SLOW FEATURES NONNEGATIVE MATRIX FACTORIZATION FOR TEMPORAL DATA DECOMPOSITION

Lazaros Zafeiriou[1], Symeon Nikitidis[1], Stefanos Zafeiriou[1] and Maja Pantic [1,2]

[1] Department of Computing, Imperial College London, UK
[2] EEMCS, University of Twente, NL
{l.zafeiriou12,s.nikitidis,s.zafeiriou,m.pantic}@imperial.ac.uk

ABSTRACT

In this paper, we combine the principles of temporal slowness and nonnegative parts-based learning into a single framework that aims to learn slow varying parts-based representations of time varying sequences. We demonstrate that the proposed algorithm arises naturally by embedding the Slow Features Analysis trace optimization problem in the nonnegative subspace learning framework and derive novel multiplicative update rules for its optimization. The usefulness of the developed algorithm is demonstrated for unsupervised facial behaviour dynamics analysis on MMI database.

Index Terms— Nonnegative Matrix Factorization, Slow Features Analysis, Facial behaviour dynamics analysis

1. INTRODUCTION

Arguably image data high dimensionality is one of the most crucial problems that every image processing algorithm has to overcome. To alleviate this problem latent feature learning methods that aim to effectively represent the high dimensional image data in a simpler and more compact form, are currently widely adopted. These methods aim to identify an appropriate subspace where a certain criterion is optimized and the latent image features are discovered by performing a linear or non-linear projection of the image. These extracted latent features can significantly decrease computational complexity and boost the performance of the succeeding image processing algorithms.

In computer vision research significant attention has been attracted in developing latent feature learning algorithms that mimic the functions of the human visual system. Two of the most popular and efficient principles that model human visual perception and have inspired significant volume of research in latent feature learning, are the temporal slowness and the parts-based representation. Nonnegative Matrix Factorization (NMF) [1], is a representative parts-based learning algorithm widely used in image processing. It is an unsupervised data matrix decomposition method that requires both the matrix being decomposed and the derived factors to contain nonnegative elements. The nonnegativity constraint leads to a parts-based representation, since it allows only additive and not subtractive combinations. The semantic interpretability of the nonnegative subspace learning is enhanced, since this conforms nicely to identifying appropriate basic elements, corresponding to the basis images, which are added to reconstruct the original data. Moreover, there is a close relationship between NMF and objects parts-based representation in the human brain, since the firing rates of the visual cortex neurons can never be negative.

Slow Feature Analysis (SFA) [2] is a latent feature learning algorithm that intuitively imitates the functionality of the receptive fields of the visual cortex [3], thus being appropriate for describing the evolution of time varying visual phenomena. The temporal slowness learning principle in SFA was motivated by the empirical observation that higher order meanings of sensory data, such as objects and their attributes, are often more persistent (i.e., change smoothly) than the independent activation of any single sensory receptor. For instance, in facial behaviour analysis it has been recently shown that SFA learning can discover mapping functions between an input image sequence that varies quickly and the corresponding high-level semantic concepts that vary slowly [4].

In this paper, we combine the principles of temporal slowness and nonnegative parts-based learning into a single combined framework. The proposed novel algorithm called Slow Features Nonnegative Matrix Factorization (SFNMF) aims to learn slow varying parts-based representations of time varying sequences for unsupervised facial behaviour dynamics analysis. More precisely, we aim to accurately capture the transitions between the temporal phases of facial Action Units (AUs). To derive SFNMF optimization problem we demonstrate that this naturally arises by embedding the SFA trace optimization problem in the nonnegative subspace learning framework, resulting in simultaneously minimizing the reconstruction error and the temporal variance of the derived latent features. As it has been shown in the literature SFA is a special case of Locality Preserving Projections (LPP) [5] acquired by defining the data neighborhood structure using their temporal variations [6]. Thus, SFNMF problem is similar to other recent NMF based algorithms that incorporate

manifold regularization constraints expressed in the form of a trace optimization problem using their graph Laplacian form [7, 8, 9].

