Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2015 August ; 2015: 3496–3499. doi:10.1109/EMBC.2015.7319146.

Decomposing Time Series Data by a Non-negative Matrix Factorization Algorithm with Temporally Constrained Coefficients

Vincent C. K. Cheung,

School of Biomedical Sciences of The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR, China

Karthik Devarajan,

Fox Chase Cancer Center, Philadelphia, PA, 19111 USA

Giacomo Severini,

Department of Physical Medicine and Rehabilitation of the Harvard Medical School, Boston, MA, 02129 USA

Andrea Turolla, and

IRCCS San Camillo Hospital Foundation, Venice, Italy

Paolo Bonato

Department of Physical Medicine and Rehabilitation of the Harvard Medical School, Boston, MA, 02129 USA

Abstract

The non-negative matrix factorization algorithm (NMF) decomposes a data matrix into a set of non-negative basis vectors, each scaled by a coefficient. In its original formulation, the NMF assumes the data samples and dimensions to be independently distributed, making it a less-than-ideal algorithm for the analysis of time series data with temporal correlations. Here, we seek to derive an NMF that accounts for temporal dependencies in the data by explicitly incorporating a very simple temporal constraint for the coefficients into the NMF update rules. We applied the modified algorithm to 2 multi-dimensional electromyographic data sets collected from the human upper-limb to identify muscle synergies. We found that because it reduced the number of free parameters in the model, our modified NMF made it possible to use the Akaike Information Criterion to objectively identify a model order (i.e., the number of muscle synergies composing the data) that is more functionally interpretable, and closer to the numbers previously determined using *ad hoc* measures.

I. INTRODUCTION

The non-negative matrix factorization algorithm (NMF) is an unsupervised learning method that decomposes high-dimensional non-negative data into two non-negative matrices – one

comprising column vectors that can be regarded as basis vectors in the data space, and another comprising rows of coefficients that scale the basis vectors [1]. Because of its non-negativity constraint, the NMF tends to return sparsely represented basis vectors that are more easily interpretable in practical applications [1]. Thus, it has been increasingly applied in analysis of large-scale data in areas such as neuroscience, computational biology, and processing of biomedical images [2].

The NMF was originally formulated as an algorithm with update rules that iteratively decrease the squared error between the data matrix and its reconstruction by the decomposed matrices. It further assumes that the data samples and data dimensions are both independently distributed [3] [4]. Because of this independence assumption, one may argue that it is a less-than-ideal algorithm for decomposing any time series data, which invariably contain dependencies across samples collected at different times.

In this paper, we explore how the NMF update rules may be modified to account for temporal dependencies in the data set. We introduced a very simple constraint, across temporal samples, within the coefficient matrix, and incorporated this constraint into the NMF update rules. We then evaluated the performance of this algorithm in human electromyographic (EMG) data, a type of multi-dimensional time series data that has been popularly analyzed by the NMF in the motor control field for the extraction of muscle synergies [5]. We found that the incorporation of this coefficient constraint made it possible to use the Akaike Information Criterion (AIC) to objectively determine a model order that is more functionally interpretable.

II. THE ALGORITHM

Let $D = [\mathbf{d}_1 \dots \mathbf{d}_j \dots \mathbf{d}_T]$ be a series of M -dimensional non-negative column vectors. The NMF algorithm models the data as a linear combination of N M -dimensional non-negative basis vectors, N M , such that

$$\mathbf{d}_j = \sum_{k=1}^N C_{kj} \mathbf{w}_k + \varepsilon, \quad (1)$$

where $\mathbf{w}_k$ is the k th basis vector, C_{kj} is the non-negative coefficient for the k th basis vector at data point j , and ε is any noise unexplained by the model. Let $W = [\mathbf{w}_1 \dots \mathbf{w}_k \dots \mathbf{w}_N]$, so that $D \approx WC$, where C is an $N \times T$ matrix with C_{kj} being the matrix entry at row k , column j . When ε follows a Gaussian distribution with constant variance [3], the W and C matrices may be estimated by the following multiplicative update rules [1]:

$$C_{kj}^{i+1} = C_{kj}^i \frac{[(W^i)^T D]_{kj}}{[(W^i)^T W^i C^i]_{kj}}; \quad (2)$$

$$W_{ak}^{i+1} = W_{ak}^i \frac{[D(C^{i+1})^T]_{ak}}{[W^i C^{i+1} (C^{i+1})^T]_{ak}}, a=1 \dots M,$$

where the superscript i denotes the iteration number, and the superscript T denotes the matrix transpose. We shall call the above update rules the Traditional NMF.

