Modeling time warping in tensor decomposition

Bertrand Rivet, Jeremy E Cohen

Modeling time warping in tensor decomposition

Bertrand Rivet and Jeremy E. Cohen
Univ. Grenoble Alpes, Gipsa-Lab, F-38000 Grenoble, France
CNRS, GIPSA-Lab, F-38000 Grenoble, France

Abstract—**Taking into account subject variability in data mining is one of the great challenges of modern biomedical engineering. In EEG recordings, the assumption that time sources are exactly shared by multiple subjects, multiple recordings of the same subject, or even multiples instances of the sources in one recording is especially wrong. In this paper, we propose to deal with shared underlying sources expressed through time warping in multiple EEG recordings, in the context of ocular artifact removal. Diffeomorphisms are used to model the time warping operators. We derive an algorithm that extracts all sources and diffeomorphism in the model and show successful simulations, giving a proof of concept that subject variability can be tackled with tensor modeling.**

I. INTRODUCTION

Electroencephalographic (EEG) recordings of a single subject's neural activity are data sets with two dimensions, namely time and sensors. A wide variety of multilinear methods have been developed in the recent years that try to extract meaningful sources from these data that describe patterns of neurons or systematic responses to certain stimuli. In particular, to perform source separation in a blind context where no a priori knowledge on the cortex nor on the sources is available, tensor decomposition techniques applied on tensorization of two way arrays or multiple recordings of the same subject have proven a powerful tool in EEG [1]. Indeed the approximate multilinearity assumption made in tensor decompositions seems to fit with underlying complex physical models of brain electrical activity. However a major obstacle to source separation in EEG is the poor signal to noise ratio due to numerous artifacts in the signal. Notably, eye movement is known to generate a signal more powerful than the signal emitted by sources of interest. In this work, source separation is used to estimate these artifacts in the EEG recordings.

For signal generated by eye-movements, it is assumed that the signal $x_{ik}(t)$ recorded by the ith sensor at the kth trial is the summation of R source signals $b_r(t)$ up to scaling factors a_{ir} that are independent of the trial. However, this is not necessarily the case since temporal changes can occur due to time shift inherent to differences in speed and amplitude of saccades. The same issue of variability between several measurements of the same phenomenon can be encountered when quantifying the concentration of a chemical substance by measuring a chromatographic signal. In this case, temporal responses of certain chemicals may have various shapes depending on temperature and injection of the mixture which is done by hand. Another example is audio data, where the Doppler effect can cause time-warping.

In this paper, we tackle the problem of tensor factorization when the underlying profiles change shape from trial to trial. In this case, the canonical polyadic (CP) decomposition model is no longer valid and leads to poor estimation results [2]. We introduce a registered CP (rCP) decomposition model which assumes that the second mode is not only shifted from trial to trial as in [3] but can be any time warping modeled by an orientation-preserving diffeomorphism.

The paper is structured as follows. In Section II, the used notations are introduces. Section III briefly recalls the CP decomposition and related methods to deal with profiles change in one mode. The proposed registered CP decomposition is derived in Section IV, next we validate the usefulness of our rCP decomposition (Section V). Finally, Section VI concludes this paper.

II. NOTATIONS

In the following, $\mathcal{X} \in \mathbb{R}^{M \times L \times K}$ denotes a third order tensor (here a three way array) whose (m, l, k)th entry is x_{mlk}. The kth matrix slice of $\mathcal{X}$ is denoted by $\mathbf{X}_k = \mathcal{X}(:, :, k)$ (*i.e.* it contains all x_{mlk} by keeping k constant). The column vector $\mathbf{y}_r$ is the rth column of the matrix $\mathbf{Y}$ and y_r is the rth element of $\mathbf{y}$. The transposition operator is denoted by $\cdot'$, the pseudo-inverse matrix is denoted by $\cdot^\dagger$ and $\otimes$ and $\boxtimes$ are the outer and Kronecker products, respectively. $\| \cdot \|_2$ and $\| \cdot \|_F$ are the ℓ_2 and Frobenius norms, respectively. $\mathrm{Diag}(\cdot)$ is the diagonal operator that only keeps the diagonal element of the matrix input. Finally, $\circ$ is the composition operator: $x \circ y(t) = x\big(y(t)\big)$ and $\dot{f}$ is the derivative of f.