2. A BRIEF REVIEW OF NMF & SFA

Without losing generality let us assume that NMF is applied for the decomposition of N images stored is the nonnegative data matrix $\mathbf{X} \in \Re_+^{F \times N}$ whose columns are F-dimensional feature vectors obtained by scanning each image row-wise. NMF attempts to find two low rank matrices $\mathbf{V}$ and $\mathbf{W}$ that minimize the reconstruction error subject to nonnegativity constraints:

$$\min_{\mathbf{V},\mathbf{W}} ||\mathbf{X} - \mathbf{V}\mathbf{W}||_F^2 \qquad (1)$$
$$\text{s.t.} \quad v_{i,k} \geq 0 \quad , w_{k,j} \geq 0, \quad \forall i, j, k.$$

where matrix $\mathbf{V} \in \Re_+^{F \times M}$ ($M << F$) contains the basis images, while $\mathbf{W} \in \Re_+^{M \times N}$ contains the appropriate linear combination coefficients that reconstruct each original image and $||.||_F$ is the matrix Frobenius norm. Using an appropriately designed auxiliary function, it has been shown in [10] that the following multiplicative rules update $w_{k,j}$ and $v_{i,k}$, yielding the desired factors, while guarantee a non increasing behavior of the cost function:

$$w_{k,j}^{(t)} = w_{k,j}^{(t-1)} \frac{[\mathbf{V}^{(t-1)^T}\mathbf{X}]_{k,j}}{[\mathbf{V}^{(t-1)^T}\mathbf{V}^{(t-1)}\mathbf{W}^{(t-1)}]_{k,j}}, \qquad (2)$$

$$v_{i,k}^{(t)} = v_{i,k}^{(t-1)} \frac{[\mathbf{X}\mathbf{W}^{(t)^T}]_{i,k}}{[\mathbf{V}^{(t-1)}\mathbf{W}^{(t)}\mathbf{W}^{(t)^T}]_{i,k}}. \qquad (3)$$

Assuming that the decomposed data are not static images but a time varying sequence, as for instance the frames $\mathbf{x}_t \in \Re^F$ of a video sequence where $t \in [1, N]$ denotes time, SFA seeks to determine appropriate projection bases stored in the columns of matrix $\mathbf{V} \in \Re^{F \times M}$, that extract the slowest varying features. To do so, SFA attempts to minimize the variance of the approximated first order time derivative of the latent variables $\mathbf{W} \in \Re^{M \times N}$ subject to zero mean, unit covariance and decorrelation constraints:

$$\min_{\mathbf{V}} \quad tr[\dot{\mathbf{W}}\dot{\mathbf{W}}^T] \qquad (4)$$
$$\text{s.t. } \mathbf{W}\mathbf{1} = \mathbf{0}, \ \mathbf{W}\mathbf{W}^T = \mathbf{I}$$

where $tr[.]$ is the matrix trace operator, $\mathbf{1}$ is a $N \times 1$ vector with all its elements equal to $\frac{1}{N}$ and $\mathbf{I}$ is an $M \times M$ identity matrix. Matrix $\dot{\mathbf{W}}$ approximates the first order time derivative of $\mathbf{W}$, evaluated using the forward latent variable differences as $\dot{\mathbf{W}} = \mathbf{W}\mathbf{P}$ where $\mathbf{P}$ is an $N \times (N-1)$ matrix with elements $\mathbf{P}_{i,i} = -1$ and $\mathbf{P}_{i+1,i} = 1$.

Considering the linear case where the latent space can be derived by projecting the input samples on a set of basis as

$\mathbf{W} = \mathbf{V}^T\mathbf{X}$ and assuming that input data have been normalized such as to have zero mean, the SFA problem in (4) can be reformulated to the following trace optimization problem:

$$\min_{\mathbf{V}} \quad tr[\mathbf{V}^T\mathbf{A}\mathbf{V}], \text{ s.t. } \mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{I} \qquad (5)$$

where $\mathbf{B}$ is the data covariance matrix and $\mathbf{A}$ is a covariance matrix evaluated using the forward temporal differences of the input data, contained in matrix $\dot{\mathbf{X}}$:

$$\mathbf{A} = \frac{1}{N-1}\dot{\mathbf{X}}\dot{\mathbf{X}}^T = \frac{1}{N-1}\mathbf{X}\mathbf{L}\mathbf{X}^T, \quad \mathbf{B} = \frac{1}{N}\mathbf{X}\mathbf{X}^T, \quad (6)$$

where $\mathbf{L} = \mathbf{P}\mathbf{P}^T$. As it has been shown in [2] the solution of (5) can be found from the Generalized Eigenvalue Problem $\mathbf{A}\mathbf{V} = \mathbf{B}\mathbf{V}\mathbf{\Lambda}$ where the columns of the projection matrix $\mathbf{V}$ are the generalized eigenvectors associated with the M-lowest generalized eigenvalues contained in the diagonal matrix $\mathbf{\Lambda}$.