The Traditional NMF assumes that both the data samples and data dimensions are independently distributed [3] [4]. The new algorithm proposed here aims to impose a predefined structure across the columns of the C matrix. If, for instance, D is a time series of M -dimensional vectors so that each column of D represents a data point collected at a specific time, the structure we will incorporate into the NMF algorithm amounts to adding a temporal constraint to components of C . Here, as a demonstration of how such a temporal structure may be imposed, we will employ a very simple constraint. For the coefficients of every basis vector (i.e., for every row of C), we require every other coefficient component be the average of the previous and following components. Thus, the coefficients are forced by the algorithm to observe a simple smoothness constraint.

Without loss of generality, assume T to be an odd integer. We propose that the C and W matrices may be estimated from D by the following modified NMF algorithm with an explicit constraint on C :

$$C_{kj}^{i+1} = C_{kj}^i \frac{[(W^i)^T D]_{kj}}{[(W^i)^T W^i C^i]_{kj}} \text{ for } j=1, 3, 5 \dots, T; \quad (3)$$

$$C_{kj}^{i+1} = \frac{C_{k(j-1)}^{i+1} + C_{k(j+1)}^{i+1}}{2} \text{ for } j=2, 4, 6 \dots T-1;$$

$$W_{ak}^{i+1} = W_{ak}^i \frac{[D(C^{i+1})^T]_{ak}}{[W^i C^{i+1} (C^{i+1})^T]_{ak}}.$$

We shall call the above update rules the NMF with Interpolated Coefficients, or NMF_i.

III. MODEL ORDER DETERMINATION

The implementation of NMF requires the number of basis vectors (i.e., the number of columns in the W matrix) to be known *a priori*. When applying the NMF for data analysis, the desired model order would have to be determined by some other model selection criteria. In the motor neuroscience field, most previous studies employing the NMF for the extraction of muscle synergies (i.e., the basis vectors in muscle space) from multi-channel EMG data have relied on *ad hoc* procedures to identify the correct number of synergies. These

procedures include finding the model order at which the R^2 of data reconstruction plateaus [6], or finding the smallest number of basis vectors that allows the R^2 to be greater than a certain predefined threshold [7]. While these methods have been shown to be very robust in simulated data sets [8], defining where the R^2 plateaus or deciding what the R^2 threshold should be remains a rather arbitrary decision.

To circumvent this arbitrariness in model-order selection, we have previously proposed that the AIC can be used to objectively determine the best model order for the NMF [4]. Grounded on information theory, the AIC offers a balance between the goodness-of-fit of the model to the data and the number of parameters in the model. According to this criterion, the model with the minimum AIC value is the preferred model [9]. For the NMF algorithm with Gaussian noise assumption, the AIC is given by

$$AIC = 2\left(\frac{E}{2\sigma^2} + \psi\right), \quad (4)$$

where E is the minimum reconstruction error, σ^2 is the estimated variance of the data set to be decomposed, and ψ is the total number of free parameters in the factorization model. Here, all data sets we analyzed were variance-normalized in each muscle, and thus σ^2 was assumed to be 1. The AIC for the Traditional NMF is given by

$$AIC = \left(\sum_{aj} [D_{aj} - (WC)_{aj}]^2\right) + 2N(M+T). \quad (5)$$

For the NMF_i, the interpolated coefficients are not counted as free parameters. The AIC is reduced to

$$AIC = \left(\sum_{aj} [D_{aj} - (WC)_{aj}]^2\right) + 2N\left(M + \frac{T+1}{2}\right). \quad (6)$$

