Robust Dynamical Component Analysis via Multivariate Variational Denoising

Monika Warmuth
Faculty of Engineering Sciences
Ansbach University of Applied Sciences
Ansbach, Germany
Monika.Warmuth@hs-ansbach.de

Philipp Romberger
Faculty of Engineering Sciences
Ansbach University of Applied Sciences
Ansbach, Germany
Philipp.Romberger@hs-ansbach.de

Christian Uhl
Faculty of Engineering Sciences
Ansbach University of Applied Sciences
Ansbach, Germany
Christian.Uhl@hs-ansbach.de

***Abstract*—This paper combines a variational denoising approach with a dimensionality reduction of multivariate time series by dynamical component analysis (DyCA), a dimensionality reduction method for high-dimensional dynamical signals governed by a low-dimensional system of ordinary differential equations (ODEs). While DyCA has been successfully applied to these types of signals in the past, it has not been robust to noise or incomplete data. The proposed approach simultaneously denoises the given time series and reduces the dimensionality of the data using DyCA while retaining the most important dynamic structures.**

***Index Terms*—multivariate signal analysis, dimensionality reduction, dynamical component analysis, variational denoising, Roessler attractor, EEG**

I. Introduction

In the processing of real-world signals, one often encounters the problem of data being contaminated with additive Gaussian noise or missing entries in the data. In multivariate signal processing, this can occur when sensors are defective or when, due to time constraints, fewer data points can be sampled during the measurement than would actually be necessary for the evaluation of the data. The task of reconstructing the original multivariate signal from data with missing entries or noisy data is further compounded by another challenge common in multivariate signal processing: dimensionality reduction.

In this paper, we focus on multivariate signals exhibiting a certain kind of underlying deterministic dynamics which can be described by a system of linear and nonlinear ordinary differential equations. It has been recognized that the recently introduced dimensionality reduction method dynamical component analysis (DyCA) is well suited to handle the before-mentioned type of multivariate signals [1], [2], [3]. In this context, it also outperforms conventional dimensionality reduction methods such as principal component analysis (PCA) [4] and independent component analysis (ICA) [5], [6]. DyCA finds the most important underlying dynamic structures with respect to the linear part of the system of ordinary differential equations (ODEs) by projecting the data onto a low-dimensional subspace and solving a least-squares problem that is equivalent to a generalized eigenvalue problem.

This work was supported by the German Federal Ministry of Education and Research (BMBF, Funding number: 05M20WBA)

Although DyCA can reduce the dimension of deterministic dynamic data, the method is not robust to even low levels of noise or missing data entries, which has recently been shown in [7]. Common multivariate denoising methods are e.g. multivariate wavelet denoising [8], [9] or methods based on (multivariate) empirical and variational mode decomposition [10], [11], [12] that can be understood as multivariate extensions of their corresponding univariate version. In contrast to the above-mentioned methods working in the frequency domain, our approach is based on the time domain. In this paper we combine a variational regularized L^2-denoising and reconstruction problem that is well-known in signal and image processing with the least-squares DyCA minimization problem for dimensionality reduction. This results in a coordinate-descent like approach that simultaneously reconstructs and denoises the data in one step and computes DyCA projection vectors in a second step. The advantage of this approach is the mutual improvement of the two procedures. On the one hand, DyCA benefits from non-noisy and complete data. On the other hand, the signal reconstruction can be improved by the consideration of the underlying dynamics, coupling the searched unknowns through the system of ODEs. The idea is inspired by [13], where the authors integrated an optical flow term into an image reconstruction problem.

The paper is structured as follows. In sections II and III we provide the basic theory of dynamical component analysis and derive a joint variational signal denoising and dimensionality reduction approach using DyCA. In section IV the proposed method is tested on simulated Roessler attractor data and real world epileptic EEG data and compared with other common denoising methods. The results will be discussed in section V.

