section the application of DyCA to a motion sense dataset publicly available [42] is presented. The data consists of sensor measurements with smartphones' accelerator and gyroscope sensors during various activities. The time series is 11-dimensional and recorded at 25 samples per second. During some activities, e.g. jogging, the recorded motion can be described as being quasi-periodic (Fig. 2, black line). Therefore we assume the data can be described by a system of (non-)linear ordinary differential equations as given by (2) and (3). This hypothesis is tested by applying DyCA to the data.

As a preprocessing step the signals are high-pass filtered with a cut-off frequency of 0.5 Hz. The time series is partitioned into non overlapping windows of 8 seconds length. The projections of DyCA, PCA and ICA are computed on the first window of the time series. They are used in a second step to project all windows of the time series. The results for a 48 s long cutout (leaving out the first and last window) of the original time series are displayed in Fig. 3. Note that due to illustration limitations, only the first three dimensions of the subspace are drawn. The dimension of the subspace obtained by DyCA depends on the chosen threshold for the eigenvalues according to (31). For this example the threshold was chosen to include the three eigenvalues closest to 1. The structure of the trajectories obtained by DyCA show a more intelligible structure than those obtained by PCA and ICA.

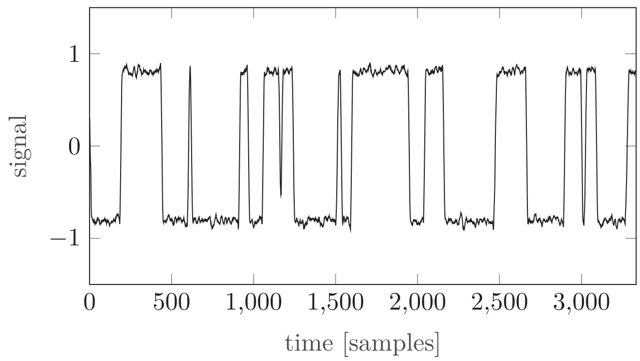
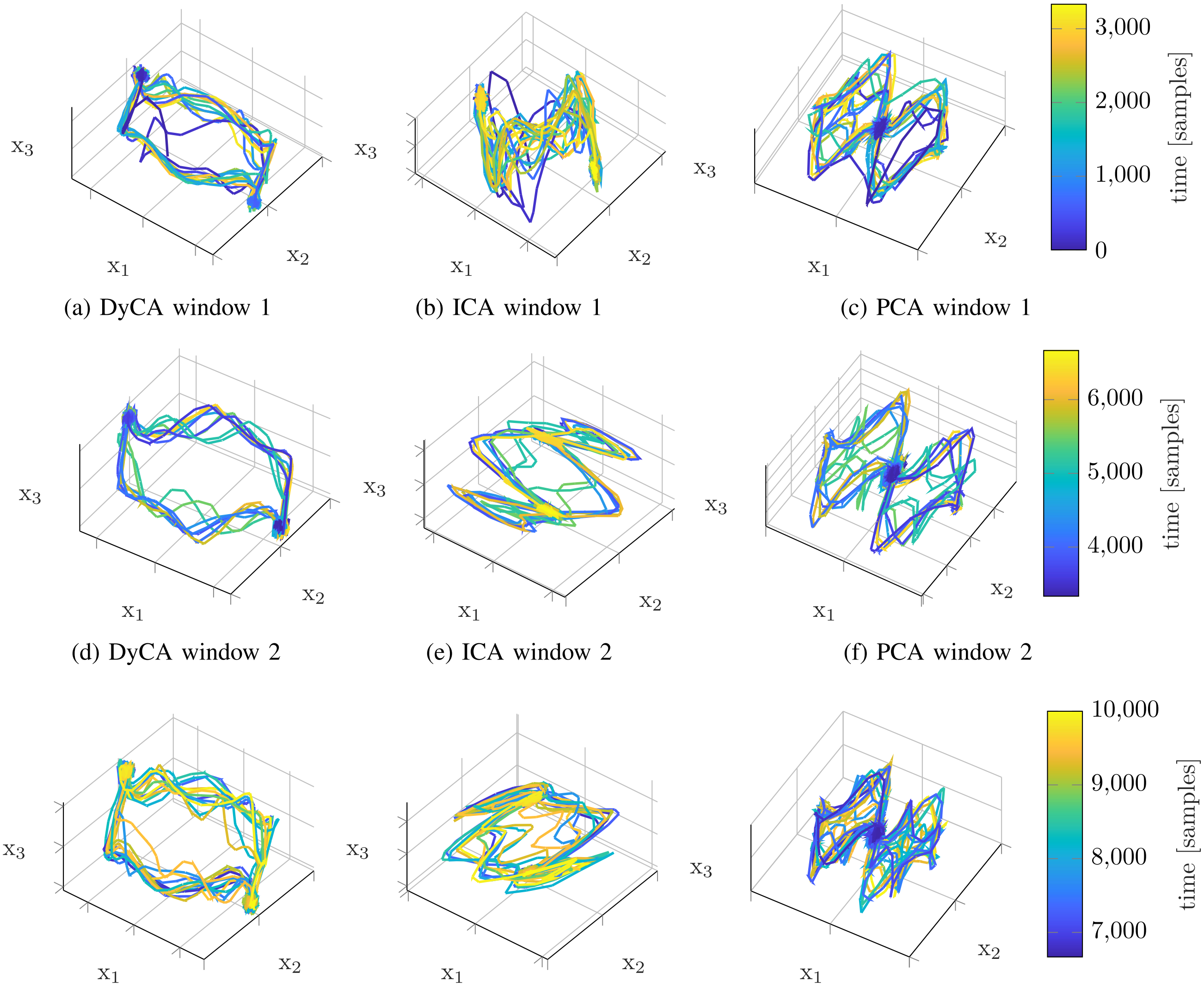


