# Sparse Symmetric Nonnegative Matrix Factorization Applied to Face Recognition

Hennadii Dobrovolskyi[1] , Nataliya Keberle[2] , Yehor Ternovyy[3]

Chair of Computer Science, Zaporizhzhya National University, 66 Zhukovskogo str, Zaporizhzhya, Ukraine

[1] gen.dobr@gmail.com    [2] nkeberle@gmail.com    [3] yehorternovyy@gmail.com

*Abstract* **– The task of Sparse Symmetric Nonnegative Matrix Factorization(SSNMF) is formulated as optimization problem and solved numerically with the method of projected gradients descent. The adjustable sparsity level allows to emphasize the most significant object features. Clustering of the Yale Faces data set shows that SSNMF provides the same level of quality as common clustering approaches.**

*Keywords* **– face recognition; symmetric nonnegative matrix factorization; sparsity condition**

## I. Introduction

A lot of problems in data analysis, data mining, machine learning, signal and image processing, statistics and graph theory can be transformed into non-negative matrix factorization (NMF) problem [1], [2] providing low-dimensional approximation of a high-dimensional data. Common NMF problem is to find $W$ and $H$ such that $A \approx WH^T$, where $A \in R^{n \times m}$ is a nonnegative matrix, $W \in R^{m \times r}$, $H \in R^{n \times r}$, $W_{ij} \geq 0$, $H_{ij} \geq 0$, and $r$ is some integer value $r \ll min(n, m)$. Added nonnegativity constraints lead to a better interpretability of the factors, e.g., when analyzing images or documents [3].

Typically the NMF is formulated as optimization problem

$$\left\| A - WH^T \right\|_F^2 \rightarrow min, W_{ij} \geq 0, H_{ij} \geq 0 \qquad (1)$$

where $\left\| Z \right\|_F^2$ denotes squared Frobenius norm of a matrix $Z$.

Non-negative sparse coding [4] is a method for decomposing multivariate data into non-negative sparse components. The goal of Sparse NMF is to find a decomposition in which one of the factors is sparse, meaning that it has few nonzero elements. This means that it depends only on a few significant parameters improving human interpretability of the results. Usually the sparsity is achieved with adding extra term to objective function to be minimized [4]:

$$\left\| A - WH^T \right\|_F^2 + \lambda \sum_{i,j} |H_{ij}| \rightarrow min,$$
$$W_{ij} \geq 0, H_{ij} \geq 0 \qquad (2)$$

or

$$\left\| A - WH^T \right\|_F^2 + \lambda \sum_{i,j} |W_{ij}| \rightarrow min,$$
$$W_{ij} \geq 0, H_{ij} \geq 0 \qquad (3)$$

where the balance between sparseness and factorization accuracy depends on the parameter $\lambda$. What must be addressed here is which factor matrix, $W$ or $H$, is selected as the candidate on which the sparseness constraint is imposed. It depends on the nature of a real-world problem being solved.

The presented study considers a special case of NMF - Symmetric NMF (SNMF), where the input matrix $A$ is symmetric. For instance, its' elements $A_{ij}$ can present a similarity between $i$-th and $j$-th items of a data set. The similarity measure can be quite different in different cases allowing great variability of the method.

The SNMF optimization problem can be obtained from (2) by setting $W = H$:

$$\left\| A - HH^T \right\|_F^2 + \lambda \sum_{i,j} |H_{ij}| \rightarrow min, H_{ij} \geq 0 \qquad (4)$$

SNMF is used successfully in different fields and was proved to provide the fine results as well as standard clustering techniques such as k-means [5], normalized cut, spectral clustering [6], [7], etc. However to the best of our knowledge there is no published attempts to achieve sparsity in SNMF providing more understandable and compact results.

The objective of the current study is to implement and test the method of Sparse SNMF (SSNMF) to face recognition, and to compare renowned clustering algorithms such as k-means and spectral clustering with SSNMF.

In this paper we formulate and solve the problem of Sparse Symmetric Nonnegative Matrix Factorization (SSNMF). The obtained solution can be used as tool of dimensionality reduction and clustering that leads to more compact object representation because of sparsity. Using Yale Faces database [1] we show that SSNMF can compete with common approaches like spectral clustering. Its distinctive feature is a controlled level of sparsity

---

[1] Yale Face Database – http://vision.ucsd.edu

introduced into common Symmetric Nonnegative Matrix Factorization (SNMF) that turns the results into more compact and interpretable form highlighting the most significant feature combinations.

## II. METHOD DESCRIPTION

Sparse SNMF problem can be formulated in the following way:

$$\left\|A - HH^T\right\|_F^2 + \lambda \sum_{i,j} |H_{ij}| \to min,$$
$$H_{ij} \geq 0, A_{ij} = A_{ji} \tag{5}$$

where the parameter $\lambda$ affects both sparsity level and factorization accuracy.

Dimensionality of the factor $H_{ij}$ is $N \times r$, where $N$ is number of elements in dataset and $r$ is a suggested number of clusters.

The matrix $A$ is assumed to be symmetric similarity matrix.