III. RELATED TENSOR FACTORIZATIONS

Let $\mathcal{X}$ be the three way array of the recorded data such that $\mathbf{X}_k$ is the kth trial or sample. More precisely, on EEG data, the raw recordings are split according to the saccades so that only one saccade is present in each trial, providing a data matrix $\mathbf{X}_k$ which are then stacked to form $\mathcal{X}$, where M is the number of sensors, L the number of time samples and K the number of trials. Assuming an additive gaussian i.i.d. noise, the recordings are expressed as

$$\mathcal{X} = \mathcal{T} + \mathcal{E}, \tag{1}$$

where $\mathcal{T}$ and $\mathcal{E}$ are the signal and noise arrays, respectively. The signal array $\mathcal{T}$ can be factorized to reveal the underlying latent factors. It is assumed that the shapes of sources in the second mode can change from sample to sample. The goal of this paper is to incorporate *a priori* knowledge on the variations among the shapes into tensor decomposition models, and to pave the way for further research in modeling subject variability in tensor decomposition.

Differences between shapes of second mode sources are modeled as a shape warping from a latent set of 'original' sources $b_r(t)$:

$$\tilde{b}_{kr}(t) = b_r\big(\gamma_{rk}(t)\big), \tag{2}$$

where $\gamma_{rk}(t)$ is an orientation-preserving diffeomorphism (see Fig. 1).

In addition to $\mathcal{X}$, a set of P others signals $\mathbf{Y}^{(p)} \in \mathbb{R}^{M \times K}$ are simultaneously recorded that allows additional *a priori* information on the latent sources of the second mode to be obtained. In the case of EEG ocular artifacts removal, GAZE signals (*i.e.* eye motions) are simultaneously recorded. The fusion of these data sets will be the topic of further research, since here the second data set serves only to be used as an *a priori* on the latent sources b_r or diffeomorphisms γ_{kr}.

A. Canonical polyadic decomposition

The CP decomposition represents an array $\mathcal{T}$ by a sum of rank-one tensors as

$$\mathcal{T} = \sum_{r=1}^{R} \lambda_r \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r. \quad (3)$$

Equivalently, the kth matrix slice is expressed as

$$\mathbf{T}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}', \quad (4)$$

where $\mathbf{D}_k$ is a diagonal matrix whose entries are the kth row of $\mathbf{C}\Lambda$, with Λ the diagonal matrix whose entries are λ_r. The loading matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ and the scaling factors $\boldsymbol{\lambda}$ are estimated as

$$\hat{\boldsymbol{\lambda}}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}} = \arg \min_{\boldsymbol{\lambda}, \mathbf{A}, \mathbf{B}, \mathbf{C}} \frac{1}{2} \|\mathcal{X} - \mathcal{T}\|_F^2 \quad (5)$$

subject to $\forall r$, $\|\mathbf{a}_r\|_2 = \|\mathbf{b}_r\|_2 = \|\mathbf{c}_r\|_2 = 1$.

Consequently, the CP decomposition model assumes that each trial can be factorized as a matrix product of spatial map $\mathbf{A}$ and temporal shapes $\mathbf{B}$ up to scaling factors. An underlying hypothesis of CP decomposition model is therefore that the spatial maps and temporal shapes are the same for each trial: *i.e.* $\gamma_{kr} = \delta_0$, with δ_0 the delta Dirac distribution at 0. However, this latter assumption is not always satisfied if the trials are misaligned and/or if the temporal shapes presents some time warping between them, like temporal shifts or time contractions/dilatations. Especially for outliers among time sources generated by eye movement, variability in eye saccades leads to variability in the temporal responses that are not taken into account by the CP decomposition. The issue is similar to the time retention shift issue encountered with chromatographic data [2].