IV. EXPERIMENTAL DATA

We evaluated the utility of the NMF_i algorithm and the performance of its associated AIC for model order selection in two different EMG data sets previously described in [7]. Both data sets comprised multi-channel upper-limb surface EMGs collected from the unaffected arm of human stroke survivors during different tasks. The Traditional NMF and NMF_i were employed to extract muscle synergies, or muscle groupings posited to be fundamental control modules of the human motor system, from the EMGs. The number of muscle synergies composing the data of each session was determined using the AIC.

Data set 1 (Boston Data)

A total of 35 sessions of EMG recordings from 15 subjects were included in this data set. Briefly, stroke survivors recruited from the Spaulding Rehabilitation Hospital, Boston, MA,

were asked to perform a reaching task using the unaffected arm. The task consisted of movement, at a self-selected speed, from an initial position to 1 of 12 possible targets in three-dimensional space, and then from the target back to the initial position. In each session, 3–4 blocks of trials were performed. Each block consisted of 12 trials (1 trial per target).

As the subject performed the behavioral tasks, EMG activities of 14 muscles were collected at 3,000 Hz using surface bipolar electrodes (Motion Lab Systems, MA-300). The muscles recorded included infraspinatus (Infraspin); latissimus dorsi (LatDor); rhomboid major (RhomMaj); superior trapezius (TrapSup); pectoralis major, clavicular head (PectClav); deltoid, anterior (DeltA), medial (DeltM), and posterior (DeltP) parts; triceps brachii, lateral head (TrLat); biceps brachii, short (BicShort) and long (BicLong) heads; brachialis (Brac); brachioradialis (BrRad); and pronator teres (PronTer).

Data set 2 (Venice Data)

A total of 36 sessions of EMG recordings from 28 subjects were included in this data set. Participants were recruited from the IRCCS San Camillo Hospital Foundation, Venice, Italy. The motor behaviors tested consisted of 7 virtual-reality tasks (10–11 trials per task) performed on the Virtual Reality Rehabilitation System (Khymeia Group, Noventa Padovana, Italy). Surface bipolar EMGs (Biopac Systems; 1,000 Hz) were recorded from all muscles listed in the Boston Data, except LatDor, TrapSup, and BicShort; in addition, triceps brachii, medial head (TrMed) and supinator (Supin) were also recorded.

All procedures were approved by the Ethics Committees of the participating hospitals.

Initializing NMF

Before muscle synergy analysis, the collected EMGs were pre-processed as described before [7]. To equalize EMG amplitude differences between muscles, the data from each muscle were normalized to unit variance. For all extractions involving the Traditional NMF, both the W and C matrices were initialized with random components, uniformly distributed between 0 and the maximum value the D matrix. For extractions involving NMF_i, in every row of the initializing C each even-column component was linearly interpolated as the average of the values at the preceding and following columns. For EMG data sets with an even number of columns, the last column was treated like an odd column. Convergence was defined as having 20 consecutive iterations that resulted in a change of EMG-reconstruction $R^2 < 0.001\%$. We repeated synergy extraction 5 times for every session of EMG data, each time with different initializing matrices. The repetition with the lowest AIC was then selected for any subsequent analysis.

V. RESULTS

A. Convergence of NMF_i

Convergence of the update rules of the Traditional NMF (2) has been previously proven [10]. For the NMF_i update rules (3), even though we are unable to supply a formal proof of convergence here, in every instance of its application in both the Boston and Venice Data,

we observed a monotonic increase in the EMG-Reconstruction R^2 as the W and C matrices were updated, with our convergence criterion satisfied in less than 1000 iterations.

We show below some examples of the evolution of the R^2 value during muscle-synergy extraction using the NMFi, in both the Boston (Fig. 1A) and Venice (Fig. 1B) Data Sets.