The problem (5) is similar to sparse NMF [4]. However, the distinguishing features are as follows: first - added is the requirement of symmetry of matrix to be approximated; second – the symmetric NMF objective function does not contain the sparsity restriction term $\lambda \sum_{i,j} |H_{ij}|$. So the formulation of the problem (5) is new.

Similarly to other optimization problems it can be solved with projected gradient descent approach [8] which consists of following update rule

$$H_{ab}^{(n+1)} = max \left(0, H_{ab}^{(n)} + \delta \nabla_{ab}\right) \tag{6}$$

where $\nabla_{ab}$ is a gradient of objective function (5) that takes into account symmetry of the matrix $A$:

$$\nabla_{ab} = 4 \left[\sum_j \left(\sum_p H_{ap}H_{jp} - A_{aj}\right) H_{jb}\right] + 1 \tag{7}$$

and parameter $\delta$ is a variable step size which is gradually decreased during iterations.

After the factor $H$ is found the cluster number for element $j$ can be found by searching position of maximal element in $j$-th row of the matrix $H$.

## III. APPLICATION OF SSNMF TO FACE RECOGNITION

Identification of person based on facial image, unlike other biometric identification indicators (fingerprints, irises etc.), does not require physical contact with a scanning device. Considering the rapid development of digital technologies it is most suitable for mass application.

Face recognition is a difficult task due to the following reasons:

- person's face is not a static object and it has high degree of changeability in appearance (for example by form and skin color);
- different conditions of light (luminosity), which defined by type, direction and number of light sources;
- overlapping of persons by other objects in the scene;
- the need to localization and face recognition with random position in space.

Authentication process based on a given facial image refers to finding one of mutually exclusive known classes (persons) which the facial image belongs to. Process can decline if the provided facial image does not belong to known classes.

Face recognition system consists of modules of face localization at photo, image alignment based on anthropometric points, retrieving a vector of face features from image, classifier training and authentication a facial image.

Identification process is divided into three steps. At the first step we preprocess images by alignment and crop facial image by using an algorithm called Face Landmark Estimation [9]. At the second step we extract facial image features from "aligned" images and save data to CSV file. At the third step clustering and prediction procedure is performed to identify the most relevant class the facial image belongs to.

### A. Face detection

The first need is a robust feature set that allows the face of person to be discriminated cleanly, even in cluttered backgrounds under difficult illumination. DLib C++ library [2] provides methods for implementing objects search on an image using descriptors of histogram oriented gradients (HOGs) [10], in particular human faces. To detect faces on the resulting HOG image, after processing it with a detector, we need to find a section of the image that is most similar to the known HOG structure obtained from the group of persons used for training.

### B. Face features extraction

Facial landmarks are used to localize and represent salient regions of the face, such as eyes, eyebrows, nose, mouth, jawline. Shape predictor attempts to localize key points of interest along the shape. The pre-trained facial landmark detector inside Dlib library is used to estimate the location of 68 (x,y)-coordinates that map to facial structures on the face. Face recognition model inside Dlib library is used to compute the 128 dimensional vector that describes the face in image identified by shape. Also, the model has an accuracy of 99.38% on the standard Labeled Faces in the Wild benchmark. This is comparable to other state-of-the-art models and means that, given two face images, it correctly predicts if the images are of the same person 99.38% of the time [11].

### C. Application of SSNMF to clustering face images

To evaluate SSNMF clustering approach we have used the Yale Face Database containing 165 grayscale images in GIF format of 15 individuals. There are 11 images

---

per subject, one per different facial expression or configuration: center-light, with glasses, happy, left-light, with no glasses, normal, right-light, sad, sleepy, surprised, and wink. The method of SSNMF described above requires a reasonable definition of the similarity matrix. Here we use a heat kernel regarding Euclidean distance between two feature vectors as "energy":

$$A_{ij} = exp\left(-\beta \left\| f_i - f_j \right\|\right) \qquad (8)$$

where $f_i$ is a feature vector of $i$-th object of dataset. Below we set $\beta = 1$ .

As soon as the similarity matrix $A$ is defined the SS-NMF procedure (6), (7) can be run. To illustrate clustering quality the content of some clusters is shown on Fig. 1 - Fig. 3
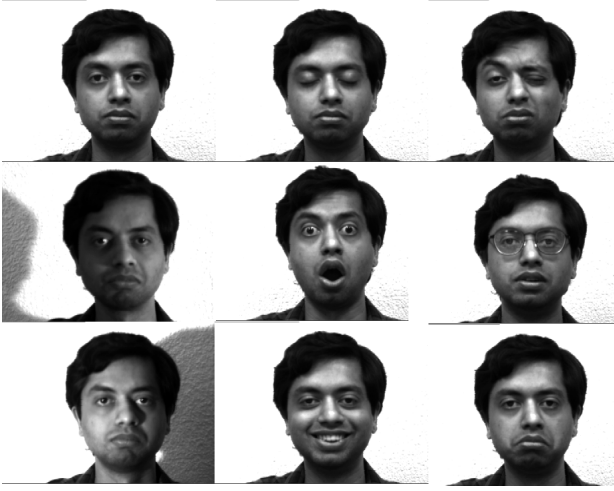