B. PARAFAC2, SCREAM and Shift-PARAFAC models

To overcome these limitations of the CP decomposition, the PARAFAC2 [4] model was introduced. It is usually defined by

$$\forall k, \quad \mathbf{T}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}_k' \quad (6)$$

where $\mathbf{D}_k$ is a diagonal matrix with entries the kth row of $\mathbf{C}\Lambda$. This model differs from (4) only in the fact that the second loading matrix can be different for each trial. To ensure identifiability, the following constraints are added: $\mathbf{B}_k'\mathbf{B}_k = \mathbf{I}$, for all k. This model thus assumes that the second loading matrices share the same correlation between trials:

$$\hat{\mathbf{A}}, \{\hat{\mathbf{B}}_k\}, \hat{\mathbf{C}} = \arg \min \frac{1}{2} \|\mathcal{X} - \mathcal{T}\|_F^2, \quad (7)$$
$$\text{s.t. } \forall k, \ \mathbf{B}_k'\mathbf{B}_k = \mathbf{I},$$

with $\mathcal{T}$ defined by (6). This model better suits the assumed data structure in (2) but it does not take into account the distortion of the time axis. In other words, PARAFAC2 assumes a
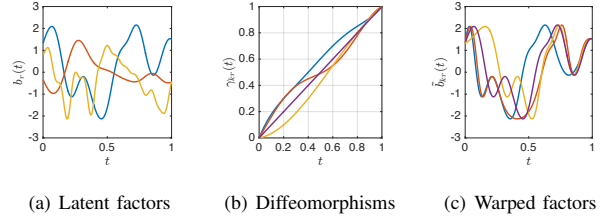

(a) Latent factors   (b) Diffeomorphisms   (c) Warped factors

Fig. 1. Typical functions b_r, warping diffeomorphisms γ_{kr} and warped factors $\tilde{b}_{kr}$. Fig. 1(a) shows 3 latent factors. Fig. 1(b) shows 4 diffeomorphisms γ_{kr} and Fig. 1(c) shows $b_1 \circ \gamma_{k1}$, *i.e.* the application of the 4 diffeomorphisms on the latent factor b_1 in blue in Fig. 1(a).

constant autocorrelation of the sources in $\mathbf{B}_k$ which is a wrong assumption if the variability among sources is due to time warping.

Another closely related model is the SCREAM [5] model which was introduced to model some data with shifts and shape changes in one mode, but provided auxiliary matrix data for which a linear regression model depending on the latent sources is valid. It is computed as a PARAFAC2 model coupled to a linear regression model with mixing matrix factor $\mathbf{C}$.

Finally when sources in $\mathbf{B}_k$ are identical but shifted in time with delays τ_{kr}, *i.e.* $\gamma_{kr} = \delta_{\tau_{kr}}$, parametrization of the delay in the indexes of columns $\mathbf{B}_r$ leads to a model called Shift-PARAFAC [3]. For EEG and eye artifact extraction, again this model cannot be applied since a simple time shift between sources is not a satisfactory model.

IV. PROPOSED REGISTERED CP DECOMPOSITION

In this paper, it is assumed that the second loading factors $\mathbf{b}_r$ come from the sampling of functions b_r and are equal from trial to trial up to orientation-preserving diffeomorphisms $\gamma_{\cdot r}$: *i.e.* $\forall r$ and $\forall (l, k)$, $b_{kr} \circ \gamma_{kr}^{-1} = b_{lr} \circ \gamma_{lr}^{-1}$. Consequently, the kth trial $\mathbf{T}_k$ is expressed as

$$\forall k, \quad \mathbf{T}_k = \mathbf{A}\mathbf{D}_k\tilde{\mathbf{B}}_k', \quad (8)$$

where $\tilde{\mathbf{B}}_k = [\boldsymbol{\Gamma}_{k1}\mathbf{b}_1, \cdots, \boldsymbol{\Gamma}_{k,R}\mathbf{b}_R]$. The matrices $\boldsymbol{\Gamma}_{kr}$ are the warping matrices (i.e. matrix representations of diffeomorphism operators).

A. Registered CP algorithm

The registered CP (rCP) decomposition aims at optimizing the following criterion:

$$\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \{\hat{\gamma}_{rk}\}_{r,k} = \arg \min_{\mathbf{A}, \mathbf{B}, \mathbf{C}, \{\gamma_{rk}\}_{r,k}} \frac{1}{2} \|\mathcal{X} - \mathcal{T}\|_F^2, \quad (9)$$

with $\mathcal{T}$ defined by (8). To optimize this criterion, an alternative least square (ALS) algorithm is used [6]: it aims at estimating sequentially each parameter with all the other parameters fixed. Consequently, we propose the following step:

