IMAGE REPRESENTATION USING SUPERVISED AND UNSUPERVISED LEARNING METHODS ON COMPLEX DOMAIN

Manh-Quan Bui[1], Viet-Hang Duong[1], Yung-Hui Li[1], Tzu-Chiang Tai[2], and Jia-Ching Wang[1]

[1]Dept. of Computer Science and Information Engineering, National Central University, Jhongli, Taiwan
[2]Dept. of Computer Science and Information Engineering, Providence University, Taichung, Taiwan

ABSTRACT

Matrix factorization (MF) and its extensions have been intensively studied in computer vision and machine learning. In this paper, unsupervised and supervised learning methods based on MF technique on complex domain are introduced. Projective complex matrix factorization (PCMF) and discriminant projective complex matrix factorization (DPCMF) present two frameworks of projecting complex data to a lower dimension space. The optimization problems are formulated as the minimization of the real-valued functions of complex variables. Motivated by independence among extracted features, Fisher linear discriminant is used as hard constraint on supervised model. Experimental results on facial expression recognition (FER) show improved classification performance in comparison to real-valued features of both unsupervised and supervised NMFs.

Index Terms— Complex matrix factorization, discriminant feature, image representation, NMF, LDA

1. INTRODUCTION

Image representation (IR) is one of the most important issues in computer vision and pattern recognition. The common object of an IR system is to transform the input signal into a new representation which reduces its dimensionality and explicates its latent structures. One of fundamental tools for IR is matrix factorization (MF) techniques that decomposes the original matrix into two or more matrix factors. Among matrix factorization methods, nonnegative matrix factorization (NMF) [1, 2] is known for its parts-based representation for images which has psychological and physiological interpretation. As a variant of NMF, a low-dimensional compact representation on complex domain, exemplar-embed complex matrix factorization (EE-CMF) [3] reconstructs a target matrix based on a complex matrix factorization model. There are two stages in structured EE-CMF for image representation including transforming the pixel intensive values to complex numbers and decomposing the transferred complex data matrix into an exemplar-embed base and a new coefficient complex matrix. EE-CMF shows its potential and promising performance for data representation.

However, both NMF and EE-CMF are the unsupervised learning algorithms which are not able to incorporate the label information. In order to extract the features that better for image representation, some extensions of NMF were proposed to utilize the association between the observations and its class information. Graph regularized nonnegative matrix factorization (GNMF) [4, 5], a semi-supervised NMF, encoded the geometrical information of the data space by constructing the nearest neighbor graph. Supervised learning model, discriminant NMF (DNMF) [6] tried to maximize the discriminative ability of learned features through integrating Fisher's criterion [7] into NMF. Encouraged by the great success of the supervised NMFs as well as MF techniques in complex field for image representation, we develop two complex matrix factorization models referred to as projective complex matrix factorization (PCMF) and discriminant projective complex matrix factorization (DPCMF). The proposed approaches adopt the concept of the cosine dissimilarity metric [8] to minimize the error reconstruction on decomposing a real matrix and integrate the Fisher's criterion to get the optimal discriminate feature in a supervised case. It is witnessed that the cosine divergence in the real field is exploited implicitly by its equivalence with squared Euclidean divergence in the complex domain [3, 9-11]. An objective function of real variables with the cosine dissimilarity distance is simplified by a corresponding function of complex variables. The proposed PCMF and DPCMF algorithms for image representation are designed with two main phases similar to the work in [3]. Firstly, the raw pixel intensive values of the original images are transferred to complex numbers via Euler formula [12] and vectored into a complex data matrix. Secondly, complex matrix decomposition is processed based on optimizing the real-valued function of complex variables.