FIGURE 4. Time series generated by SDE, real part of the time series.

Using the inverse projection as described in Section II-E, a signal reconstructed from the subspace is obtained. This reconstruction can be seen for one window in Fig. 2 by the red dotted line and demonstrates a good match with the original signal (black line, relative error of reconstruction: 84%). I.e. fitting the linear part (3) of the underlying dynamics leads to a possible decomposition (1) of the multivariate data. When calculating the relative error of the reconstruction for all subjects and windows of the dataset, the reconstruction by the deterministic amplitudes reproduces on average 74% of the original data (802 windows of 8 seconds length, mean value: 74%, standard deviation: 8%).

C. EARTHS MAGNETIC FIELD REVERSAL

In this example a univariate time series generated by a model derived in [43] based on a stochastic differential equation



(a) DyCA window 1 (b) ICA window 1 (c) PCA window 1

(d) DyCA window 2 (e) ICA window 2 (f) PCA window 2

(g) DyCA window 3 (h) ICA window 3 (i) PCA window 3

FIGURE 5. DyCA, ICA and PCA trajectories of 3 successive windows of the magnetic time series, 3 dimensions of every subspace are drawn, color indicates time evolution.

(SDE) is investigated. The real-world background for this model is the irregular switching behavior of the earths magnetic field discovered by paleomagnetic studies.

A similar switching behavior could be reproduced with an experimental turbulent dynamo in [44]. Related to this experiment the model in [43] was derived. In this model the magnetic field A is assumed as the sum of two components, a dipolar D and a quadrupolar Q. The field A is then defined as $A = D + iQ$. The two modes are governed by the following differential equation:

$$\dot{A} = \mu A + \nu\bar{A} + \beta_1 A^3 + \beta_2 A^2\bar{A} + \beta_3 A\bar{A}^2 + \beta_4\bar{A}^3 + f. \tag{43}$$

The stochastic forcing term f is used to model turbulent fluctuations on the dynamo and is given by:

$$f = (br_1\zeta_1 + ibr_3\zeta_3)\,\mathrm{Re}(A) + (br_2\zeta_2 + ibr_4\zeta_4)\,\mathrm{Im}(A)$$