1) Update $\mathbf{A}$ given $\hat{\mathbf{B}}$, $\hat{\mathbf{C}}$, $\{\hat{\gamma}_{kr}\}$

$$\hat{\mathbf{A}} \leftarrow \left(\sum_{k=1}^{K} \mathbf{X}_k \hat{\tilde{\mathbf{B}}}_k \hat{\mathbf{D}}_k \right) \left(\sum_{k=1}^{K} \hat{\mathbf{D}}_k \hat{\tilde{\mathbf{B}}}_k' \hat{\tilde{\mathbf{B}}}_k \hat{\mathbf{D}}_k \right)^{-1}. \quad (10)$$

2) Update $\mathbf{C}$ given $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$, $\{\hat{\gamma}_{kr}\}$

$$\hat{\mathbf{D}}_k \leftarrow \text{Diag}\left(\hat{\mathbf{A}}^\dagger \mathbf{X}_k \hat{\tilde{\mathbf{B}}}_k^{\dagger'}\right). \qquad (11)$$

3) i) Estimate $\tilde{\mathbf{B}}_k$, for all k, given $\hat{\mathbf{A}}$, $\hat{\mathbf{C}}$, $\{\hat{\gamma}_{kr}\}$

$$\hat{\tilde{\mathbf{B}}}_k \leftarrow \mathbf{X}_k' \hat{\mathbf{A}}^\dagger \hat{\mathbf{D}}_k^\dagger. \qquad (12)$$

ii) Factorize each matrix $\left[\hat{\tilde{\mathbf{b}}}_{1r}, \cdots, \hat{\tilde{\mathbf{b}}}_{Kr}\right]$ as $\Gamma_r(\mathbf{I}_K \boxtimes \mathbf{b}_r)$, where $\Gamma_r = [\Gamma_{1r}, \cdots, \Gamma_{Kr}]$ is the concatenation of warping matrices for the rth component of the second loading factor, by

$$\hat{\mathbf{b}}_r, \{\hat{\gamma}_{kr}\}_k = \arg\min_{\mathbf{b}_r, \{\gamma_{kr}\}_k} \sum_{k=1}^K d\left(\Gamma_{kr}\mathbf{b}_r, \tilde{\mathbf{b}}_{kr}\right)^2, \qquad (13)$$

where $d(\cdot)$ is a distance defined in the next paragraph.

iii) Then, $\hat{\tilde{\mathbf{B}}}_k$ is updated as

$$\forall k, \quad \hat{\tilde{\mathbf{B}}}_k = \left[\hat{\Gamma}_{k1}\hat{\mathbf{b}}_1, \cdots, \hat{\Gamma}_{kR}\hat{\mathbf{b}}_R\right]. \qquad (14)$$

B. Factorization of warped loadings

The underlying structure of the warped loadings $\tilde{\mathbf{B}}_k$ is recovered by estimating both the generating function b_r and the warping diffeomorphisms γ_{kr} [7]. Let $\mathbb{F}$ be the set of absolutely continuous functions defined on $[0, 1]$ and $\mathbb{G}$ be the set of orientation-preserving diffeomorphisms of the unit interval $[0, 1]$ to $[0, 1]$.

The aim of step 3) defined by (13) is to find a function $b_r \in \mathbb{F}$ so that it is the best approximation of all functions $\tilde{b}_{kr}$ up to warping functions $\gamma_{kr} \in \mathbb{G}$. To do this, let us use an elastic distance between functions, *i.e.* a distance in the quotient space $\mathbb{F}/\mathbb{G}$ which is the set of orbits of the type $[f] = \{f \circ \gamma \mid \forall \gamma \in \mathbb{G}\}$, for any $f \in \mathbb{F}$. The used elastic distance between two functions f_1 and f_2 of $\mathbb{F}$ is defined by the Fisher-Rao Riemannian metric [8], [9]:

$$d(f_1, f_2) = \inf_{\gamma \in \mathbb{G}} d_{FR}(f_1, f_2 \circ \gamma).$$

However, as pointed out by [7], it has been shown [10] that this distance can be efficiently computed by transforming the functional space $\mathbb{F}$ into $\mathbb{L}^2$, the set of square integrable functions, using the square-root slope function (SRSF): $q = \text{sign}(\dot{f})\sqrt{|\dot{f}|}$, with $\text{sign}(\cdot)$ the sign function. Indeed, in $\mathbb{L}^2$, the Fisher-Rao distance becomes the Euclidian distance: $d_{FR}(f_1, f_2) = \|q_1 - q_2\|$. Consequently, the latent loading b_r is estimated as the Karsher mean of $\{\tilde{b}_{1r}, \cdots, \tilde{b}_{Kr}\}$ defines from its SRSF as

