21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Semidefinite Programming Based Preconditioning for More Robust Near-Separable Nonnegative Matrix Factorization

Nicolas Gillis[1] and Stephen A. Vavasis[2]

***Abstract*— Nonnegative matrix factorization (NMF) under the separability assumption can provably be solved efficiently, even in the presence of noise, and has been shown to be a powerful technique in document classification and hyperspectral unmixing. This problem is referred to as near-separable NMF and requires that there exists a cone spanned by a small subset of the columns of the input nonnegative matrix approximately containing all columns. In this talk, we present a preconditioning based on semidefinite programming introduced in 'Semidefinite programming based preconditioning for more robust near-separable nonnegative matrix factorization' (Gillis and Vavasis, arXiv:1310.2273, 2013), making the input matrix well-conditioned. This in turn can improve significantly the performance of near-separable NMF algorithms which is illustrated on the popular successive projection algorithm (SPA). The new preconditioned SPA is provably more robust to noise, and outperforms SPA on several synthetic data sets.**

I. INTRODUCTION

Nonnegative matrix factorization (NMF) is a powerful dimensionality reduction technique as it automatically extracts sparse and meaningful features from a set of nonnegative data vectors. Given n nonnegative m-dimensional vectors gathered in a nonnegative matrix $M \in \mathbb{R}^{m \times n}_+$ and a factorization rank r, NMF computes two nonnegative matrices $W \in \mathbb{R}^{m \times r}_+$ and $H \in \mathbb{R}^{r \times n}_+$ such that $M \approx WH$. Unfortunately, NMF is NP-hard in general [1]. However, if the input data matrix M is r-separable, that is, if there exists an index set $\mathcal{K}$ with cardinality r and an r-by-n nonnegative matrix H such that $M = M(:, \mathcal{K})H$, then the problem can be solved in polynomial time [2]. In the presence of noise, this problem is referred to as near-separable NMF. Although the separability condition is rather strong, it makes sense in several applications, e.g., in text mining [3], hyperspectral unmixing [4] and blind source separation [5].

In this talk, we present the results from [6]: we show how to make near-separable NMF algorithms more robust using preconditioning. We focus on a simple yet effective near-separable NMF algorithm, namely, the successive projection algorithm.

*This work was supported in part by a grant from the U.S. Air Force Office of Scientific Research and a Discovery Grant from the Natural Science and Engineering Research Council (Canada).

[1]N. Gillis is with the Department of Mathematics and Operational Research, Faculté Polytechnique, Université de Mons, Rue de Houdain 9, 7000 Mons, Belgium. nicolas.gillis@umons.ac.be

[2]S. Vavasis is with the Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. vavasis@math.uwaterloo.ca

II. SUCCESSIVE PROJECTION ALGORITHM

The successive projection algorithm is a simple but fast and robust recursive algorithm for solving near-separable NMF; see Algorithm SPA. At each step of the algorithm, the column of the input near-separable matrix $\tilde{M}$ with maximum ℓ_2 norm is selected, and then $\tilde{M}$ is updated by projecting each column onto the orthogonal complement of the columns selected so far. It was first introduced in [7], and later proved to be robust in [8].

Assumption 1: The matrix $\tilde{M}$ is a near-separable matrix if $\tilde{M} = W[I_r \ H']\Pi + N$ with W full rank, $H'(:, i) \in \{x \in \mathbb{R}^r \mid x \geq 0, \sum_i x_i \leq 1\}$ for all i, Π a permutation, and $\max_i ||N_{:i}||_2 \leq \epsilon$ for some $\epsilon > 0$.