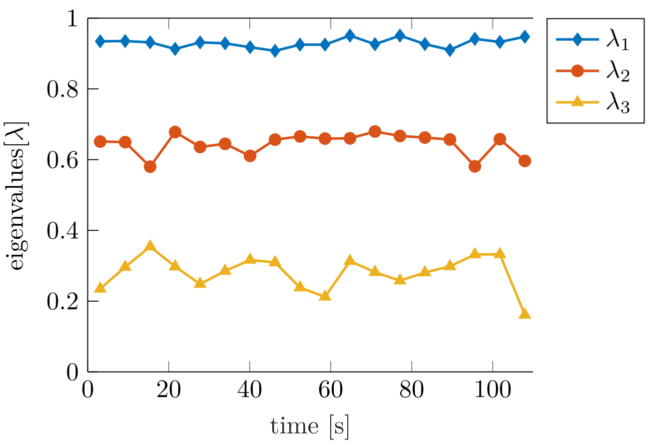
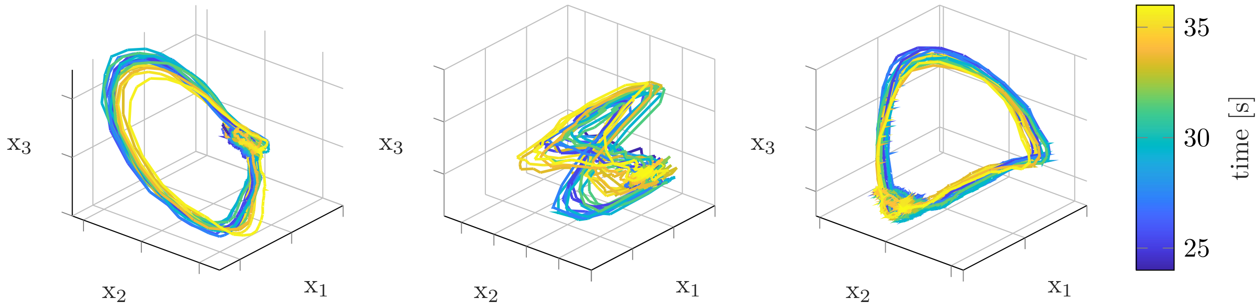


FIGURE 6. Eigenvalues of DyCA calculated on each window of 6000 samples length (12 seconds).



(a) DyCA (b) ICA (c) PCA

FIGURE 7. The results of different dimensionality reduction methods on ECG-data data recorded from healthy patients, 3 dimensions of every subspace are drawn, color indicates time evolution.

Variables ζ_i are Gaussian random variables. The other parameters are given by $\mu = 1$, $v = 0$, $\beta_1 = -0.4605i$, $\beta_2 = -1 + 0.12i$, $\beta_3 = 0.4395i$, $\beta_4 = -0.06 - 0.12i$, $br_1 = br_4 = 0.25$ and $br_2 = br_3 = 0.07$.

The simulated time series (component D) is shown in Fig. 4 with the typical stochastic switching between two states.

Takens time-delay coordinates [45] have been used to transfer the signal into a multi-variate signal and, as in the previous section, the time series is partitioned into three non overlapping windows of 3333 samples length. The projections are obtained by applying DyCA, ICA and PCA on the three windows, the results are shown in Fig. 5. For this example, the ICA algorithm used differential entropy (negentropy) as a measurement of non-gaussianity. For every window the DyCA trajectories (Fig. 5(a), 5(d), 5(g) have a clear structure in contrast to varying results obtained by ICA (Fig. 5(b), 5(e), 5(h). Furthermore DyCA does not require the selection of suitable components for each window to achieve satisfactory results.

The PCA trajectories (Fig. 5(c), 5(f), 5(i) show a different but similar structure for each window. But they do not show the transition between two different states as clearly as shown in the DyCA case. This characteristic is "better" represented by DyCA due to its aim of describing the dynamics.

D. ECG-DATA

Finally an ECG data set publicly available [46] is investigated. It contains electrocardiograms of various patients recorded at 500 samples per second with a conventional 12 lead ECG setup. For this example ECG-recordings from healthy subjects are used.

As a pre-processing step, linear trends were removed from the signal. The time series is partitioned into non overlapping windows of 6000 samples (12 seconds) length. For each of these windows the DyCA-algorithm is applied and the three largest eigenvalues are displayed in Fig. 6.

For each window we obtain one eigenvalue close to 1 and a second eigenvalue in the interval [0.55; 0.7]. This indicates a possible modeling based on a set of coupled ODEs with one linear equation. One of the first models based on a set of ODEs was developed by Zeeman [47] considering a three-dimensional set with two linear equation, and has been developed into different directions, the most prominent one by McSharry *et al.* [48], but also other interesting approaches, like an extension of the Zeeman model [49] and a model based on fractional dynamics [50]. None of these models has exactly one linear ODE and therefore do not confirm our finding at first sight. This discrepancy will be investigated in future work and might be due to a linear approximation of the non-linear equations or even the potential to model the signal with an alternative set of coupled differential equations.