$$\hat{q}_r = \arg\min_{q \in \mathbb{L}^2} \sum_{k=1}^K \left(\inf_{\gamma_k \in \mathbb{G}} \left\| q - (q_k \circ \gamma_k)\sqrt{\dot{\gamma}_k} \right\|^2\right), \qquad (15)$$

where q_k is the SRSF associated with $\tilde{b}_{kr}$. The set of warping diffeomorphisms γ_{kr} are then estimated as

$$\hat{\gamma}_{kr} = \arg\inf_\gamma \left\| \tilde{q}_{kr} - (\hat{q}_r \circ \gamma)\sqrt{\dot{\gamma}} \right\|^2, \qquad (16)$$

by dynamical programming [11]. Finally, b_r is expressed as

$$\hat{\mathbf{b}}_r = \left(\sum_{k=1}^K c_{kr}^2 (\mathbf{a}_r' \mathbf{a}_r)\hat{\Gamma}_{kr}'\hat{\Gamma}_{kr}\right)^{-1}$$
$$\times \sum_{k=1}^K \left(c_{kr}\hat{\Gamma}_{kr}'\left(\mathbf{X}_k' - \sum_{l \neq r} c_{kl}\hat{\Gamma}_{kl}\mathbf{b}_l\mathbf{a}_l'\right)\mathbf{a}_r\right). \qquad (17)$$

It is worth noting that some trivial underdetermination are inherent to the assumption (2). Indeed, one can *blindly* only recover the latent sources b_r and the diffeomorphisms γ_{kr} up to diffeomorphisms $\tilde{\gamma}_{kr}$ since

$$\tilde{b}_{kr} = b_r \circ \gamma_{kr} = \left(b_r \circ \tilde{\gamma}_{kr}^{-1}\right) \circ \left(\tilde{\gamma}_{kr} \circ \gamma_{kr}\right).$$

C. Using priors on b_r or γ_{kr}

The above scheme assumes that no prior information is known about neither the latent factor shapes b_r nor the warping diffeomorphisms γ_{kr}. However, in some experiments, such prior information can be available. This is for instance the case if GAZE&EEG data are recorded synchronously. Indeed, the GAZE recordings can be used either to initialize the iterative ALS algorithm or as prior on either b_r or γ_{kr}.

If all warping diffeomorphisms γ_{kr} are known, then the third step (13) resumes to (17) by substituting $\hat{\Gamma}_{kr}$ by its actual value Γ_{kr}. Similarly, if latent factors b_r are known, then the third step (13) resumes to (16) by substituting $\hat{q}_r$ with q_r, the SRSF of b_r.

V. RESULTS

Both numerical simulations and an illustration on actual GAZE&EEG data are presented.

A. Numerical simulations

In each simulation, the number of components R is fixed to 3. The length of the first, second and third loading factors $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$ are 10, 250 and 50, respectively. The first and third loadings factors $\mathbf{A}$ and $\mathbf{C}$ are generated randomly from a standard normal distribution. Each component $\mathbf{b}_r$ of the second loading factor $\mathbf{B}$ is generated from a Gaussian process [12] with a zero mean function and a covariance function $k_r(t, t') = \exp\left(-(t-t')^2/2l_r^2\right)$, with l_r uniformly distributed in $[.02; .22]$, so that it is smooth enough with still some variations. The warping diffeomorphisms are generated as

$$\gamma_{kr}(t) = \int_0^t \alpha \exp\left(w_{kr}(u)\right)du,$$

where α is a normalisation factor so that $\gamma_{kr}(1) = 1$. Functions w_{kr} are randomly drawn as Gaussian processes with a zero mean value function and a covariance function $k_{kr}^{(w)}(t, t') = \sigma^2 \exp\left(-(t - t')^2/2\lambda_{kr}^2\right)$, with $\sigma = .8$ and λ_{kr} uniformly distributed in $[.1; .4]$. Typical functions b_r and warping diffeomorphisms γ_{kr} are plotted in Fig. 1. Entries of tensor $\mathcal{E}$ are drawn from a zero mean normal distribution with a standard deviation equal to .5. For each numerical experiment, 50 configurations are randomly drawn by first regenerating random parameters l_r, λ_{kr} and all warping diffeomorphisms γ_{kr} and then each factors $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ and additive noise $\mathcal{E}$.