In summary, the contributions of this paper are three-folds:

1. We introduce the supervised and unsupervised learning frameworks for image representation (PCMF and DPCMF). Our proposed algorithms are more efficient and flexible than real NMF models and can provide much

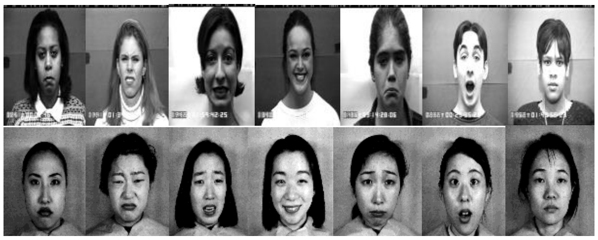

Fig. 1. Sample images of seven facial expressions from the CK+ dataset [18] (first row) and JAFFE dataset [19] (second row).

intuitive recognition results. Without limiting the sign of data, the developed methods are able to be applied on real-world applications, particularly the field of complex-valued data processing, such as communication and acoustic, etc.

2. To the best of our knowledge, DPCMF is the first work on the supervised learning using complex matrix factorization in which the label information is used as hard constraints. In this supervised learning framework, label knowledge is integrated into loss function, making the learned representation as discriminative as possible.

3. Our approaches are practically effective on facial expression recognition (FER) task. Experimental results on two popular facial expression image datasets, JAFFE and CK+, confirm the effectiveness of the proposed models comparing with the complex EE-CMF and the representative unsupervised/supervised NMF algorithms.

2. EULER FORMULATION FOR SPACE TRANSFORMATION

It is proved that the cosine dissimilarity metric in the real domain is robust to suppress outliers and there is an equivalence between the cosine-based distance measures on the real field and the squared Euclidean norm on the complex field [3, 9-11]. According to these observations, we first perform space transformation as described as follows.

Consider the $N\times M$ matrix $\mathbf{X}$ containing pixel intensive values of M original images. Each image was vectored as a column of N elements in $\mathbf{X}$. Similar to the works in [3, 9-11], we construct a map E using the Euler's formulation [12] to convert the real matrix $\mathbf{X}$ into its complex space as follows:

$$E:\mathbb{R}^N \to \mathbb{C}^N \ \text{ s.t } E(\mathbf{X}_j)=\mathbf{Z}_j=\frac{1}{\sqrt{2}}e^{i\gamma\pi\mathbf{X}_j};\ \ \forall j=1..M \quad (1)$$

As a result, the real matrix $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M]$ are converted to a complex matrix which then can be projected to the lower dimension subspace by some complex matrix decomposition methods. We will propose two methods of complex matrix factorization in the next sections.

3. THE PROPOSED METHODS

In this section, we introduce the methods of complex matrix factorization with unsupervised and supervised learning techniques. We assume that the training data are given as a matrix $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_M]$, where $\mathbf{D}_i \in \mathbb{C}^N$, and M is the total number of training samples. Let n_i be the number of vectors in the i^{th} class, and C be the number of classes.

3.1 Projective complex matrix factorization

The unsupervised projective complex matrix factorization (PCMF) projects high-dimensional complex samples $\mathbf{D}$ onto a lower-dimensional subspace spanned by a basis $\mathbf{B}$ and considers $\mathbf{B}^T\mathbf{D}$ as their encoding such that $\mathbf{D}\approx \mathbf{B}\mathbf{B}^T\mathbf{D}$.

The objective function of PCMF is given by:

$$\min_{\mathbf{B}} f_{\text{PCMF}}(\mathbf{B}) = \min \frac{1}{2}\left\|\mathbf{D} - \mathbf{B}\mathbf{B}^H\mathbf{D}\right\|_F^2 \quad (2)$$

It is noticed that

$$\left\|\mathbf{D} - \mathbf{B}\mathbf{B}^H\mathbf{D}\right\|_F^2 = Trace[(\mathbf{D}^H\mathbf{D} - 2\mathbf{D}^H\mathbf{B}\mathbf{B}^H\mathbf{D} + \mathbf{D}^H\mathbf{B}\mathbf{B}^H\mathbf{B}\mathbf{B}^H\mathbf{D})] \quad (3)$$