II. Dynamical Component Analysis (DyCA)

In this section we will provide a brief introduction to dynamical component analysis which is discussed in detail in [1]. Let $q(t) \in \mathbb{R}^N$ be a multivariate time series with $t = t_1, t_2, \ldots, t_T$, $T \geq N$. It is assumed that, under

ideal circumstances, the signal $q(t)$ can be decomposed into components

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

with linearly independent vectors $w_i \in \mathbb{R}^N$ and $n \leq N$, whereas the deterministic amplitudes $x_i(t)$ obey a set of linear and nonlinear ordinary differential equations

$$\dot{x}_i(t) = \sum_{k=1}^{n} a_{i,k} x_k(t) \qquad \text{for } i = 1, ..., m \tag{2}$$

$$\dot{x}_i(t) = f_i(x_1(t), ..., x_n(t)) \quad \text{for } i = m+1, ..., n \tag{3}$$

with $m \geq \frac{n}{2}$. In terms of matrix notation we can rewrite (1) as

$$Q = WX, \tag{4}$$

with $Q \in \mathbb{R}^{N \times T}$, $W \in \mathbb{R}^{N \times n}$, and $X \in \mathbb{R}^{n \times T}$. As an additional condition we must assume that the data matrix Q and its time derivative $\dot{Q}$ are of full rank N. This is usually the case, as in practice the data often exhibits at least small amounts of noise. DyCA now aims at computing a generalized left inverse $W^- \in \mathbb{R}^{n \times N}$ of W, i.e. $W^- W = I_n$, such that

$$X = W^- Q. \tag{5}$$

The assumption of linearly independent vectors w_i leads directly to the fact that the rows $\{u_1^\top, ..., u_n^\top\}$ of W^- are linearly independent as well. Hence, the amplitudes $x_i(t)$ can be calculated by the scalar product

$$x_i(t) = u_i^\top q(t) = q(t)^\top u_i \tag{6}$$

with time derivative $\dot{x}_i(t) = \dot{q}(t)^\top u_i$. DyCA estimates these vectors u_i first and approximates the corresponding modes w_i in a second step. Therefore we define $v_i := \sum_{k=1}^{n} a_{i,k} u_k$ for $i = 1, ..., m$ and write (2) as follows:

$$\dot{q}(t)^\top u_i = \sum_{k=1}^{n} a_{i,k} q(t)^\top u_k = q(t)^\top v_i \tag{7}$$

To obtain estimations for the vectors u_i, v_i we define the error functions

$$D(u_i, v_i) := \frac{\frac{1}{T} \sum_{j=1}^{T} \|\dot{q}(t_j)^\top u_i - q(t_j)^\top v_i\|^2}{\frac{1}{T} \sum_{j=1}^{T} \|\dot{q}(t_j)^\top u_i\|^2} \tag{8}$$

and compute approximation vectors $\widetilde{u}_i$, $\widetilde{v}_i$, $i = 1, ..., m$ as a solution of the least squares problem $\arg\min_{u_i, v_i} \sum_{i=1}^{m} D(u_i, v_i)$ subject to $u_1, ..., u_m$ being linearly independent. After some basic computations which can be found in [1] and defining $C_0 := \frac{1}{T} Q Q^\top$, $C_1 := \frac{1}{T} \dot{Q} Q^\top$, and $C_2 := \frac{1}{T} \dot{Q} \dot{Q}^\top$, (8) reduces to the solution of a generalized eigenvalue problem

$$C_1 C_0^{-1} C_1^\top u = \lambda C_2 u \tag{9}$$

with $\widetilde{v}_i = C_0^{-1} C_1^\top \widetilde{u}_i$ for the corresponding eigenvectors $\widetilde{u}_i$. For the generalized eigenvalues λ_i we obtain

$$\min_{u_i, v_i \in \mathbb{R}^N} D(u_i, v_i) = 1 - \lambda_i \tag{10}$$

i.e. the linear approximation of (8) is well suited for $\lambda_i \approx 1$. Choosing a minimal subset of vectors $\widetilde{u}_i$, $\widetilde{v}_i$ such that $\widetilde{n} = \dim(\mathrm{span}\{\widetilde{u}_1, ..., \widetilde{u}_{\widetilde{m}}, \widetilde{v}_1, ..., \widetilde{v}_{\widetilde{m}}\})$ then yields an estimate $\widetilde{W}^-$ for the matrix W^-. An estimate $\widetilde{W}$ of W can be computed as $\widetilde{W} = \frac{1}{T} Q \widetilde{X}^\top C_{\widetilde{X}}^{-1}$ with $C_{\widetilde{X}} = \frac{1}{T} \widetilde{X} \widetilde{X}^\top$. All in all, the complete DyCA algorithm can be found in [1]. Taking a closer look at equation (2) at this point, one might notice that applying a linear transformation to the linear part of the ODE system keeps the structure of the data invariant. This degree of freedom leads to an ambiguity in the choice of the estimates for the matrix W and the amplitudes $x_i(t)$, resulting in rotated DyCA trajectories compared to the original data.