3. SLOW FEATURES NONNEGATIVE MATRIX FACTORIZATION

Next we first present the SFNMF optimization problem and demonstrate that this naturally arises by embedding the SFA trace optimization problem in the nonnegative subspace learning framework. Subsequently, we derive multiplicative update rules for SFNMF optimization.

3.1. SFNMF Optimization Problem

Assuming that we have whitened our data such as to have zero mean and unit variance, which can be performed by applying PCA on $\mathbf{X}$ [2], the constrained trace optimization problem in (5) is equivalent to simultaneously requiring the minimization of term $tr[\mathbf{V}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{V}]$ and the maximization of term $tr[\mathbf{V}^T\mathbf{X}\mathbf{X}^T\mathbf{V}]$. This can be expressed by the following minimization problem:

$$\min_{\mathbf{V}} \quad tr[\mathbf{V}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{V}] - tr[\mathbf{V}^T\mathbf{X}\mathbf{X}^T\mathbf{V}]. \qquad (7)$$

However, optimizing term $tr[\mathbf{V}^T\mathbf{X}\mathbf{X}^T\mathbf{V}]$ with respect to $\mathbf{V}$ is equivalent to optimizing $\frac{1}{2}||\mathbf{X} - \mathbf{V}\mathbf{V}^T X||_F^2$ since:

$$\frac{1}{2}||\mathbf{X} - \mathbf{V}\mathbf{V}^T X||_F^2 = tr[\mathbf{X}\mathbf{X}^T] - tr[\mathbf{V}^T\mathbf{X}\mathbf{X}^T\mathbf{V}] \qquad (8)$$

and term $tr[\mathbf{X}\mathbf{X}^T]$ is independent of $\mathbf{V}$. Consequently, considering the linear case where $\mathbf{W} = \mathbf{V}^T\mathbf{X}$ and incorporating nonnegativity constraints on the elements of matrices $\mathbf{V}$ and $\mathbf{W}$ we derive the proposed SFNMF problem that aims to identify slow varying basic components:

$$\min_{\mathbf{V},\mathbf{W}} \quad \frac{1}{2}||\mathbf{X} - \mathbf{V}\mathbf{W}||_F^2 + \lambda tr[\mathbf{W}(\mathbf{L}^+ - \mathbf{L}^-)\mathbf{W}^T] \ (9)$$
$$\text{s.t.} \qquad v_{i,k} \geq 0 \quad , w_{k,j} \geq 0, \quad \forall i, j, k,$$

where λ is a positive constant. Moreover, since matrix $\mathbf{L}$ contains both positive and negative elements we have expressed it as the difference of the nonnegative matrices $\mathbf{L}^+$ and $\mathbf{L}^-$ to ensure that the subsequently derived update rules cannot assign negative values to the updated elements.

3.2. Multiplicative Update Rules for SFNMF Optimization

To solve the SFNMF constrained optimization problem in (9) we introduce the Lagrangian multipliers $\boldsymbol{\phi} = [\phi_{i,k}] \in R^{F \times M}$ and $\boldsymbol{\psi} = [\psi_{j,k}] \in R^{M \times N}$ each one associated with constraints $v_{i,k} \geq 0$ and $w_{k,j} \geq 0$, respectively. Thus the Lagrangian function $\mathcal{L}$ is formulated as:

$$
\begin{aligned}
\mathcal{L} &= \frac{1}{2}tr[\mathbf{X}\mathbf{X}^T] - tr[\mathbf{V}\mathbf{W}\mathbf{X}^T] + \frac{1}{2}tr[\mathbf{V}\mathbf{W}\mathbf{W}^T\mathbf{V}^T] \\
&+ \lambda tr[\mathbf{W}(\mathbf{L}^+ - \mathbf{L}^-)\mathbf{W}^T] + tr[\boldsymbol{\phi}\mathbf{V}^T] + tr[\boldsymbol{\psi}\mathbf{W}^T]. (10)
\end{aligned}
$$