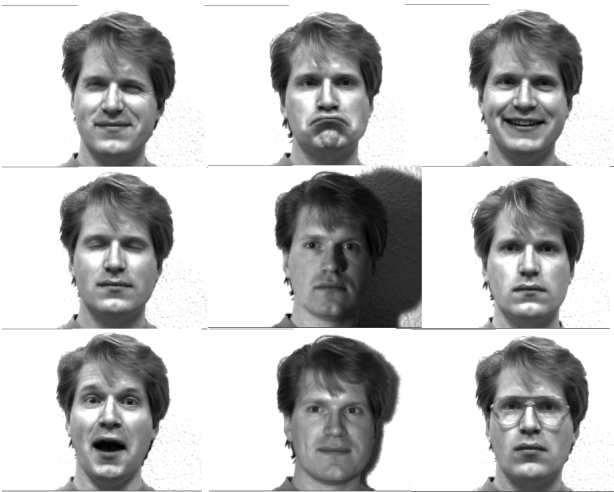Figure 1. Face images assigned to cluster 1



Figure 2. Face images assigned to cluster 19

The complete review of the clusters shows that in most cases the same person images are assigned to the
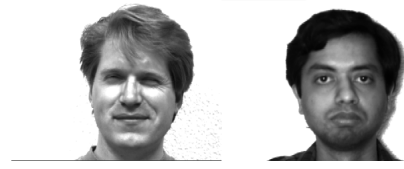


Figure 3. Face images assigned to cluster 10

same cluster though sometimes the different persons can form small separate clusters because of feature extraction inaccuracy.

To estimate clustering quality we calculate Adjusted Rand Index (ARI), homogeneity score and completeness score [12]. Adjusted Rand Index (ARI) [13] in statistics, and in particular in data clustering, is a measure of the similarity between two data clusterings. It may be interpreted as the chance of grouping of elements in the first clustering as well as in the second one.

Homogeneity shows the chance that each cluster contains only members of a single class and completeness is related to the probability of the fact that the members of a given class are assigned to the same cluster.

Table 1 shows the ARI, homogeneity score and completeness score for SSNMF, k-means and spectral clustering applied to Yale Faces database. The k-means and spectral clustering were performed using Scikit-Learn Python library [3].

Table 2 shows the same measures for SSNMF and different level of sparsity.

Table I. CROSS EVALUATION OF SSNMF, k-MEANS AND SPECTRAL CLUSTERING FOR FACE RECOGNITION

| Method | Scores | | |
|---|---|---|---|
| | *ARI* | *Homogeneity* | *Completeness* |
| SSNMF ($\lambda = 0.1$) | 0.746 | 0.987 | 0.830 |
| k-means | 0.729 | 0.979 | 0.822 |
| spectral clustering | 0.656 | 0.920 | 0.774 |

Table II. CLUSTERING QUALITY AT DIFFERENT LEVELS OF SPARSITY

| Sparsity parameter | Scores | | |
|---|---|---|---|
| | *ARI* | *Homogeneity* | *Completeness* |
| $\lambda = 0$ | 0.753 | 0.983 | 0.833 |
| $\lambda = 0.01$ | 0.723 | 0.987 | 0.837 |
| $\lambda = 0.02$ | 0.758 | 0.978 | 0.842 |
| $\lambda = 0.03$ | 0.721 | 0.987 | 0.819 |
| $\lambda = 0.04$ | 0.726 | 0.981 | 0.820 |
| $\lambda = 0.10$ | 0.746 | 0.987 | 0.830 |

## IV. CONCLUSIONS

The suggested sparse modification of symmetric matrix factorization can compete in clustering quality with other well-known algorithms. The bottleneck of the method is

[3]http://scikit-learn.org

its low scalability because it requires at least $O(N^2)$ data storage to keep the similarity matrix. But at the same time it has the wide range of applications utilizing any type of object similarity. Also it allows adjusting the level of sparsity on account of factorization accuracy leading to easier understanding and robustness.

## REFERENCES

[1] N. Gillis, "Introduction to nonnegative matrix factorization," *arXiv preprint arXiv:1703.00663*, 2017.

[2] M. Udell, C. Horn, R. Zadeh, S. Boyd *et al.*, "Generalized low rank models," *Foundations and Trends in Machine Learning*, vol. 9, no. 1, pp. 1–118, 2016.

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

[4] P. O. Hoyer, "Non-negative sparse coding," in *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*. IEEE, 2002, pp. 557–565.

[5] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA., 1967, pp. 281–297.

[6] C. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering," in *Proceedings of the 2005 SIAM International Conference on Data Mining*. SIAM, 2005, pp. 606–610.

[7] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336–1353, 2013.

[8] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural computation*, vol. 19, no. 10, pp. 2756–2779, 2007.

[9] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874.

[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[11] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[12] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *Journal of Machine Learning Research*, vol. 11, no. Oct, pp. 2837–2854, 2010.

[13] L. Hubert and P. Arabie, "Comparing partitions," *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.