III. A Joint Signal Denoising and Dimensionality Reduction Model

Dynamical component analysis as derived in the previous section is not robust to additive noise so far, working only for low noise levels around 35dB [7]. To provide a solution to this matter we revisit the idea of [13], and consider a measured signal $s(t)$ with

$$s(t) = \sum_{i=1}^{n} x_i(t) w_i + n(t) \tag{11}$$

where we assume $n(t)$ to be Gaussian random noise and seek to reconstruct the original signal $q(t) = \sum_{i=1}^{n} x_i(t) w_i$ out of

$$s(t) = q(t) + n(t). \tag{12}$$

For this purpose, we expand a standard variational regularized L^2-denoising problem by additional DyCA minimization terms and therefore solve a minimization problem of the general form

$$\begin{aligned} \underset{q, u_i, v_i \in \mathbb{R}^N}{\arg\min} \frac{1}{T} \sum_{j=1}^{T} \Bigg[& \frac{1}{2} \|q(t_j) - s(t_j)\|_2^2 + \frac{\beta}{2} \|\partial_t q(t_j)\|_2^2 \\ & + \frac{\gamma}{2} \sum_{i=1}^{m} \|\partial_t q(t_j)^\top u_i - q(t_j)^\top v_i\|_2^2 \Bigg] \end{aligned} \tag{13}$$

subject to u_i being linearly independent and $\partial_t q$ being an alternative operator notation for the time derivative $\dot{q}$ of q. To facilitate later computations, we assumed the denominator of the DyCA minimization problem to be equal to one which is equivalent to minimizing (8). For reasons of simplicity and due to the summands of the DyCA term being all of the same nature, we neglect the index i in the following. We would like to point out that this approach is also suitable for the reconstruction of the searched signal from incomplete data.

In analogy to [13] and to simplify the later numerical implementation, we use an alternating coordinate-descent like approach for (13) changing between minimization with respect to q in one step and u and v in a second step while fixing

the other variables. The alternating minimization approach of model (13) then reads:

$$q^{k+1} = \arg\min_q \frac{1}{T}\sum_{j=1}^{T}\frac{1}{2}\|q(t_j)-s(t_j)\|_2^2 + \frac{\beta}{2}\|\partial_t q(t_j)\|_2^2 + \frac{\gamma}{2}\|(u^k)^\top\partial_t q(t_j) - (v^k)^\top q(t_j)\|_2^2$$

$$u^{k+1}, v^{k+1} = \arg\min_{u,v} \frac{\gamma}{2T}\sum_{j=1}^{T}\|u^\top\partial_t q^{k+1}(t_j) - v^\top q^{k+1}(t_j)\|_2^2$$

While the second subproblem can be solved by the standard DyCA method as introduced in section II, we seek a solution for the first subproblem by using the Chambolle-Pock algorithm [14]. In the following we shall focus on the first subproblem which we reformulate to

$$q^{k+1} = \arg\min_q F(Kq) + G(q) \tag{14}$$

with suitable operators F, K, and G. For reasons of readability we neglect the dependence of q on the time t_j. Since the data term does not contain any operator acting on q, we assign it to G and write

$$G(q) := \frac{1}{T}\sum_{j=1}^{T}\frac{1}{2}\|q-s\|_2^2. \tag{15}$$

Both of the other terms contain an operator acting on q which is the gradient or time derivative ∂_t for the regularization term and $((u^k)^\top\partial_t - (v^k)^\top)$ for the DyCA term. Thus, we write

$$F(Kq) = \frac{1}{T}\sum_{j=1}^{T}\frac{\beta}{2}\|\partial_t q\|_2^2 + \frac{\gamma}{2}\|(u^k)^\top\partial_t q - (v^k)^\top q\|_2^2 \tag{16}$$

with an underlying operator

$$Kq = \begin{pmatrix}\partial_t \\ (u^k)^\top\partial_t - (v^k)^\top\end{pmatrix} q. \tag{17}$$