Consequently, the optimization problem in (9) is equivalent to the minimization of the Lagrangian. To minimize $\mathcal{L}$, we first obtain its partial derivatives with respect to $v_{i,j}$ and $w_{i,j}$ and set them equal to zero:

$$
\frac{\partial \mathcal{L}}{\partial v_{i,k}} = -[\mathbf{X}\mathbf{W}^T]_{i,k} + [\mathbf{V}\mathbf{W}\mathbf{W}^T]_{i,k} + \phi_{i,k} = 0 \quad (11)
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial w_{k,j}} &= [\mathbf{V}^T\mathbf{V}\mathbf{W}]_{k,j} - [\mathbf{V}^T\mathbf{X}]_{k,j} + 2\lambda[\mathbf{W}\mathbf{L}^+]_{k,j} \\
&- 2\lambda[\mathbf{W}\mathbf{L}^-]_{k,j} + \psi_{k,j} = 0. \quad (12)
\end{aligned}
$$

Using the KKT conditions it holds that $\phi_{i,k}v_{i,k} = 0$ and $\psi_{k,j}w_{k,j} = 0$. Thus, solving equation (11) for $v_{i,k}$ leads to the multiplicative update of conventional NMF algorithm shown in (3), since the incorporated term is independent of $\mathbf{V}$. On the other hand, solving (12) for $w_{k,j}$ we derive the proposed multiplicative update rule:

$$
w_{k,j}^{(t)} = w_{k,j}^{(t-1)} \frac{[\mathbf{V}^{(t-1)^T}\mathbf{X}]_{k,j} + +2\lambda[\mathbf{W}^{(t-1)}\mathbf{L}^-]_{k,j}}{[\mathbf{V}^{(t-1)^T}\mathbf{V}^{(t-1)}\mathbf{W}^{(t-1)}]_{k,j} + 2\lambda[\mathbf{W}^{(t-1)}\mathbf{L}^+]_{k,j}}. \quad (13)
$$

Since the derived update rule for $\mathbf{V}$ is the same as in (3) we can recall the proof in [10] to show that the cost function of SFNMF is non-increasing under this update. Regarding the proposed update in (13) the detailed proof is omitted here due to space limitations. However, this can be easily derived similarly to that in [7].

4. EXPERIMENTAL RESULTS

We compared the proposed method against NMF, GNMF [7] and SFA for unsupervised facial behaviour analysis. More precisely, we investigated how effectively each method can detect the transitions between the temporal phases during different facial AUs activation. In general, when activating an AU, the following temporal phases are recorded: Neutral, when the face is relaxed, Onset, when the action initiates, Apex, when the muscles reach the peak intensity and Offset when the muscles begin to relax. The action finally ends with Neutral. Experiments were conducted on the publicly available MMI database [11] which consists of more than 400 videos annotated in terms of facial AUs activations and their temporal phases. All facial images acquired by MMI video sequences have been aligned and scaled to a fixed size of 169×171 pixels. Image alignment was performed by warping the facial images based on 68 landmark points obtained by tracking each subjects facial expression formation.

For each method in the comparison we performed a validation step using part of the available data in order to fine tune the involved parameters. Thus, for GNMF we considered a 5-nearest neighbors graph to capture the local geometric structure of data, a $0 - 1$ weighting system for defining the weight matrix and set parameter λ that regulates the contribution of the two parts in GNMF cost function to 150. Finally, for all algorithms we considered projection to a subspace of equal dimensionality which was set to 50, while the proposed method, NMF and GNMF algorithms were iteratively trained until convergence determined by monitoring the objective function improvement between successive iterations.

To facilitate the comparison between the considered algorithms and the ground truth, we map the recovered latent space by each method to the temporal phases of AUs. This is done by finding for each method the slowest varying latent feature of the 50 extracted. To identify this we compute the first order derivative for each obtained latent variable and select the one that minimizes: $\operatorname{argmin}_i \mathbf{w}_i\mathbf{L}\mathbf{w}_i^T$. We should note that since SFA introduces an ordering to the derived latent variables sorted by their temporal slowness, we simply acquire the first identified latent feature which corresponds to the slowest varying one.

Fig. 1 shows the performance of the examined methods in terms of capturing the AU temporal phases on two video sequences displaying the activation of two different AUs. More precisely, the results presented in Fig. 1(a) correspond to a video sequence where the subject performs AU 27 (i.e. mouth stretch), while results shown in Fig. 1(b) correspond to the activation of AU 43 (i.e. eyes closed). In each plot the ground truth instances when the AUs temporal phases transition appear are highlighted with red marks. As can be observed in both videos the proposed method outperforms both GNMF and SFA since it detects the temporal phases more accurately. Moreover, NMF was not able to detect AU phases transition on both videos.