B. The NMFi models returned smaller AIC values

To identify the optimal number of muscle synergies composing the EMGs of each data session, we successively increased the number of synergies extracted from 1 to the number of muscles recorded (14 for Boston Data; 13 for Venice Data), and then plotted the AIC values from the extractions against the model order. For the Boston Data, when Traditional NMF was used the AIC reached a minimum, on average, at 2 synergies (Fig. 2A, blue); when NMFi was used the AIC reached a minimum at 3 synergies instead (Fig. 2A, red). Similarly, for the Venice Data, the AIC selected a desired model order of 2 synergies for Traditional NMF, but 4 for NMFi (Fig. 2C). Thus, when AIC was used as the objective model selection criterion, a higher model order was indicated when NMFi was employed to extract the muscle synergies.

More importantly, in every EMG session, the AIC value at the optimal model order obtained from NMFi was smaller than that obtained from Traditional NMF (Fig. 2B, 2D). This suggests that the NMFi model with the smoothness constraint imposed onto the coefficients was a better description of our human EMG data.

C. The NMFi models had better goodness-of-fit

At the optimal model order determined by AIC, the synergies and coefficients extracted by NMFi also described the EMGs better than those extracted by Traditional NMF. For the Boston Data, the Traditional NMF achieved an EMG-Reconstruction R^2 of $56.27 \pm 5.39\%$ (mean $\pm$ SD) at the number of synergies selected by AIC; the NMFi, on the other hand, achieved an R^2 of $60.39 \pm 4.69\%$. Similarly, for the Venice Data, the R^2 values for the Traditional NMF and NMFi were $41.37 \pm 10.28\%$ and $65.26 \pm 6.03\%$, respectively.

Not surprisingly, the numbers of muscle synergies determined for NMFi, as compared with the numbers for Traditional NMF, were closer to the numbers previously determined using the criterion of selecting the minimum number of synergies yielding an R^2 of $>80\%$ [7]. As shown in Fig. 3, the distribution of model order for NMFi falls between that for Traditional NMF and that determined by the 80%- R^2 criterion.

D. Muscle synergies returned by NMFi

Despite the additional smoothness constraint imposed onto the coefficients during updating, at the model order selected for NMFi, the Traditional NMF and NMFi returned virtually identical muscle synergies. We quantified the similarity between the synergies obtained from the two algorithms using the scalar product value. For the Boston Data, the scalar product for every pair of synergies extracted by the two algorithms reached 0.9960 ± 0.0077 , and for the Venice Data, 0.9793 ± 0.0850 . An example below (Fig. 4) from one Venice EMG session highlights this similarity.

VI. DISCUSSION AND CONCLUSION

When the Traditional NMF was applied to the EMG data, the AIC selected a very small number of muscle synergies ($= 2$) in both data sets (Fig. 2). At this number, not only was the goodness-of-fit of the model to the data very poor, but it was also difficult to biomechanically and functionally interpret the extracted muscle synergies, especially for tasks as complex as multi-directional reaching. We speculate that, one reason for this small optimal model order is that, because of the temporal dependencies present in the EMGs, there were redundant free parameters in the C matrix in the Traditional NMF model. As a result, the increase in the number of free parameters (ψ in (4)) did not contribute enough to the decrease in reconstruction error as the number of synergies extracted increased, thus forcing the AIC to trough early. In the NMF_i, because of the additional coefficient constraint, ψ increased at a slower rate that matched the decrease in error better as model complexity increased. This resulted in both a higher and more reasonable optimal model order that is closer to previously determined model order (Fig. 3), and lower AICs (Fig. 2).

We realize that the EMG-reconstruction R^2 achieved by NMF_i at the optimal model order was still on the low side (60–65%), and that this number determined by AIC was also lower than previously determined number using the 80%- R^2 criterion (Fig. 3). However, our analysis illustrates what even a very simple constraint on the coefficients can do to improve the performance of the AIC without sacrificing the quality of muscle-synergy identification (Fig. 4). Our results strongly suggest that when an appropriate constraint is incorporated into the NMF update rules, the AIC should select a functionally interpretable model order that describes the data well. A comparison of results obtained with different constraints should illuminate the precise nature of the temporal dependencies present in the EMG data set.