The adjoint operator K^* of K can be computed as $K^*y = -\partial_t y_1 - \left((u^k)^\top\partial_t - (v^k)^\top\right)y_2$, and the convex conjugate F^* of F as $F^*(y) = \frac{1}{T}\sum_{j=1}^{T}\frac{1}{2\beta}\|y_1\|_2^2 + \frac{1}{2\gamma}\|y_2\|_2^2$. We are now able to apply the Chambolle-Pock algorithm and receive the following iteration scheme:

$$\begin{aligned}
\tilde{y}^{l+1} &= y^l + \sigma K\overline{q}^l \\
y_1^{l+1} &= \arg\min_{y_1}\left\{\frac{1}{T}\sum_{j=1}^{T}\frac{\|y_1-\tilde{y}_1^{l+1}\|^2}{2\sigma} + \frac{1}{2\beta}\|y_1\|_2^2\right\} \\
y_2^{l+1} &= \arg\min_{y_2}\left\{\frac{1}{T}\sum_{j=1}^{T}\frac{\|y_2-\tilde{y}_2^{l+1}\|^2}{2\sigma} + \frac{1}{2\gamma}\|y_2\|_2^2\right\} \\
\tilde{q}^{l+1} &= q^l - \tau K^* y^{l+1} \\
q^{l+1} &= \arg\min_q\left\{\frac{1}{T}\sum_{j=1}^{T}\frac{\|q-\tilde{q}^{l+1}\|^2}{2\tau} + \frac{1}{2}\|q-s\|_2^2\right\} \\
\overline{q}^{l+1} &= q^{l+1} + \theta(q^{l+1}-q^l)
\end{aligned}$$

The proximal mappings for y_1, y_2 can be specified directly:

$$y_1^{l+1} = \frac{\beta}{\beta+\sigma}\tilde{y}_1^{l+1}, \qquad y_2^{l+1} = \frac{\gamma}{\gamma+\sigma}\tilde{y}_2^{l+1} \tag{18}$$

The explicit form of the proximal mapping of $G(q)$ can be computed as

$$q^{l+1} = \frac{\tilde{q}^{l+1}+\tau s}{1+\tau}. \tag{19}$$

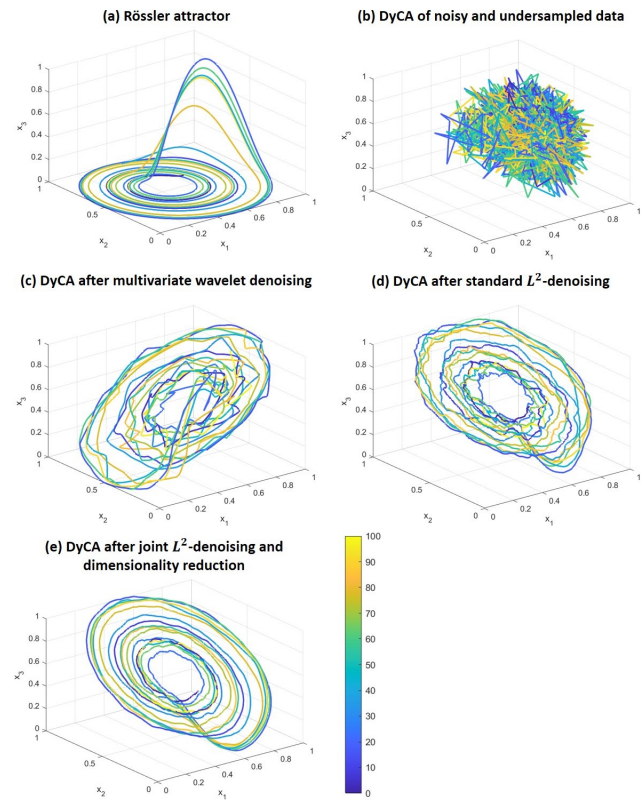

Fig. 1. DyCA trajectories computed from (a) Rössler attractor data, (b) with 1dB additive noise after application of standard DyCA, (c) DyCA after multivariate wavelet denoising, (d) DyCA after standard L^2-denoising, and (e) DyCA after the joint L^2-denoising and dimensionality reduction approach according to (13). The color bar indicates the time evolution.