3.2 Discriminant projective complex matrix factorization

The supervised discriminant projective complex matrix factorization (DPCMF) integrates Fisher's criterion [6] into PCMF to utilize the label information. In the lower-dimensional subspace, DPCMF expects minimizing the distance between any two samples of the same class and meanwhile maximizing the distance of the samples in different classes. In other words, the extracted features of the supervised learning algorithm DPCMF have better discriminant ability. The cost function of DPCMF is given by:

$$\min_{\mathbf{B}} f_{\text{DPCMF}}(\mathbf{B}) = \min \frac{1}{2}\left\|\mathbf{D} - \mathbf{B}\mathbf{B}^H\mathbf{D}\right\|_F^2 + \frac{1}{2}\alpha Trace(\mathbf{B}^H(\lambda\mathbf{S}_w - \mathbf{S}_b)\mathbf{B}) \quad (4)$$

Herein α and λ are hyperparameters to obtain the tradeoff between the accuracy of the reconstruction error and two discriminative terms, $\mathbf{S}_w$ is the within-class scatter matrix and $\mathbf{S}_b$ is the between-class scatter matrix. The formulation of $\mathbf{S}_w$ and $\mathbf{S}_b$ are defined as follows:

$$\mathbf{S}_w = \frac{1}{C}\sum_{i=1}^{C}\frac{1}{n_i}\sum_{j=1}^{n_i}(\mathbf{d}_j - \mu_i)(\mathbf{d}_j - \mu_i)^T \quad (5)$$

$$\mathbf{S}_b = \frac{1}{C(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{C}(\mu_i - \mu_j)(\mu_i - \mu_j)^T \quad (6)$$

where μ_i denotes the mean values of class i in $\mathbf{D}$ and $\mu_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{d}_j$.

3.3 Gradient descent method for optimal solutions

It can be seen that equations (2) and (4) are optimization problems of a real-valued function with one complex variable **B**. The complex gradient descent algorithm [3, 10, 13] is employed by starting with the randomly initialized matrix **B** and applying the following update rules:

$$\mathbf{B}^{(t+1)} = \mathbf{B}^{(t)} - 2\beta_t \nabla_{\mathbf{B}^*} f(\mathbf{B}^{(t)}) \tag{7}$$

The Writinger's calculus [14-15] is used to evaluate the gradient in the following form:

$$\nabla_{\mathbf{B}^*} f(\mathbf{B}) = \frac{\partial f(\mathbf{B})}{\partial(\mathrm{Re}\,\mathbf{B})} + i\frac{\partial f(\mathbf{B})}{\partial(\mathrm{Im}\,\mathbf{B})} \tag{8}$$

The resulting gradients for PCMF and DPCMF are given by equations (9) and (10), respectively

$$\nabla_{\mathbf{B}^*} f_{PCMF}(\mathbf{B}) = -2\mathbf{D}^H\mathbf{D}\mathbf{B} + \mathbf{B}\mathbf{B}^H\mathbf{D}\mathbf{D}^H\mathbf{B} + \mathbf{D}\mathbf{D}^H\mathbf{B}\mathbf{B}^H\mathbf{B} \tag{9}$$

$$\nabla_{\mathbf{B}^*} f_{DPCMF}(\mathbf{B}) = -2\mathbf{D}^H\mathbf{D}\mathbf{B} + \mathbf{B}\mathbf{B}^H\mathbf{D}\mathbf{D}^H\mathbf{B} + \mathbf{D}\mathbf{D}^H\mathbf{B}\mathbf{B}^H\mathbf{B} + \alpha(\lambda S_w - S_b)\mathbf{B} \tag{10}$$

We utilized the first-order Taylor series expansion [15] for the real-differentiable function $f(\mathbf{B})$ and the backtracking-Armijo search [17] to estimate the step size β_t.

