

Leveraging Geometry for Fast Mixing Bayesian Non-negative Matrix Factorization

Muhammad Arjumand Masood

Harvard University

Finale Doshi-Velez

Harvard University

MASOOD@G.HARVARD.EDU

FINALE@SEAS.HARVARD.EDU

Bayesian Non-negative Matrix Factorization (NMF) is a popular tool for data exploration. Its Bayesian nature has the additional advantage of improving interpretability over other NMF algorithms by providing uncertainty estimates in factorizations as well as finding the multiple NMF modes that may exist. Practical MCMC simulations incur slow mixing times and tend to get stuck on single models. In this work we leverage the underlying geometry of the NMF problem to reformulate the non-negativity constraint and parameterize the space of non-negative matrices within a given subspace. We employ this technique to propose global changes in our MCMC chain via a Metropolis-Hastings step within a Bayesian NMF Gibbs sampler to encourage the MCMC chain to explore the true posterior at a much faster rate.

Starting with an NMF $X = WA$ with $X, W, A \geq 0$, we employ the theory behind the vertex enumeration problem for polyhedra to re-characterize the non-negativity constraint of the NMF problem. We express the space of \vec{q} satisfying the non-negativity constraint $W\vec{q} \geq 0$ as $\vec{q} = \{R_W y_W | y_W \geq 0\}$. The matrix R_W is the matrix of extreme rays such that any conical sum of its columns will yield a vector that satisfies the non-negativity constraint. This allows us to propose global changes to W in the same subspace while maintaining the non-negativity condition $W' = WQ$. We ensure that each column of Q is a conic combination of our extreme rays i.e. $Q = R_W Y_W$. There exists then a one-to-one mapping between the matrix of conic combinations Y_W and non-negative matrices in the subspace spanned by W . We use this insight to make global MH proposals within the Gibbs Sampler.

To evaluate our approach, an angle-based similarity measure is developed for comparing factorizations and is used to demonstrate the faster mixing times using our methodology. In addition to effective sample size and autocorrelation measurements, we use this custom similarity measure to assess mixing in the MCMC chain. Larger angle deviations across samples in the chain is indicative of better

mixing.

The algorithm’s (ERMH) run on synthetic data with two known exact factorizations successfully finds both factorizations and shows faster mixing compared to baselines as measured by the effective sample size. Performing the algorithm on real datasets shows that we are able to obtain distinct NMF modes that may allow for different interpretations by domain experts.

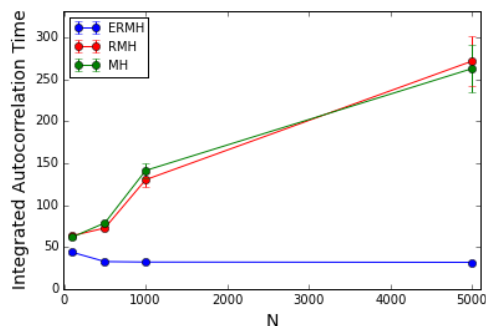


Figure 1. ERMH consistently outperforms the baselines on integrated autocorrelation time on this synthetic data set with known two exact factorizations.

Table 1. Maximum Angle Standard Deviation(in degrees). We note that our algorithm ERMH shows greater angle deviations on these real data sets compared to our baselines

	Dataset		
	AML	BBC	Faces
Gibbs	0.6 (0.2, 1.0)	0.49 (0.48, 0.51)	1.09 (0.47, 1.71)
RMH	0.7 (0.2, 1.2)	0.49 (0.47, 0.51)	1.37 (0.66, 2.07)
ERMH	2.4 (2.1, 2.8)	3.28 (3.17, 3.40)	6.59 (5.78, 7.40)

Acknowledgements

We would like to acknowledge Weiwei Pan for helpful discussions and comments

References

- Arora, Sanjeev, Ge, Rong, Halpern, Yoni, Mimno, David, Moitra, Ankur, Sontag, David, Wu, Yichen, and Zhu, Michael. A practical algorithm for topic modeling with provable guarantees. *arXiv preprint arXiv:1212.4777*, 2012.
- Brecheisen, Stefan, Kriegel, Hans-Peter, and Pfeifle, Martin. Efficient similarity search on vector sets. In *BTW*, pp. 425–443, 2005.
- Brunet, Jean-Philippe, Tamayo, Pablo, Golub, Todd R, and Mesirov, Jill P. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the national academy of sciences*, 101(12):4164–4169, 2004.
- Cichocki, Andrzej, Zdunek, Rafal, and Amari, Shun-ichi. New algorithms for non-negative matrix factorization in applications to blind source separation. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 5, pp. V–V. IEEE, 2006.
- Donoho, David and Stodden, Victoria. When does non-negative matrix factorization give correct decomposition into parts? In Thrun, S., Saul, L.K., and Schölkopf, B. (eds.), *Advances in Neural Information Processing Systems 16*, pp. 1141–1148. MIT Press, 2004.
- Févotte, Cédric, Bertin, Nancy, and Durrieu, Jean-Louis. Non-negative matrix factorization with the itakura-saito divergence: With application to music analysis. *Neural computation*, 21(3):793–830, 2009.
- Fukuda, Komei and Prodon, Alain. Double description method revisited. In *Combinatorics and computer science*, pp. 91–111. Springer, 1996.
- Gillis, Nicolas. Sparse and unique nonnegative matrix factorization through data preprocessing. *J. Mach. Learn. Res.*, 23(1): 3349–3386, January 2012.
- Greene, Derek and Cunningham, Pádraig. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proceedings of the 23rd international conference on Machine learning*, pp. 377–384. ACM, 2006.
- Laurberg, H., Christensen, M., Plumbley, M., Hansen, L., and Jensen, S. Theorems on positive data: on the uniqueness of nmf. *Computational Intelligence and Neuroscience*, 2008.
- Lee, D.D. and Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
- Lin, Chuan-bi. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10):2756–2779, 2007.
- Motzkin, Theodore S, Raiffa, Howard, Thompson, Gerald L, and Thrall, Robert M. The double description method. 1953.
- Moussaoui, Saïd, Brie, David, Mohammad-Djafari, Ali, and Carteret, Cédric. Separation of non-negative mixture of non-negative sources using a bayesian approach and mcmc sampling. *Signal Processing, IEEE Transactions on*, 54(11):4133–4145, 2006.
- Paisley, John, Blei, D, and Jordan, Michael I. Bayesian nonnegative matrix factorization with stochastic variational inference.
- Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent, et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.
- Schmidt, Mikkel N, Winther, Ole, and Hansen, Lars Kai. Bayesian non-negative matrix factorization. In *Independent Component Analysis and Signal Separation*, pp. 540–547. Springer, 2009.
- Yang, Zhirong, Yuan, Zhijian, and Laaksonen, Jorma. Projective non-negative matrix factorization with applications to facial image processing. *International Journal of Pattern Recognition and Artificial Intelligence*, 21(8):1353–1362, December 2007.
- Žitnik, Marinka and Zupan, Blaž. Nimfa: A python library for nonnegative matrix factorization. *The Journal of Machine Learning Research*, 13(1):849–853, 2012.