THE OPEN JOURNAL OF MATHEMATICAL SCIENCES AND APPLICATIONS
ISSN 0000-0000
(https://openjournal.utar.edu.my/index.php/cmsojmsa)

Semi-Orthogonal Non-negative Matrix Factorization with Sparse Constraint in Topic Modelling

Gillian Yi Han Woo [*1], Hong Seng Sim[1], Yong Kheng Goh[1], and Wah June Leong[2]

[1] *Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Malaysia*
[2] *Institute for Mathematical Research, Universiti Putra Malaysia, Malaysia*

Received: *, accepted: *, published online: *.

Abstract

This study introduces a novel method for topic modeling by combining semi-orthogonal matrix factorization with sparse constraints to improve interpretability, coherence, and scalability. Traditional techniques such as Latent Dirichlet Allocation (LDA) and Nonnegative Matrix Factorization (NMF) often face challenges in maintaining these qualities, especially with high-dimensional data. To address these issues, we propose the Spectral Proximal Method (SPM), an optimization approach that uses proximal variable metric updates with spectral diagonal scaling. SPM enforces both l_1-norm sparsity and semi-orthogonality to generate diverse and interpretable topics. The algorithm uses non-convex alternating minimization, with initialisation based on NMF to enhance computational efficiency.

Keywords: proximal variable metric; matrix factorisation; orthogonal matrix; sparse optimisation; topic modelling.

Math. Subj. Class. (2020):

1 Introduction

In modern information processing, the key technique involves factorising complex data matrices into simpler forms with lower ranks. Our research objective is to decompose large

* Corresponding author: gillianwoo@1utar.my

data matrix V, by finding two smaller matrices, W and H based on optimisation-based approximations:

$$\min_{W,H} \frac{\alpha}{2}\|V - WH^T\|_F^2 + \frac{\beta}{2}\|W^T W - I\|_F^2 + \gamma\|H\|_0 \tag{1.1}$$
$$\text{subject to } W \geq 0, \quad H \geq 0.$$

where $\alpha > 0$, $\beta > 0$, and $gamma > 0$. The objective problem has semi-orthogonal structures and sparse constraints, which expect to increase the efficiency in machine learning problems. However, this is a challenging task due to its non-convex nature, and easily ensnare in local minima (Chi et al., 2019). Hence, we apply the convex relaxation (Recht et al., 2010) and employ non-convex alternating minimization (Polson et al., 2015).

2 Literature Review

2.1 Nonnegative Matrix Factorization (NMF)

NMF (Lee and Seung, 2024) decomposes a nonnegative matrix into two lower-rank nonnegative factors, offering interpretable and parts-based representations. Orthogonal NMF (ONMF) (Ding et al., 2008) extends NMF by imposing orthogonality constraints on one factor matrix, reducing redundancy and yielding more distinct latent features. However, it does not incorporate sparsity.

2.2 Spectral Proximal Method (SPM)

In this research, we employ the proximal variable metric method with spectral diagonal updates, as proposed by (Woo et al., 2023), to solve the optimization problem (1.1), where the sparsity constraint relax to $\|H\|_1$ and the semi-orthogonal constraints imposed on matrix W. The method, derived from proximal gradient principles (Parikh and Boyd, 2014), achieves efficient updates with $O(n)$ storage.

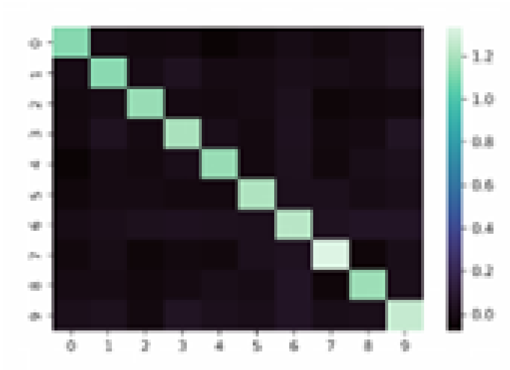
3 Results

In the experiment, we will compare SPM method with the Proximal Steepest Descent (PSD) on 20Newsgroups dataset. NMF model with euclidean distance initialises W and H for 50 iterations. Then, continue with 200 iterations for optimisation.

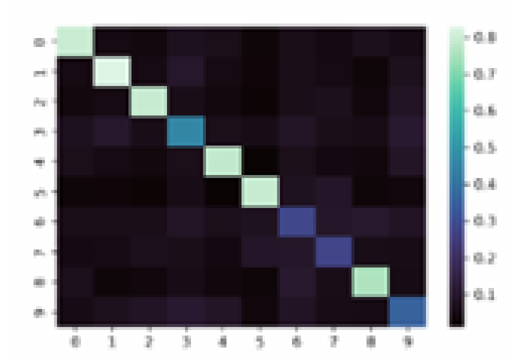
Table 1: Results in partitioning the dataset to 10 topics.

	Number of Function Calls	**Execution Time (seconds)**	**Sparsity of H (%)**	**Cosine Similarity of H**
PSD	19992562	143.0812	18.03	0.6752
SPM	355167	101.2079	39.60	0.3712

SPM exhibits superior efficiency regarding the number of function calls and execution time when compared to PSD. Both methods yield a matrix W that incorporates an orthogonal structure. Therefore, considering the trade-off between topic coherence, efficiency and the sparsity of the resulting matrix H, the SPM method emerges as the more effective approach for solving the topic modeling problem.



(a) PSD



(b) SPM

Figure 1: Orthogonality of W when in partitioning the dataset to 10 topics.

4 Conclusions

SPM effectively generates a document-topic matrix. W with semi-orthogonal properties and sparse topic-term matrix, H.

Acknowledgements

The research was supported by Universiti Tunku Abdul Rahman (UTAR) through Universiti Tunku Abdul Rahman Research Fund project number IPSR/RMC/UTARRF/2023-C2/G01.

References

Chi, Y., Lu, Y. M. and Chen, Y. (2019) 'Nonconvex optimization meets low-rank matrix factorization: An overview', *IEEE Transactions on Signal Processing* **20**(12), 5239–5269.

Ding, C. H., Li, T. and Jordan, M. I. (2008) 'Convex and semi-nonnegative matrix factorizations', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**(1), 45–55.

Lee, D. D. and Seung, H. S. (2024) 'Learning the parts of objects by non-negative matrix factorization', *Nature* **401**, 788–791.

Parikh, N. and Boyd, S. (2014) 'Proximal algorithms', *Foundations and Trends in Optimization* **1**(3), 127–239.

Polson, N. G., Scott, J. G. and Willard, B. T. (2015) 'Proximal algorithms in statistics and machine learning', *Statistical Science* **30**(4), 559–581.

Recht, B., Fazel, M. and Parrilo, P. A. (2010) 'Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization', *SIAM Review* **52**(3), 471–501.

Woo, G. Y. H., Sim, H. S., Goh, Y. K. and Leong, W. J. (2023) 'Proximal variable metric method with spectral diagonal update for large-scale sparse optimization', *Journal of the Franklin Institute* **360**(7), 4640–4660.