The projections obtained by DyCA (Fig. 7) show a distinctive pattern, so that we conclude that the DyCA is a suitable tool for reducing the dimensionality of ECG data. The projection obtained by PCA shows a similar pattern, whereas the ICA projection looks quite different. This might be due to the choice of projection vectors, selected by visual inspection of the trajectory in phase space.

IV. CONCLUSION

Dynamical Component Analysis (DyCA), a method to decompose a multivariate signal into time-dependent amplitudes and corresponding multivariate modes (40), was presented. Thereby the amplitudes are estimated with the goal of optimally obeying a set of coupled ODEs. If the signal consists of linear independent deterministic sources and noisy components and if the underlying model of the signal has more linear than non-linear differential equations, the relevant subspace of the signal can be detected by DyCA representing a blind source separation. The theory behind the method was shown and the corresponding algorithm is presented.

The presented examples demonstrate a wide variety of possible applications of DyCA in different scientific fields:

- dimension reduction to obtain low-dimensional trajectories describing the dynamics
- the DyCA eigenvalue spectrum as a feature for classification tasks
- filtering the signal by eliminating noisy components
- embedding univariate data by delay coordinates and DyCA projections
- data driven approach to model signals.

With this broad field of possible applications we are convinced that DyCA represents a valuable alternative tool besides PCA and ICA to help researches to establish a deeper insight in their investigation of complex systems.

APPENDIX

A. BASICS AND FURTHER THEORY

In this section we provide the basic theorems that are required for the derivation of DyCA. First of all, we want to remind the reader of the definition of a generalized inverse.

Definition IV.1 (Generalized Inverse): Let $A \in \mathbb{R}^{n\times m}$. Then $A^- \in \mathbb{R}^{m\times n}$ is called *generalized inverse* of A, if the following holds:

$$AA^-A = A$$

For each matrix $A \in \mathbb{R}^{n\times m}$ there exists always a generalized inverse that is, however, not necessarily unique. If $A^-A = I_m$ holds in addition to the above definition, we call A^- a *(generalized) left inverse* of A. If on the other hand it holds that $AA^- = I_n$, then A^- is called *(generalized) right inverse* of A. If A is regular, then $A^- = A^{-1}$. Regarding the rank of A and A^- the following theorem holds:

Theorem IV.2: Let $A^- \in \mathbb{R}^{m\times n}$ be a generalized inverse of $A \in \mathbb{R}^{n\times m}$. Then:

- $\operatorname{rank}(A) = \operatorname{rank}(AA^-) = \operatorname{rank}(A^-A)$
- $\operatorname{rank}(A) \le \operatorname{rank}(A^-)$

By the help of the last theorem it immediately follows that:

Lemma IV.3: Let $n \ge m$ and let $A \in \mathbb{R}^{n\times m}$ have pairwise linearly independent column vectors a_i, $i = 1, \ldots, m$. Then the row vectors $b_i^\top$, $i = 1, \ldots, m$ of a generalized left inverse A^- of A are also linearly independent.

Proof: It is obvious that the matrix A is always of rank m due to its linearly independent columns a_i, $i = 1, \ldots, m$. By theorem IV.2, for a generalized left inverse A^- of A it holds that

$$\operatorname{rank}(A^-) \ge \operatorname{rank}(A).$$

Since $A^- \in \mathbb{R}^{m\times n}$ has the maximum rank m, it holds that $\operatorname{rank}(A^-) = m$ and hence, A^- has m linearly independent row vectors $b_i^\top$. ■

The last lemma obviously holds for the matrix W as defined in (7) as its columns w_i, $i = 1, \ldots, n$, are per definition linearly independent. Hence, the row vectors $u_i^\top$, $i = 1, \ldots, n$, of a generalized left inverse W^- are also linearly independent. We would now like to provide an explanation for the linear independence of at least $n-m$ of the v_i.

Lemma IV.4: Let $A = [A_1, A_2]$ as defined in (4) with $\operatorname{rank}(A_2) = n-m$. Then at least $n-m$ of the v_i as defined in (11) are linearly independent.