IV. Application and Results

After the theoretical groundwork in the last sections, we now examine the derived method on practical examples. Here we consider both, the Rössler attractor [15] as a simulated signal and epileptic EEG data as a real world example. We compare the original data sets with the trajectories obtained after applying DyCA on the noisy and incomplete data, on multivariate wavelet denoised data [8], [9], on data obtained by a standard L^2-denoising, and with the trajectories of our joint variational approach with DyCA term as in (13).

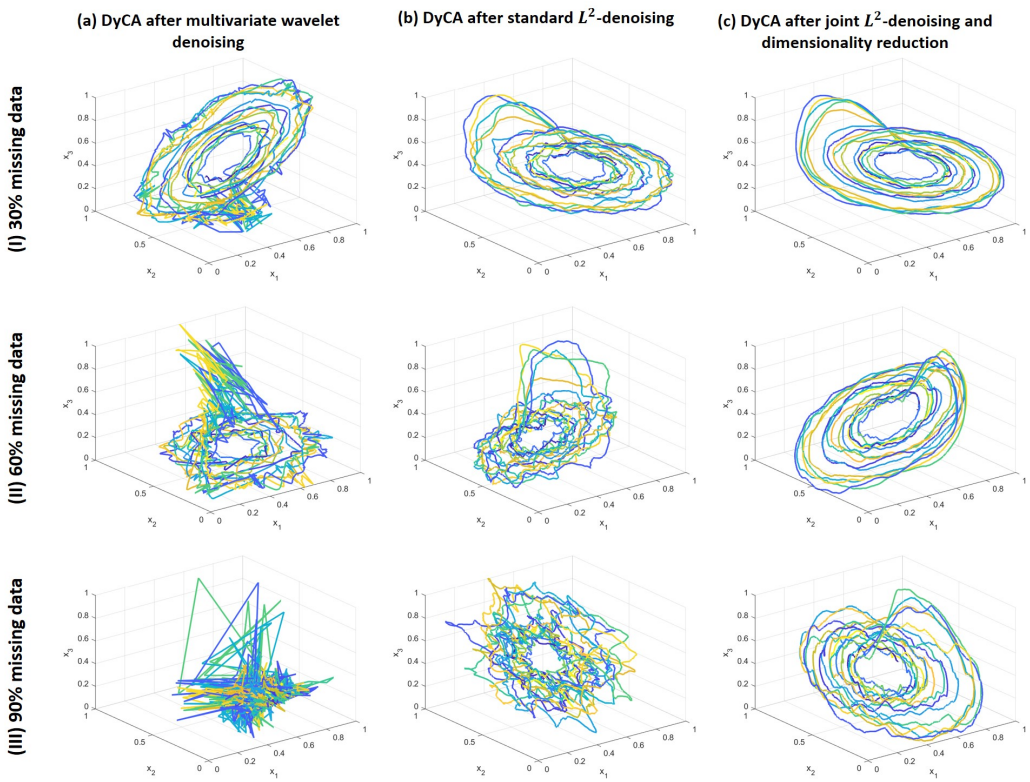

Fig. 2. DyCA Rössler trajectories computed from (a) multivariate wavelet denoised data, (b) standard L^2-denoised data, and (c) data after joint denoising and dimensionality reduction approach each with 10dB additive noise and (I) 30% missing entries, (II) 60% missing entries, and (III) 90% missing entries

A. Rössler Attractor

The well-known Rössler attractor [15] is a strange attractor defined by a system of two linear and one nonlinear ordinary differential equations

$$\begin{aligned} \dot{x}_1 &= -x_2 - x_3 \\ \dot{x}_2 &= x_1 + a x_2 \\ \dot{x}_3 &= b - c x_3 + x_1 x_2 \end{aligned} \tag{20}$$

with $a = 0.15$, $b = 0.2$, and $c = 10$. This is a synthetic signal example, where we do not simulate incompleteness of the data in a first run, but apply white noise with a signal-to-noise ratio of 1dB. Figure 1 shows the computed trajectories of the data using the first two DyCA eigenvalues λ_1, λ_2 with the corresponding DyCA projection vectors u_1, u_2 and v_1. For our tests we found that the choice of $\beta = 0.015$, $\gamma = 1$ and $\sigma = \tau = 4 \cdot 10^{-4}$ provide the best results.