In conclusion, we have presented here an NMF algorithm with an explicit constraint on the coefficients incorporated into the update rules. This modification allows the profitable use of the AIC as an objective criterion for identifying the model order. Future research should investigate NMF formulated for other constraints, and how these constrained algorithms may be derived for signal-dependent noise [4].

Acknowledgments

We thank Prof. Emilio Bizzi of MIT for his enthusiastic support of this study. We also thank Patrick Kasi, Catherine Adans-Dester, Caoimhe Bennis, Stefano Silvoni and Michela Agostini for their help with data collection.

Research supported by funds from The Chinese University of Hong Kong to V. C., US NIH grant P30 CA 06927 to K. D., Italian Ministry of Health GR-2011-02348942 to A. T., and NIH-NICHD grant R24HD050821-07.

References

1. Lee DD, Seung HS. Learning the parts of objects by nonnegative matrix factorization. *Nature*. 1999; 401:788–91. [PubMed: 10548103]
2. Devarajan K. Nonnegative matrix factorization: An analytical and interpretive tool in computational biology. *PLoS Computational Biology*. 2008; 4(7):E1000029. [PubMed: 18654623]
3. Cheung, VCK., Tresch, MC. Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. Piscataway, NJ: IEEE; 2005. Nonnegative matrix factorization algorithms modeling noise distributions within the exponential family; p. 4990-3.

4. Devarajan K, Cheung VCK. On nonnegative matrix factorization algorithms for signal-dependent noise with application to electromyography data. *Neural Computation*. 2014; 26:1128–68. [PubMed: 24684448]
5. Bizzi E, Cheung VCK. The neural origin of muscle synergies. *Front Comp Neurosci*. 2013; 7:51.
6. d'Avella A, Saltiel P, Bizzi E. Combinations of muscle synergies in the construction of a natural motor behavior. *Nature Neuroscience*. 2003; 6(3):300–308. [PubMed: 12563264]
7. Cheung VCK, Turolla A, Agostini M, Silvoni S, Bennis C, Kasi P, Paganoni S, Bonato P, Bizzi E. Muscle synergy patterns as physiological markers of motor cortical damage. *Proceedings of the National Academy of Sciences USA*. 2012; 109(36):14652–6.
8. Tresch MC, Cheung VC, d'Avella A. Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets. *Journal of Neurophysiology*. 2006; 95(4):2199–2212. [PubMed: 16394079]
9. Akaike, H. Information theory and an extension of the maximum likelihood principle. In: Petrov, BN., Csaki, F., editors. *Second international symposium on information theory*. Budapest: Akademiai Kiado; 1973. p. 267-281.
10. Lee, DD., Seung, HS. Algorithms for non-negative matrix factorization. In: Leen, TK., Dietterich, TG., Tresp, V., editors. *Advances in neural information processing systems*. Vol. 13. Cambridge, MA: MIT Press; 2001. p. 556-62.

Figure 1. The evolution of R^2 during the extraction of muscle synergies using NMF
 Shown here are randomly selected collections of R^2 curves showing the monotonic increase in the EMG-reconstruction R^2 during muscle-synergy extraction using the NMF, for different subjects, and at different numbers of synergies.

Figure 2. Using the AIC to select the number of muscle synergies

A, A plot of the AIC against model order, when the Traditional NMF (blue) and NMFi (red) were applied to the EMGs, respectively, for the Boston Data. *, model order with minimum AIC. *B*, A comparison, for Boston Data, of the minimum AIC obtained by Traditional NMF and that obtained by NMFi. Each line connects values obtained from the same EMG session. *C*, Same as *A*, but for Venice Data. *D*, Same as *B*, but for Venice Data.

Fig. 3.
Distribution of optimal model order selected by 3 methods

Figure 4. Similarity between muscle synergies extracted by the Traditional NMF and NMFi
The scalar product between each synergy pair is shown above the synergies.