Proof: To prove the above statement we write the $v_i = \sum_{k=1}^n a_{i,k}u_k$, $i = 1, \ldots, m$ in matrix notation:

$$V = UA^\top$$

with $A \in \mathbb{R}^{m\times n}$ as in (4), $U := (W^-)^\top \in \mathbb{R}^{N\times n}$ consisting of the vectors $u_1, \ldots, u_n$, and $V \in \mathbb{R}^{N\times m}$ with the column vectors $v_1, \ldots, v_m$. Furthermore it holds that $\operatorname{rank}(U) = n$ and we additionally assume that $\operatorname{rank}(A) = n-m$. Then $\operatorname{rank}(A^\top) = \operatorname{rank}(A) = n-m$. With Sylvester's rank inequality it follows that

$$\underbrace{\operatorname{rank}(U)}_{=n} + \underbrace{\operatorname{rank}(A^\top)}_{=n-m} - n \le \operatorname{rank}(UA^\top)$$
$$\le \underbrace{\min\{\operatorname{rank}(U), \operatorname{rank}(A^\top)\}}_{=n-m}$$

thus $n-m \le \operatorname{rank}(UA^\top) \le n-m$ and hence, $\operatorname{rank}(UA^\top) = \operatorname{rank}(V) = n-m$. Therefore V has $n-m$ linearly independent column vectors v_i. The total amount of linearly independent v_i depends on the actual rank of A that is minimum $n-m$ and maximum m. Indeed, it holds that if $\operatorname{rank}(A) = j$, $j = n-m, \ldots, m$, then j of the v_i are linearly independent which immediately follows by Sylvester's rank inequality as well.

The next lemmas state the prerequisites for exchanging $u_{m+1}, \ldots, u_n$ with suitable v_i in (30).

Lemma IV.5: Let Y be a $\mathbb{K}$-vector space, let $y_1, \ldots, y_n \in Y$ as well as $z = \sum_{i=1}^n \lambda_i y_i \in Y$ with $\lambda_1 \ne 0$. Then $\operatorname{span}\{y_1, \ldots, y_n\} = \operatorname{span}\{z, y_2, \ldots, y_n\}$.

Lemma IV.6 (Steinitz exchange lemma): Let $Z = \{z_1, \ldots, z_m\}$ and $Y = \{y_1, \ldots, y_n\}$ be two finite subsets of a $\mathbb{K}$-vector space and let $z_1, \ldots, z_m$ be linearly independent. If $Z \subseteq \operatorname{span}\{y_1, \ldots, y_n\}$, it holds that $m \le n$ and m elements of Y can be exchanged for the elements of Z with suitable numbering $y_1, \ldots, y_m$ in such a way that:

$$\operatorname{span}\{z_1, \ldots, z_m, y_{m+1}, \ldots, y_n\} = \operatorname{span}\{y_1, \ldots, y_n\}$$

In order to explain the solution of the least squares problems (37) and (41) we briefly recall the theory of overdetermined linear systems of equations and linear regression. Considering a matrix $A \in \mathbb{R}^{n\times m}$ and a vector $b \in \mathbb{R}^n$, a linear systems of equations

$$Ax = b$$

does in general not have a solution if it is overdetermined, i.e. if $n > m$. Therefore one tries to find a vector $x \in \mathbb{R}^m$ by means of *linear regression,* such that $Ax = b$ is still fulfilled as closely as possible. For this we write

$$Ax \approx b$$

and intend to determine the unknown $x \in \mathbb{R}^m$ in such a way that the vector $Ax \in \mathbb{R}^n$ has the smallest possible distance from the vector $b \in \mathbb{R}^n$ in the Euclidean norm. Hence, we want to find an $x \in \mathbb{R}^m$ that solves

$$\min_{x\in\mathbb{R}^m} \|b - Ax\|. \tag{44}$$

The following theorems answer the questions about existence and uniqueness of a solution of the linear regression problem (44).

Theorem IV.7: A vector $\hat{x} \in \mathbb{R}^m$ is a solution of the linear regression problem (44), if and only if it suffices the *(Gaussian) normal equation*

$$A^\top Ax = A^\top b \tag{45}$$

Theorem IV.8: The linear regression problem (44) always has a solution. The solution is unique if $rank(A) = m$.