It is known that following equation (11) is true.

$$\Delta f(\mathbf{B}) \approx \left\langle \nabla_{\mathbf{B}} f(\mathbf{B}), \Delta\mathbf{B}^* \right\rangle + \left\langle \nabla_{\mathbf{B}^*} f(\mathbf{B}), \Delta\mathbf{B} \right\rangle \tag{11}$$

Therefore, the step size β_t must satisfy $\beta_t = \mu^{s_t}$, where $0 < \mu < 1$, and s_t is the first non-negative integer such that:

$$f(\mathbf{B}^{(t+1)}) - f(\mathbf{B}^{(t)}) \le 2\sigma\,\mathrm{Re}\left\{\left\langle \nabla_{\mathbf{B}^*} f(\mathbf{B}^{(t)}), \mathbf{B}^{(t+1)} - \mathbf{B}^{(t)} \right\rangle\right\} \tag{12}$$

4. EXPERIMENTS

The unsupervised PCMF and supervised DPCMF methods were explored for data representation by learning the base $\mathbf{B}_{train}$ firstly from the training set $\mathbf{D}_{train}$ such that $\mathbf{D}_{train}=\mathbf{B}_{train}(\mathbf{B}_{train})^H\,\mathbf{D}_{train}$. The learned feature $\mathbf{C}_{train}$ then is obtained from production of $(\mathbf{B}_{train})^H$ and $\mathbf{D}_{train}$. The testing phase was carried out by projecting the tested samples $\mathbf{D}_{test}$ onto the feature space and obtaining the encode $\mathbf{C}_{test}=(\mathbf{B}_{train})^H\,\mathbf{D}_{test}$. Finally, $\mathbf{C}_{train}$ and $\mathbf{C}_{test}$ were considered as the input data of a nearest-neighbor (NN) classifier for recognition phase.

Two popular FER datasets, the extended Cohn-Kanade (CK+) [18] and the JAFFE [19], were used for evaluations. Five frames of each labeled sequence in the CK+ dataset were taken and processed as static images. One of them was collected for the training set and the remaining for the testing set. In the JAFFE dataset, one image of each expression per person was selected randomly to set up the training set and using the rest images for testing. Fig. 1 shows sample images from the CK+ and JAFFE datasets.

We validate the effectiveness of the proposed PCMF and DPCMF by comparing them to the most related methods including: the complex model EE-CMF [3], the unsupervised/supervised projective NMF [20-21], and the unsupervised/supervised NMF [1, 6].

The parameters α and λ were tested by starting with small values and changing them stepwise. The practical values were set in the range [0.1, 0.5].

4.1. Recognition Results on the CK+ Dataset

Table 1 reports the average accuracies of the proposed methods and the compared algorithms over different dimensionalities. It shows that if the projection space has higher dimensionality, the better recognition accuracy becomes better. Overall, the complex algorithms consistently outperform the real NMFs, and the method with unsupervised technique is less effectiveness than the method with supervised technique. It is seen that 97.20%, 96.98% and 96.73% recognition accuracies are achieved by DPCMF, PCMF and DPNMF, respectively. The complex EE-CMF just obtains 95.21%. Since obtaining only 34.31% recognition accuracy, DNMF is not useful for extracting discriminative features from CK+ dataset.

To clarify which facial parts are more deterministic on the recognition performance, we computed the confusion matrix of the proposed DPCMF and give results in Table 2. It is observed that, neutral can be classified well with highest accuracy (100%). Because of confusing with sadness (7.35%) and happiness (1.47%), anger is recognized with the lowest accuracy (91.18%). The other five expressions are recognized with accuracy ranged from 94.74% to 98.44%.