5. CONCLUSION

In this paper, we proposed a novel algorithm called Slow Features Nonnegative Matrix Factorization that aims to learn slow varying parts-based representations of time varying se-
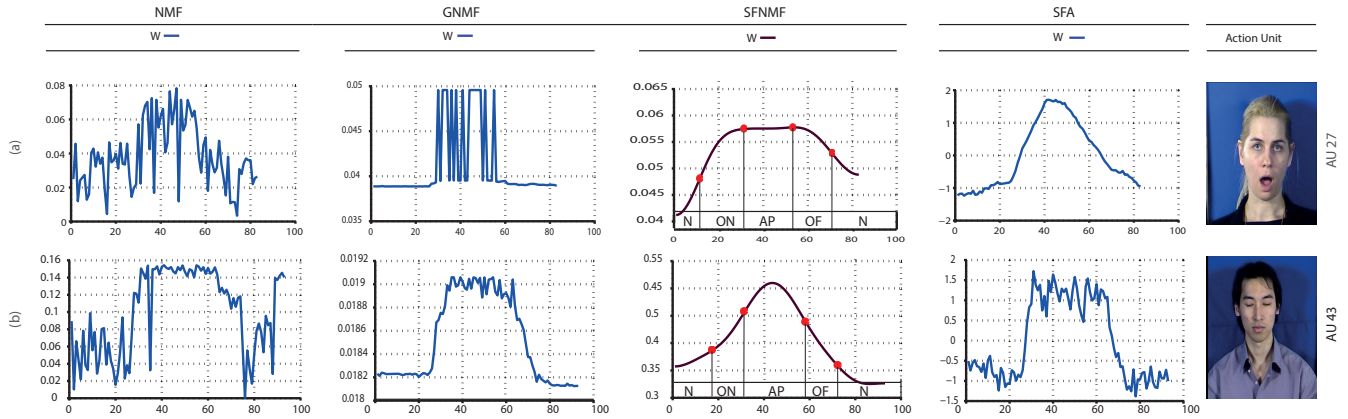
Fig. 1. Obtained results by applying the proposed method, NMF, GNMF and SFA on a video sequence from the MMI dataset displaying a subject performing: (a) Mouth stretch (AU 27) and (b) Eyes closed (AU 43). The red marks indicate the annotated ground truth where the AU temporal phase changes (N - Neutral phase, ON - Onset phase, AP - Apex phase, OF - Offset phase).

quences. The proposed method attempts to simultaneously minimize the data reconstruction error and the temporal variance of the derived latent features. For SFNMF optimization we derived novel multiplicative update rules and verified its superiority against NMF, GNMF and SFA for unsupervised facial behaviour dynamics analysis on MMI database.

7. REFERENCES

[1] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[2] Laurenz Wiskott and Terrence J Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural computation*, vol. 14, no. 4, pp. 715–770, 2002.

[3] Mathias Franzius, Henning Sprekeler, and Laurenz Wiskott, "Slowness and sparseness lead to place, head-direction, and spatial-view cells," *PLoS Computational Biology*, vol. 3, no. 8, pp. e166, 2007.

[4] Lazaros Zafeiriou, Mihalis A Nicolaou, Stefanos Zafeiriou, Symeon Nikitidis, and Maja Pantic, "Learn-ing slow features for behaviour analysis," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2013.

[5] X Niyogi, "Locality preserving projections," in *Neural information processing systems*, 2003, vol. 16, pp. 234–241.

[6] Henning Sprekeler, "On the relation of slow feature analysis and laplacian eigenmaps," *Neural computation*, vol. 23, no. 12, pp. 3287–3302, 2011.

[7] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang, "Graph regularized nonnegative matrix factorization for data representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1548–1560, 2011.

[8] Naiyang Guan, Dacheng Tao, Zhigang Luo, and Bo Yuan, "Manifold regularized discriminative non-negative matrix factorization with fast gradient descent," *Image Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 2030–2048, 2011.

[9] Symeon Nikitidis, Anastasios Tefas, Nikos Nikolaidis, and Ioannis Pitas, "Subclass discriminant nonnegative matrix factorization for facial image analysis," *Pattern Recognition*, vol. 45, no. 12, pp. 4080–4091, 2012.

[10] Daniel D. Lee and H. Sebastian Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems (NIPS)*, 2000, pp. 556–562.

[11] M. Pantic, M. F. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in