The quality of the tensor estimation is quantified by two performance indexes. The overall signal-to-noise ratio improvement (Π_O) is defined as

$$\Pi_O(\hat{\mathcal{T}}) = 10\log\left(\frac{\|\mathcal{T}\|_F^2}{\|\mathcal{T} - \hat{\mathcal{T}}\|_F^2}\right) - 10\log\left(\frac{\|\mathcal{T}\|_F^2}{\|\mathcal{T} - \hat{\mathcal{T}}_{init}\|_F^2}\right),$$

where $\mathcal{T}$, $\hat{\mathcal{T}}$ and $\hat{\mathcal{T}}_{init}$ are the true, estimated and initial tensors, respectively. The higher, the better the results. The factor performance index (Π_F), which measures the quality estimation of each factor, is defined as

$$\Pi_F = \begin{cases} 1 - \dfrac{\left|\hat{\mathbf{d}}_r' \mathbf{d}_r\right|}{\|\hat{\mathbf{d}}_r\|_2 \|\mathbf{d}_r\|_2}, & \text{if } \mathbf{d}_r \text{ is a non warped factor,} \\ \inf_{\gamma \in \mathbb{G}} d_{FR}(\hat{d}_r \circ \gamma, d_r), & \text{if } d_r \text{ is a warped factor.} \end{cases}$$

Based on the trivial underdetermination of the warped factor, its performance is measured as the elastic distance in the quotient space $\mathbb{F}/\mathbb{G}$, while the other ones are based on the correlation coefficient between the true loading and its estimate. For the factor performance index, the closer to zero, the better the results.

The numerical results are shown on Fig. 2. As one can see, the proposed rCP allows to estimate the underlying factors (Fig. 2(b), 2(c), 2(d)). Indeed, the performance indexes Π_F after the estimation are lower than the Π_F at the initialization. Both methods rCP and rCP-B provide quite similar performance but lower than rCP-G which assumes known the diffeomorphisms. Quite surprisingly, rCP-B leads to a little bite lower performance than rCP: such results will be more deeply investigated in a future work. The proposed rCP outperforms both the classical CP and PARAFAC2: after convergence, the estimated tensors are worse than the initialized one (the Π_O is negative). Indeed, for these two methods, this result is explained by a worse estimation on the second loading factor (Fig. 2(c)). For the CP, one can explain this by the miss-modeling which does not assume some shapes modifications in any loadings. This is expected on the warped loading (Fig. 2(c)). Consequently, the first one, which is theoretically not affected by a miss-modeling, and the third one (Fig. 2(d)) which models the amplitude of factors are also less accurately estimated (Fig. 2(b)) than using rCP. The more flexible PARAFAC2 model also fails to estimate the underlying latent factors $\mathbf{B}$ since the orthogonality constraint is not valid but its flexibility leads to a better estimation of the first and third loadings than CP.

B. Illustration on GAZE&EEG data

The Gaze&EEG data comes from experiments in visual search where participants had to search a target from a set of distractors [13]. GAZE and EEG data were recorded synchronously. From these data, two tensors (space × time × trials) are built synchronized on the beginning of the saccades: each tensor is the concatenation in the third dimension of the matrices in which each column is the vector of channel signals at a given time. The number of components is fixed to $R = 2$, EEG tensor is of size 63 channels, 250 time samples and 47 trials. The initializations of γ_{kr} have been done from the GAZE signals. As one can see (Fig. 3), after ocular removal by rCP, the remaining EEG is much more stationary than using CP. rCP removes almost all the artefacts while CP does not.
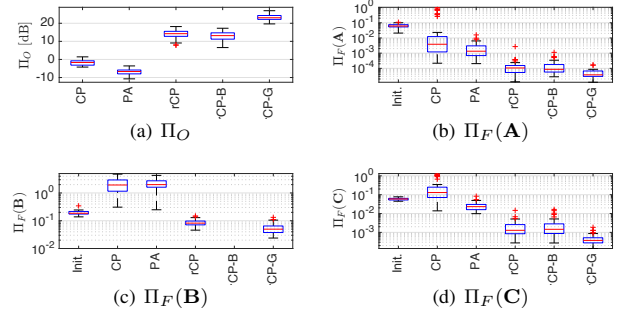