By displaying the resulting trajectories, it is visible that the joint L^2-denoising and dimensionality reduction approach outperforms all the other methods. In compliance with (10) the eigenvalues of the joint approach are $\lambda_1 = 0.9998$ and $\lambda_2 = 0.9993$ while the other eigenvalues are close to zero. None of the other denoising methods exhibits eigenvalues close to one. To examine the behavior of the algorithm in case of incomplete data, we add 10dB additive noise to the Rössler attractor data and vary the incompleteness level from 30% over 60% up to 90%. Figure 2 shows that even at a moderate incompleteness level of 30%, the joint approach (13) provides smoother trajectories than the multivariate wavelet denoising and the standard L^2-denoising approach. If the degree of incompleteness is increased to 90%, only the joint approach can detect the trajectory, although it is not as smooth as before. However, keeping in mind that we work with merely 10% of the data and that the standard DyCA is solely applicable to complete data with a noise level greater than 35dB, this is can be interpreted as a satisfactory result.

B. Epileptic EEG Data

In the past, DyCA has already been successfully applied to epileptic EEG data [2], [3]. There, the authors support the suggestion of [16] and [17] that epileptic seizure EEG data exhibits Shilnikov chaos and can be described by

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3 &= f(x_1, x_2, x_3) \end{aligned} \tag{21}$$

with f being a nonlinear smooth function. The epileptic seizure data was recorded with 25 sensors. After simulating an incompleteness level of 50%, we apply the above mentioned denoising methods before computing the DyCA trajectories and projecting the data onto a 3-dimensional subspace. Figure 3 visualizes the results using the two largest eigenvalues λ_1 and λ_2 as well as the corresponding DyCA

projection vectors u_1, u_2, and v_1. For the epileptic EEG data we used the parameters $\beta = 0.000125$, $\gamma = 1$, and $\sigma = \tau = 4 \cdot 10^{-4}$.

Also in this case our joint approach achieves better results than the other denoising and reconstruction methods. In particular, the trajectory of the multivariate wavelet denoised data loses its three-dimensionality. The first two DyCA eigenvalues are given as $\lambda_1 = 1.0000$, $\lambda_2 = 1.0000$ while the eigenvalues of the other methods are $\lambda_1 = 0.7305$, $\lambda_2 = 0.6622$ for DyCA applied on the noisy data, $\lambda_1 = 0.6905$, $\lambda_2 = 0.6227$ for the multivariate wavelet denoised data, and $\lambda_1 = 0.7700$, $\lambda_2 = 0.6709$ for the L^2-regularization denoised data.

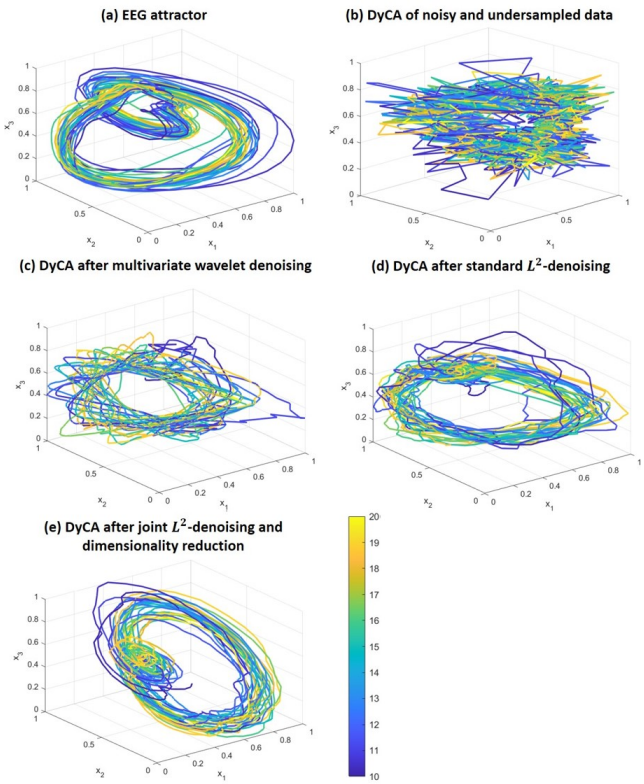

Fig. 3. DyCA trajectories computed from (a) EEG attractor, (b) data with 50% missing entries after application of standard DyCA, (c) DyCA after multivariate wavelet denoising, (d) DyCA after standard L^2-denoising, and (e) DyCA after the joint L^2-denoising and dimensionality reduction approach according to (13). The color bar indicates the time evolution.