TABLE I

FACIAL EXPRESSION RECOGNITION RATE (%) USING THE CK DATASET WITH DIFFERENT SUBSPACE DIMENSIONALITIES

No. Base	DPCMF	PCMF	EE-CMF	DPNMF	PNMF	DNMF	NMF
20	**96.78**	95.95	95.43	95.89	77.25	24.24	85.41
30	**97.31**	97.00	92.25	96.32	80.83	25.19	90.99
40	**97.15**	96.96	91.24	96.69	81.30	28.26	93.88
50	97.11	**97.15**	95.06	96.65	85.32	28.68	94.5
60	**97.31**	97.17	96.14	97.02	84.54	38.74	95.06
70	**97.44**	97.05	96.59	96.80	86.39	38.49	95.18
80	**97.27**	97.21	96.74	97.07	87.47	38.93	95.93
90	**97.25**	97.19	96.63	96.96	86.89	40.48	95.95
100	**97.20**	97.11	96.78	97.17	87.99	45.76	96.03
Ave.	**97.20**	96.98	95.21	96.73	84.22	34.31	93.66

TABLE II

CONFUSION MATRIX (%) OF 7-CLASS FACIAL EXPRESSION RECOGNITION USING DPCMF ON CK+ DATASET

	anger	disgust	fear	happiness	sadness	surprise	neutral
anger	**91.18**	0	0	1.47	7.35	0	0
disgust	3.95	**94.74**	0	0	0	1.32	0
fear	0	0	**98.33**	0	0	1.67	0
happiness	1.00	2.00	0	**97.00**	0	0	0.00
sadness	0	0	0	0	**98.44**	0.00	1.56
surprise	0	0	0	0	2.88	**97.12**	0
neutral	0	0	0	0	0	0	**100**

TABLE III
FACIAL EXPRESSION RECOGNITION RATE (%) USING THE JAFFE DATASET WITH DIFFERENT SUBSPACE DIMENSIONALITIES

No. Base	DPCMF	PCMF	EE-CMF	DPNMF	PNMF	DNMF	NMF
20	**70.42**	69.58	66.99	63.01	50.00	15.31	65.24
30	**70.98**	69.02	66.36	66.78	56.01	14.90	68.11
40	**72.31**	71.26	**72.31**	69.58	57.83	15.10	70.84
50	**72.66**	71.40	72.03	69.58	60.21	15.59	71.68
60	72.24	**72.45**	**72.45**	70.49	57.34	15.66	71.12
70	**72.45**	72.38	72.31	70.07	60.56	15.24	69.79
80	**73.01**	71.75	72.59	71.75	62.03	15.38	26.15
90	72.17	**72.80**	71.68	71.54	63.64	15.45	16.01
100	72.73	72.24	**73.57**	72.31	61.54	17.48	18.60
Ave.	**72.11**	71.43	71.14	69.46	58.80	15.57	53.06

TABLE IV
CONFUSION MATRIX (%) OF 7-CLASS FACIAL EXPRESSION RECOGNITION USING DPCMF ON JAFFE DATASET

	anger	disgust	fear	happiness	sadness	surprise	neutral
anger	**60.00**	20.00	5.00	5.00	0	10.00	0
disgust	5.26	**68.42**	5.26	0	10.53	10.53	0
fear	0	4.55	**45.45**	0	4.55	36.36	9.09
happiness	0	0	9.52	**66.67**	14.29	9.52	0
sadness	0	0	5.00	0	**90.00**	0	5.00
surprise	9.52	9.52	4.76	0	19.05	**57.14**	0
neutral	0	0	10.00	0	5.00	0	**85.00**

4.2. Recognition Results on the JAFFE Dataset

In JAFFE dataset, we used the same parameters as those used in the CK+ dataset. We found that the images in the JAFFE are more challenging. Based on the results on the recognition accuracy (Table III) and the confusion matrix (Table IV), matrix factorization on complex domain are much robust in capturing the texture and salient features of the face. Similar to the results in section 4.1, the highest recognition rates, 72.11%, 71.43% and 71.14%, are obtained by DPCMF, PCMF and EE-CMF. The unsupervised learning algorithms DNMF and NMF have a tendency to extract features with a large dimension, and are not suitable for overfitting case (number of bases are bigger 70). The fear expression has lowest recognition rate (just 45.45%) because it has more overlapping (36/36%) with surprise emotion.