Problem (37) actually intends to solve the linear system of equations $Q = \widetilde{W}\widetilde{X}$ that can be considered to be overdetermined by regarding its transpose:

$$Q^\top = \widetilde{X}^\top \widetilde{W}^\top$$

According to theorem IV.7 we multiply this equation by $\frac{1}{T}\widetilde{X}$ yielding

$$\frac{1}{T}\widetilde{X}Q^\top = \frac{1}{T}\widetilde{X}\widetilde{X}^\top \widetilde{W}^\top.$$

As $C_{\widetilde{X}} := \frac{1}{T}\widetilde{X}\widetilde{X}^\top$ is invertible, we obtain $\widetilde{W}^\top = \frac{1}{T}C_{\widetilde{X}}^{-1}\widetilde{X}Q^\top$. Transposing a second time yields

$$\widetilde{W} = \frac{1}{T}Q\widetilde{X}^\top C_{\widetilde{X}}^{-1}.$$

It can be shown by Sylvester's rank inequality that $rank\ (\widetilde{X}^\top) = \widetilde{n}$ and according to theorem IV.8 the solution is unique. Problem (41) can be solved analogously.

ACKNOWLEDGMENT

The authors would like to thank Knut Hüper for fruitful discussions and the two anonymous reviewers for their comments and valuable hints on related work.

REFERENCES

[1] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019.

[2] B. Seifert, K. Korn, S. Hartmann, and C. Uhl, "Dynamical component analysis (DyCA): Dimensionality reduction for high-dimensional deterministic time-series," in *Proc. IEEE 28th Int. Workshop Mach. Learn. Signal Process.*, 2018, pp. 1–6.

[3] K. Korn, B. Seifert, and C. Uhl, "Dynamical component analysis (DyCA) and its application on epileptic EEG," in *ICASSP Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019.

[4] B. Seifert, D. Adamski, and C. Uhl, "Analytic quantification of Shilnikov chaos in epileptic EEG data," *Front. Appl. Math. Statist.*, vol. 4, 2018, Art. no. 559.

[5] K. Hasselmann, "PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns," *J. Geophys. Res.*, vol. 93, no. D9, 1988, Art. no. 11015.

[6] F. Kwasniok, "The reduction of complex dynamical systems using principal interaction patterns," *Physica D: Nonlinear Phenomena*, vol. 92, nos. 1/2, pp. 28–60, 1996.

[7] F. Kwasniok, "Reduced atmospheric models using dynamically motivated basis functions," *J. Atmospheric Sci.*, vol. 64, no. 10, pp. 3452–3474, 2007.

[8] K. Pearson, "LIII. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901.

[9] I. T. Jolliffe, *Principal Component Analysis*, New York, NY: Berlin, Germany; Springer New York, 1986.

[10] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson, "Empirical orthogonal functions and related techniques in atmospheric science: A review," *Int. J. Climatol.*, vol. 27, no. 9, pp. 1119–1152, 2007.

[11] C. A. L. Pires and A. Hannachi, "Independent subspace analysis of the sea surface temperature variability: Non-gaussian sources and sensitivity to sampling and dimensionality," *Complexity*, vol. 2017, nos. 2/3, pp. 1–23, 2017.

[12] A. Gavrilov, A. Seleznev, D. Mukhin, E. Loskutov, A. Feigin, and J. Kurths, "Linear dynamical modes as new variables for data-driven ENSO forecast," *Climate Dyn.*, vol. 52, nos. 3/4, pp. 2199–2216, 2019.

[13] D. Mukhin, A. Gavrilov, A. Feigin, E. Loskutov, and J. Kurths, "Principal nonlinear dynamical modes of climate variability," *Sci. Rep.*, vol. 5, 2015, Art. no. 15510.

[14] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 30, no. 1, pp. 179–196, 1995.

[15] E. Vanhatalo, M. Kulahci, and B. Bergquist, "On the structure of dynamic principal component analysis used in statistical process monitoring," *Chemometrics Intell. Lab. Syst.*, vol. 167, pp. 1–11, 2017.

[16] Y. Dong and S. J. Qin, "A novel dynamic PCA algorithm for dynamic data modeling and process monitoring," *J. Process Control*, vol. 67, pp. 1–11, 2018.

[17] W. Li, H. Yue, S. Valle-Cervantes, and S. Qin, "Recursive PCA for adaptive process monitoring," *J. Process Control*, vol. 10, no. 5, pp. 471–486, 2000.