V. Conclusion

In this paper an algorithm was derived that overcomes the problem of non-robustness of DyCA to additive noise or missing data entries. By an iterative scheme consisting of the primal-dual Chambolle-Pock algorithm [14] on the one hand and the standard dynamical component analysis algorithm [1] on the other hand, it is now possible to apply DyCA to noisy and incomplete data sets and to simultaneously reconstruct the signal and compute DyCA projection vectors. Simulated and real world data examples are discussed in this paper and show that the joint approach allows to denoise the data and reduce the dimensionality from 10 and 25 to 3 while preserving the main characteristics of the underlying dynamics. For signals that match the assumptions of DyCA, the joint model outperforms other conventional denoising methods like multivariate wavelet denoising or a standard L^2 regularization.

Future research will be done applying the presented algorithm on other data sets. Furthermore, data will be examined that are affected not only by additive noise or missing data entries, but rather by an underlying mathematical operator such as the Fourier transform or the Radon transform. In addition to the already known tasks of data denoising, data completion and dimensionality reduction, the signal must first be reconstructed from the available operator data. The approach presented in this paper can be easily expanded to such a scenario.

References

[1] C. Uhl, M. Kern, M. Warmuth, and B. Seifert, "Subspace Detection and Blind Source Separation of Multivariate Signals by Dynamical Component Analysis (DyCA)," *IEEE Open Journal of Signal Processing*, vol. 1, pp. 230–241, 2020.

[2] B. Seifert, K. Korn, S. Hartmann, and C. Uhl, "Dynamical component analysis (dyca): Dimensionality reduction for high-dimensional deterministic time-series," in *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2018, pp. 1–6.

[3] K. Korn, B. Seifert, and C. Uhl, "Dynamical component analysis (dyca) and its application on epileptic eeg," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1100–1104.

[4] K. Pearson, "On lines and planes of closest fit to a system of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 6,2, pp. 559–572, 1901.

[5] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[6] A. Hyvärinen, J. Karhunen, E. Oja, and S. Haykin, *Independent Component Analysis*, JOHN WILEY & SONS, INC, New York, USA, 2001.

[7] P. Schell, F. Anselstetter, M. Warmuth, and C. Uhl, "Applications and Limitations of Dynamical Component Analysis (DyCA)," in *ARC 2021 Applied Research Conference (accepted)*, 2021.

[8] M. Aminghafari, N. Cheze, and Poggi. J.-M., "Multivariate denoising using wavelets and principal component analysis," *Computational Statistics & Data Analysis*, vol. 50, no. 9, pp. 2381–2398, 2006.

[9] K. Naveed and N. Rehman, "Wavelet based multivariate signal denoising using Mahalanobis distance and EDF statistics," *arXiv:2005.11616*, 2020.

[10] N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 466, no. 2117, pp. 1291–1302, 2010.

[11] N. Rehman and H. Aftab, "Multivariate Variational Mode Decomposition," *IEEE Transactions on Signal Processing*, vol. 67, no. 23, pp. 6039–6052, 2019.

[12] P. Cao, H. Wang, and K. Zhou, "Multichannel Signal Denoising Using Multivariate Variational Mode Decomposition With Subspace Projection," *IEEE Access*, vol. 8, pp. 74039–74047, 2020.

[13] M. Burger, H. Dirks, and C.-B. Schönlieb, "A Variational Model for Joint Motion Estimation and Image Reconstruction," *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 94–128, 2018.

[14] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 120–145, 2011.

[15] O.E. Rössler, "An equation for continous chaos," *Physics Letters A*, vol. 57, no. 5, pp. 397–398, 1976.

[16] 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, pp. 208101, 2006.

[17] R. Friedrich and C. Uhl, "Spatio-temporal analysis of human electroencephalograms: Petit-mal epilepsy," *Physica D*, vol. 98, no. 1, pp. 171–182, 1996.