Fig. 2. Performance of the proposed rCP. Comparison of performance indexes for several assumptions from initialization (Init.), usual CP (CP), PARAFAC2 (PA), proposed registered CP (rCP), rCP with known latent factors $\mathbf{B}$ (rCP-B) and rCP with known warping functions γ_{kr} (rCP-G). On each box, the red mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to extreme values and the outliers are plotted individually as red crosses.
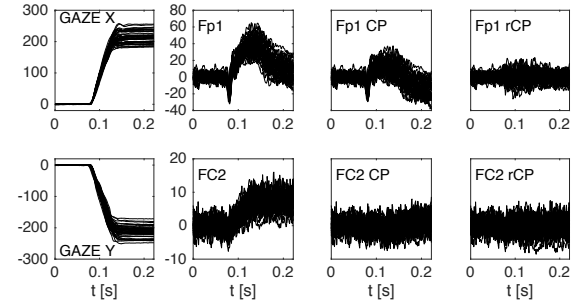

Fig. 3. Illustration on EEG data. Columns from left to right: GAZE signals, Fp1 electrode and FC2 electrode raw EEG, ocular artefacts removal by CP and by rCP.

VI. CONCLUSION

In this article, the factorization of a three-way array has been proposed for the case that one mode presents shape variations. The proposed registered-CP decomposition processes the mode with shape variations as orientation-preserving diffeomorphisms. It is thus a generalization of the Shift-PARAFAC method since it considers a larger class of shape variations and it is related to the PARAFAC2 model, but different constraints are imposed. An ALS scheme has been proposed to estimate the loadings and the latent shapes with the associated diffeomorphisms. The numerical simulations show the potential of the rCP which outperforms both the CP and the PARAFAC2 decompositions.

The current algorithm suffers from its computational cost due to the implementation using dynamic programming. Thus some efficient algorithms must be proposed to increase the efficient use of the rCP algorithm. In perspective, some applications like chromatographic data require additional assumptions on some loadings such as non-negativity, those cases will be considered in the future. Finally, a theoretical study must be made to provide some sufficient/necessary conditions of identifiability of the proposed rCP model.

REFERENCES

[1] A. Cichocki, D. Mandic, L. D. Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. PHAN, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, March 2015.

[2] R. Bro, C. A. Andersson, and H. A. Kiers, "PARAFAC2–Part II. Modeling chromatographic data with retention time shifts," *Journal of Chemometrics*, vol. 13, no. 3-4, pp. 295–309, 1999.

[3] R. A. Harshman, S. Hong, and M. E. Lundy, "Shifted factor analysis–Part I: Models and properties," *Journal of chemometrics*, vol. 17, no. 7, pp. 363–378, 2003.

[4] R. A. Harshman, "Parafac2: Mathematical and technical notes," *UCLA Working Papers in Phonetics*, vol. 22, pp. 30–44, 1972.

[5] F. Marini and R. Bro, "SCREAM: A novel method for multi-way regression problems with shifts and shape changes in one mode," *Chemometrics and Intelligent Laboratory Systems*, vol. 129, no. 0, pp. 64 – 75, 2013, multiway and Multiset Methods.

[6] P. Comon, X. Luciani, and A. L. F. de Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.

[7] J. D. Tucker, W. Wu, and A. Srivastava, "Generative models for functional data using phase and amplitude separation," *Computational Statistics & Data Analysis*, vol. 61, pp. 50 – 66, 2013.

[8] C. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of Calcutta Mathematical Society*, vol. 37, pp. 81–91, 1945.

[9] C. R. Rao, *Breakthroughs in Statistics: Foundations and Basic Theory*. Springer New York, 1992, ch. Information and the Accuracy Attainable in the Estimation of Statistical Parameters, pp. 235–247.

[10] A. Bhattacharyya, "On a measure of divergence between two multinomial populations," *Sankhy: The Indian Journal of Statistics (1933-1960)*, vol. 7, no. 4, pp. 401–406, 1946.

[11] D. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 1995.

[12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[13] E. Kristensen, A. Guérin-Dugué, and B. Rivet, "Comparison between Adjar and xDAWN algorithms to estimate eye-fixation related potentials distorted by overlapping," in *Proc. International IEEE EMBS Conference on Neural Engineering (NER)*, Montpellier, France, April 2015, pp. 976–979.