3.2 Non-negative Low-Rank Learning

Zhirong Yang, He Zhang, Zhanxing Zhu, and Erkki Oja

Enforcing nonnegativity in linear factorizations [4] has proven to be a powerful principle for multivariate data analysis, especially sparse feature analysis, as shown by the well-known Nonnegative Matrix Factorization (NMF) algorithm by Lee and Seung [2]. Their method minimizes the difference between the data matrix $\mathbf{X}$ and its non-negative decomposition $\mathbf{WH}$. Yuan and Oja [11] proposed the Projective NMF (PNMF) method which replaces $\mathbf{H}$ in NMF with $\mathbf{W}^T\mathbf{X}$. Empirical results indicate that PNMF is able to produce more spatially localized, part-based representations of visual patterns.

Recently, we have extended and completed the preliminary work with the following new contributions [5]: (1) formal convergence analysis of the original PNMF algorithms, (2) PNMF with the orthonormality constraint, (3) nonlinear extension of PNMF, (4) comparison of PNMF with two classical and two recent algorithms [10, 1] for clustering, (5) a new application of PNMF for recovering the projection matrix in a nonnegative mixture model, (6) comparison of PNMF with the approach of discretizing eigenvectors, and (7) theoretical justification of moving a term in the generic multiplicative update rule. Our in-depth analysis shows that the PNMF replacement has positive consequences in sparseness of the approximation, orthogonality of the factorizing matrix, decreased computational complexity in learning, close equivalence to clustering, generalization of the approximation to new data without heavy re-computations, and easy extension to a nonlinear kernel method with wide applications for optimization problems. We have later demonstrated that combining orthogonality and negativity works well in graph partitioning [3].

In NMF, the matrix difference was originally measured by the Frobenius matrix norm or the unnormalized Kullback-Leibler divergence (I-divergence). Recently we have significantly extended NMF to a much larger variety of divergences with theoretically convergent algorithms. In [7], we have presented a generic principle for deriving multiplicative update rules, as well as a proof of the convergence of their objective function, that applies for a large variety of linear and quadratic NMF problems. The proposed principle only requires that the NMF approximation objective function can be written as a sum of a finite number of monomials, which is a mild assumption that holds for many commonly used approximation error measures. As a result, our method turns the derivation, which seemingly requires intense mathematical work, into a routine exercise that could be even readily automated using symbolic mathematics software. In our practice [8], both theoretical and practical advantages indicate that there would be good reasons to replace the I-divergence with normalized Kullback-Leibler for NMF and its variants. The PNMF method can also be generalized to the α-divergence family [6].

Automatic determination of the low-rank in NMF is a difficult problem. In [9], we have presented a new algorithm which can automatically determine the rank of the projection matrix in PNMF. By using Jeffrey's prior as the model prior, we have made our algorithm free of human tuning in finding algorithm parameters. Figure 3.1 visualizes the learned basis of the *Swimmer* dataset.

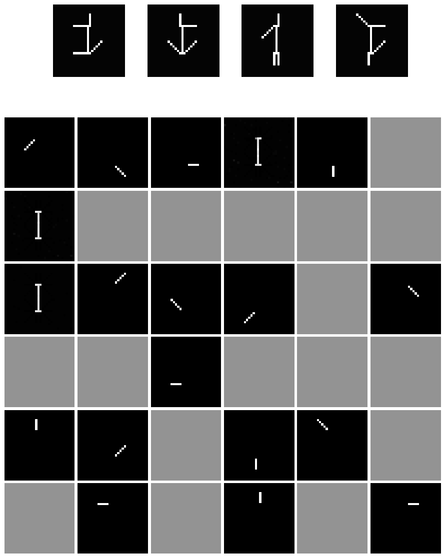

Figure 3.1: (Top) Some sample images of Swimmer dataset; (Bottom) 36 basis images of Swimmer dataset. The gray cells correspond to matrix columns whose L_2-norms are zero or very close to zero.

References

[1] I. Dhillon, Y. Guan, and B. Kulis. Kernel kmeans, spectral clustering and normalized cuts. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 551–556, Seattle, WA, USA, 2004.

[2] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[3] E. Oja and Z. Yang. Orthogonal nonnegative learning for sparse feature extraction and approximate combinatorial optimization. *Frontiers of Electrical and Electronic Engineering in China*, 5(3):261–273, 2010.

[4] P. Paatero P and U. Tapper. Positive matrix factorization: A nonnegative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.

[5] Z. Yang and E. Oja. Linear and nonlinear projective nonnegative matrix factorization. *IEEE Transaction on Neural Networks*, 21(5):734–749, 2010.

[6] Z. Yang and E. Oja. Projective nonnegative matrix factorization based on α-divergence. *Journal of Artificial Intelligence and Soft Computing Research*, 1(1):7–16, 2011.

[7] Z. Yang and E. Oja. Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization. *IEEE Transactions on Neural Networks*, 22(12):1878–1891, 2011.

[8] Z. Yang, H. Zhang, Z. Yuan, and E. Oja. Kullback-leibler divergence for nonnegative matrix factorization. In *Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN2011)*, pages 250–257, Espoo, Finland, 2011.

[9] Z. Yang, Z. Zhu, and E. Oja. Automatic rank determination in projective nonnegative matrix factorization. In *Proccedings of the 9th International Conference on Latent Variable Analysis and Signal Separation (LVA 2010)*, pages 514–521, St. Malo, France, 2010.

[10] S. X. Yu and J. Shi. Multiclass spectral clustering. In *Proceedings of the 9th IEEE International Conference on Computer Vision*, volume 2, pages 313–319, 2003.

[11] Z. Yuan and E. Oja. Projective nonnegative matrix factorization for image compression and feature extraction. In *Proceedings of 14th Scandinavian Conference on Image Analysis (SCIA)*, pages 333–342, Joensuu, Finland, 2005.