[18] I. Portnoy, K. Melendez, H. Pinzon, and M. Sanjuan, "An improved weighted recursive PCA algorithm for adaptive fault detection," *Control Eng. Pract.*, vol. 50, pp. 69–83, 2016.

[19] J. He, L. Balzano, and J. C. S. Lui, "Online robust subspace tracking from partial information," Sep. 2011, *arXiv:1109.3827v2*.

[20] P. Narayanamurthy and N. Vaswani, "Provable dynamic robust PCA or robust subspace tracking," in *Proc. IEEE Int. Symp. Inf. Theory*, 2018, pp. 376–380.

[21] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jun. 2018.

[22] P. Narayanamurthy, V. Daneshpajooh, and N. Vaswani, "Provable subspace tracking from missing data and matrix completion," *IEEE Trans. Signal Process.*, vol. 67, no. 16, pp. 4245–4260, Jun. 2019.

[23] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4/5, pp. 411–430, 2000.

[24] H. Shen and K. Huper, "Generalised FastICA for independent subspace analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. ICASSP '07*, vol. 4, 2007, pp. IV–1409–IV–1412.

[25] A. Hyvärinen, "Independent component analysis: Recent advances," *Philosophical Trans. Ser. A, Math. Phys. Eng. Sci.*, vol. 371, no. 1984, Feb. 2013, Art. no. 20110534.

[26] A. Hyvärinen, J. Karhunen, E. Oja, and S. Haykin, *Independent Component Analysis*, New York, NY, USA: Wiley, 2001.

[27] S. Makeig *et al.*, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, pp. 690–694, 2002.

[28] B. Hunyadi *et al.*, "ICA extracts epileptic sources from fMRI in EEG-negative patients: A retrospective validation study," *PLoS ONE*, vol. 8, no. 11, 2013, Art no. e78796.

[29] A. Bhatnagar, K. Gupta, U. Pandharkar, R. Manthalkar, and N. Jadhav, "Comparative analysis of ICA, PCA-based EASI and wavelet-based unsupervised denoising for EEG signals," in *Computing, Communication Signal Processing*, ser. Adv. Intell. Syst. Comput., B. Iyer, S. Nalbalwar, and N. P. Pathak, Eds., Singapore: Springer Singapore, 2019, vol. 810, pp. 749–759.

[30] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *J. Fluid Mech.*, vol. 656, pp. 5–28, 2010.

[31] G. M. Goerg, "Forecastable component analysis," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. II–64–II–72.

[32] N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proc. Roy. Soc. A: Math., Phys. Eng. Sci.*, vol. 466, no. 2117, pp. 1291–1302, 2010.

[33] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[34] N. u. Rehman and H. Aftab, "Multivariate variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 67, no. 23, pp. 6039–6052, Dec. 2019.

[35] L. Stankovic, M. Brajović, M. Dakovic, and D. Mandic, "On the decomposition of multichannel nonstationary multicomponent signals," *Signal Process.*, vol. 167, Aug. 2019, Art. no. 107261.

[36] B. Thirion and O. Faugeras, "Dynamical components analysis of fMRI data through kernel PCA," *NeuroImage*, vol. 20, no. 1, pp. 34–49, 2003.

[37] D. Clark, J. A. Livezey, and K. E. Bouchard, "Unsupervised discovery of temporal structure in noisy data with dynamical component analysis," *Adv. Neural Inf. Process. Syst.*, 2019, pp. 14 267–14 278.