4.2. Recognition Results on the occluded CK+ images

The effectiveness of the proposed approaches is also validated by their performance on the occluded CK+ images. Two images among five static frames per expression of each person were processed so that they had occlusion. Some occluded images are shown in Fig. 2. Creating occlusion/unocclusion on training/testing images were designed as two cases in occluded experiments. The detailed recognition results are shown in Table V and VI in which the best performances belong to DPCMF and PCMF. It can be observed that the occlusion on training set makes the recognition more difficultly than the occluded testing set.

anger disgust fear happiness sadness surprise neutral

Fig. 2. Cropped face images of six facial expressions and neutral with occlusions from the CK+ dataset

TABLE V
FACIAL EXPRESSION RECOGNITION RATE (%) USING THE OCCLUDED CK + IIMAGES (CASE OF OCCLUSION ON TRAINING SET)

No. Base	DPCMF	PCMF	EE-CMF	DPNMF	PNMF	DNMF	NMF
20	45.90	48.26	**53.28**	41.49	16.03	21.49	45.12
30	59.17	55.76	**59.17**	47.82	15.54	21.49	50.22
40	**59.83**	57.85	59.01	51.74	17.30	21.49	52.17
50	**68.43**	60.99	60.06	50.58	16.14	21.49	49.68
60	**71.74**	60.66	61.60	51.79	17.30	21.49	55.73
70	**74.33**	59.34	62.15	56.80	15.48	21.49	60.57
80	**76.53**	60.61	63.80	58.07	15.26	21.49	61.48
90	**77.85**	60.72	63.47	55.87	15.37	21.49	63.95
100	**80.28**	58.35	64.85	57.41	15.59	13.22	65.25
Ave.	**68.23**	58.06	60.82	52.40	16.00	20.57	56.02

TABLE VI
FACIAL EXPRESSION RECOGNITION RATE (%) USING THE OCCLUDED CK + IIMAGES (CASE OF OCCLUSION ON TESTING SET)

No. Base	DPCMF	PCMF	EE-CMF	DPNMF	PNMF	DNMF	NMF
20	66.32	65.95	**73.26**	56.24	34.38	24.55	50.62
30	72.23	**73.31**	68.77	59.26	35.87	25.41	58.39
40	**77.81**	77.27	68.48	66.12	43.93	26.98	62.27
50	**80.62**	79.71	68.71	71.86	44.05	35.45	65.29
60	**83.80**	82.69	72.2	75.25	45.91	40.33	70.37
70	**84.83**	83.18	71.31	75.33	46.61	36.36	70.33
80	**86.36**	84.67	73.91	77.56	49.88	40.00	73.31
90	**85.54**	85.17	73.38	79.79	52.64	40.95	73.39
100	**86.24**	85.12	73.2	79.71	51.36	37.98	75.25
Ave.	**80.42**	79.67	71.47	71.24	44.96	34.22	66.58

5. CONCLUSION

This paper has presented two efficient algorithms for robust image representation. The novel approaches take advantages on complex matrix factorization to learn subspace. The combination with Fisher's criteria provides a supervised model which is reliable and stable to extract the meaningful features and make the classification task much easier. The proposed models uncover the low-dimensional structures hidden in the high-dimensional data and gets rid of the data redundancy, and thus significantly enhance the recognition performance. The future works are focused on extending the proposed approaches to the nonlinear representation and also testing their performance on various type of dataset, particularly on complex data such as Fourier feature of acoustics data.

6. REFERENCES

[1] 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.

[2] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Advances in Neural Information Processing Systems*, pp. 556-562, 2001.

[3] V. H. Duong, Y. S. Lee, J. J. Ding, B. T. Pham, P. T. Bao, and J. C. Wang, “Exemplar-embed complex matrix factorization for facial expression recognition,” in *Proc. of the 42nd International Conference on Acoustics, Speech and Signal Processing* (*ICASSP’07)*, 2017, p. 1837-1841.

[4] D. Cai, X. He, X. Wu, and J. Han, “Nonnegative matrix factorization on manifold,” in *Proc. IEEE Int’l Data Mining (ICDM ’08)*, pp. 63-72, 2008.

[5] D. Cai, X. F. He, J. W. Han, and T. S. Huang, “Graph regularized nonnegative matrix factorization for data representation,” *IEEE Trans. Patt. Anal. and Mach. Inte.*, vol. 33, no. 8, pp. 1548-1560, 2011.

[6] S. Zafeiriou, A. Tefas, I. Buciu, I. Pitas, “Exploiting discriminant Information in nonnegative matrix factorization with application to frontal face verification,” *IEEE Transactions on Neural Networks*, vol. 17, pp. 683-695, 2006.

[7] K. Fukunaga, “Statistical Pattern Recognition,”*Acadamic*, 1990.

[8] A. J. Fitch, A. Kadyrov, W. J. Christmas, and J. Kittler, “Fast robust correlation,” *IEEE Transactions on Image Processing*, vol. 14, no.8, pp. 1063-1073, 2005.

[9] S. Liwicki, G. Tzimiropoulos, S Zafeiriou, and M. Pantic, “Euler principal component analysis,” *Int. J. Comput. Vis*., vol. 101, pp. 498-518, 2013.

[10] V. H. Duong, Y. S. Lee, B. T. Pham, S. Mathulaprangsan, P. T. Bao, and J. C. Wang, “Complex matrix factorization for face recognition,” 2016, [Online] Available:
https://arxiv.org/ftp/arxiv/papers/1612/1612.02513.pdf

[11] V. H. Duong, M. Q. Bui, J. J. Ding, Y. S. Lee, B. T. Pham, P. T. Bao, and J. C. Wang, “A new approach of matrix factorization on complex domain for data representation,” *IEICE Trans. on Information and Systems,* vol. 100, no. 12, p. 3059-3063, 2017.

[12] Moskowitz and Martin, *A course in complex analysis in one variable*, World Scientific Publishing Co., pp. 7, 2002.

[13] C. J. Lin, “Projected gradient methods for non-negative matrix factorization,” *Neural Comput*., vol. 19, pp. 2756-2779, 2007.

[14] M. F. Amin, M. I. Amin, A. Al-Nuaimi, and K. Murase, “Wirtinger calculus based gradient descent and Levenberg Marquardt learning algorithms in complex-valued neural networks,” in *Proc. Int. Conf. Neural Information Processing ((ICONIP)*, 2011, pp. 550-559.

[15] A. Hjorunges and D. Gesbert, “Complex-valued matrix differentiation: techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740-2746, 2007.

[16] Michiel Hazewinkel, *Taylor series*, Encyclopedia of Mathematics, Springer Science+Business Media B.V. / Kluwer Academic Publishers, 2001.

[17] J. Nocedal, S. J. Wright, *Numerical Optimization*, Springer-Verlag, 2000.

[18] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Mathews, “The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression,” *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010, pp. 94-101.

[19] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with Gabor wavelets,” in *IEEE Int. Conf. Auto. Face, Gest. Recog.*, 1998, pp. 200-205

[20] Z. Yuan and E. Oja, “Projective nonnegative matrix factorization for image compression and feature extraction,” in *Scandinavian Conf. Image Analysis*, 2005, pp. 333–342.

[21] N. Guan, X. Zhang, Z. Luo,D. Tao, X. Yang, “Discriminant projective non-negative matrix factorization,” *PLoS ONE*, vol. 8, no.12, e83291, 2013.