[38] R. Friedrich and C. Uhl, "Spatio-temporal analysis of human electroencephalograms: Petit-mal epilepsy," *Physica D: Nonlinear Phenomena*, vol. 98, no. 1, pp. 171–182, 1996.
[39] L. van Veen and D. T. J. Liley, "Chaos via Shilnikov's saddle-node bifurcation in a theory of the electroencephalogram," *Phys. Rev. Lett.*, vol. 97, no. 20, 2006, Art. no. 208101.
[40] C. Frühauf, S. Hartmann, B. Seifert, and C. Uhl, "Determinism testing of low-dimensional signals embedded in high-dimensional multivariate time series," in *Chaos and Complex Systems*, S. G. Stavrinides and M. Ozer, Eds., Berlin, Germany: Springer, 2020, vol. 68, pp. 3–14.
[41] Kaplan and Glass, "Direct test for determinism in a time series," *Phys. Rev. Lett.*, vol. 68, no. 4, pp. 427–430, 1992.
[42] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi, "Mobile sensor data anonymization," in *Proceedings of the International Conference on Internet of Things Design and Implementation*, O. Landsiedel and K. Nahrstedt, Eds. New York, NY, USA: ACM, Apr. 2019, pp. 49–58.
[43] F. Pétrélis and S. Fauve, "Chaotic dynamics of the magnetic field generated by dynamo action in a turbulent flow," *J. Phys. Condens. Matter*, vol. 20, no. 49, 2008, Art. no. 494203.
[44] M. Berhanu *et al.*, "Magnetic field reversals in an experimental turbulent dynamo," *Europhys. Lett.*, vol. 77, no. 5, 2007, Art. no. 59001.
[45] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L.-S. Young, Eds. Berlin, Heidelberg: Springer, 1981, vol. 898, pp. 366–381.
[46] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet," *Biomedizinische Technik*, Band 40, Ergänzungsband 1, S 317, 1995
[47] E. C. Zeeman, "Differential equations for the heartbeat and nerve impulse," in *Dynamical Systems*, Amsterdam, The Netherlands; NY, USA: Elsevier, 1973, pp. 683–741.
[48] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Bio-Medical Eng.*, vol. 50, no. 3, pp. 289–294, Mar. 2003.
[49] A. Ayatollahi, N. J. Dabanloo, D. C. McLernon, V. J. Majd, and H. Zhang, "A comprehensive model using modified zeeman model for generating ECG signals," *Iranian J. Elect. Electron. Eng.*, vol. 1, no. 2, pp. 88–93, Apr. 2005.
[50] S. Das and K. Maharatna, "Fractional dynamical model for the generation of ECG like signals from filtered coupled van-der pol oscillators," *Comput. Methods Program. Biomed.*, vol. 112, no. 3, pp. 490–507, Dec. 2013.

CHRISTIAN UHL received the M.Sc. degree in physics from the University of Oregon, Eugene, OR, USA, in 1989, from the University of Stuttgart, Germany, in 1992 and the Ph.D. degree from the Institute of Theoretical Physics and Synergetics, University of Stuttgart, Germany, in 1995.

He was a Postdoctoral Researcher with the Max Planck Institute of Cognitive NeuroScience, Leipzig, Germany and as a Research Scientist with the Philips Research Laboratories, Aachen, Germany. He is a Professor of computer science and mathematics and the Head of the Center for Signal Analysis of Complex Systems (CCS), with the University of Applied Sciences, Ansbach, Germany. His research interests include self-organization, pattern formation and detection, dynamical systems, and multivariate signal processing.

MORITZ KERN received the B.Eng. degree in biomedical engineering from the University of Applied Sciences, Ansbach, in 2019, during his work as a Research Assistant with the Center for Signal Analysis of Complex Systems (CCS). He is currently working toward the M.Sc. degree in applied computer science. His research interests include application and implementation of signal processing approaches in various scientific fields, especially time series analysis with respect to dynamic characteristics in biomedical applications.

MONIKA WARMUTH received the B.Sc. degree in industrial mathematics from the University of Applied Sciences Würzburg-Schweinfurt, Germany, in 2017 and the M.Sc. degree in computational mathematics from Julius-Maximilians-Universität Würzburg, Germany, in 2019. She is currently working toward the Ph.D. degree in mathematics with Julius-Maximilians-Universität Würzburg. She is a Research Associate with the Center for Signal Analysis of Complex Systems (CCS), the University of Applied Sciences, Ansbach, Germany. Her research interests include dimensionality reduction of multivariate time series, as well as noise removal and data reconstruction of high-dimensional dynamical systems.

BASTIAN SEIFERT (Member, IEEE) received the B.Sc. degree in mathematics from Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, in 2013, the M.Sc. and Ph.D. degrees in mathematics from Julius-Maximilians-Universität Würzburg, Germany, in 2015 and 2020, respectively. He was a Research Associate with the Center for Signal Analysis of Complex Systems (CCS), the University of Applied Sciences, Ansbach, Germany. Currently, he is a Postdoc with the Eidgenössische Technische Hochschule Zürich (ETH), Zürich, Switzerland. His research interests include algebraic signal processing, dimensionality reduction, and applied mathematics.