Theorem 1 ([8], Th. 3): Let $\tilde{M}$ satisfy Ass. 1. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(W)}{\sqrt{r}\kappa^2(W)}\right)$, then SPA identifies the columns of W up to error $\mathcal{O}\left(\epsilon\, \kappa^2(W)\right)$, that is,

Algorithm SPA Successive Projection Algorithm [7], [8]

Input: Near-separable matrix $\tilde{M}$, rank r.
Output: Set of r indices $\mathcal{K}$ such that $\tilde{M}(:,\mathcal{K}) \approx W$.

```
Let R = M̃, K = {}.
for k = 1 : r do
    p = argmax_j ||R_{:j}||_2.
    R = (I − R_{:p} R_{:p}^T / ||R_{:p}||_2^2) R.
    K = K ∪ {p}.
end for
```

Let $R = \tilde{M}$, $\mathcal{K} = \{\}$.
for $k = 1 : r$ **do**
  $p = \text{argmax}_j \, ||R_{:j}||_2$.
  $R = \left(I - \frac{R_{:p} R_{:p}^T}{||R_{:p}||_2^2} \right) R$.
  $\mathcal{K} = \mathcal{K} \cup \{p\}$.
end for

the index set $\mathcal{K}$ identified by SPA satisfies

$$\max_{1 \leq j \leq r} \min_{k \in \mathcal{K}} \left\| W(:,j) - \tilde{M}(:,k) \right\|_2 \leq \mathcal{O}\left(\epsilon \, \kappa^2(W)\right),$$

where $\kappa(W) = \frac{\sigma_{\max}(W)}{\sigma_{\min}(W)}$ is the condition number of W.

III. PRECONDITIONING

Assume for simplicity that $m = r$. If we were given the full rank r-by-r matrix W, we could premultiply the input near-separable matrix $\tilde{M} = W[I_r, H'] + N$ with W^{-1} and obtain

$$\tilde{M}' = W^{-1}\tilde{M} = I_r[I_r, H'] + W^{-1}N.$$

The matrix $\tilde{M}'$ is also near-separable, and is now perfectly conditioned, although the noise $\epsilon' = \max_i ||W^{-1}N_{:i}||_2$ might have increased but by a factor of at most $\sigma_{\min}^{-1}(W)$. Under the same conditions as in Theorem 1, one can check that $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(W)}{\sqrt{r}}\right)$ would imply that SPA identifies the columns of W up to error $\mathcal{O}\left(\epsilon \, \kappa(W)\right)$. This is a significant improvement compared to Theorem 1, both for the noise level ϵ (improvement of a factor $\kappa^{-2}(W)$) and the error (improvement of a factor $\kappa(W)$).

Of course, the matrix W is unknown, otherwise the problem would be solved. However, it can be shown that a minimum volume ellipsoid problem allows to approximate W^{-1} (up to orthogonal transformations). Given a matrix $\tilde{M} \in \mathbb{R}^{r \times n}$ of rank r, we can formulate the minimum volume ellipsoid centered at the origin and containing the columns $\tilde{m}_i$ $1 \leq i \leq n$ of matrix $\tilde{M}$ as follows

$$A^* = \text{argmin}_{A \in \mathbb{S}_+^r} \log\det(A)^{-1} \quad \text{s.t. } \tilde{m}_i^T A \tilde{m}_i \leq 1 \quad \forall i. \qquad (1)$$

This problem is SDP representable [9, p.222]. In [6], it is shown that, if $\tilde{M}$ satisfies Ass. 1, then

$$A^* \approx \left(WW^T\right)^{-1},$$

hence factoring A^* (e.g., using the Cholesky decomposition) allows to recover W^{-1} approximately (up to orthogonal transformations). In fact, in the noiseless case (that is, for $N = 0$), $A^* = (WW^T)^{-1}$.

Finally, the new proposed algorithm, referred to as preconditioned SPA, works as follows:

1) Solve (1) in order to approximately compute $P \approx W^{-1}$.
2) Premultiply $\tilde{M}$ with the preconditioning to obtain $\tilde{M}' = P\tilde{M}$.
3) Apply SPA on the pre-conditioned matrix $\tilde{M}'$.

We have the following robustness result:

Theorem 2 ([6], Th. 3): Let $\tilde{M}$ satisfy Ass. 1. If $\epsilon \leq \mathcal{O}\left(\frac{\sigma_{\min}(W)}{r\sqrt{r}}\right)$, then preconditioned SPA identifies a subset $\mathcal{K}$ so that $\tilde{M}(:,\mathcal{K})$ approximates the columns of W up to error $\mathcal{O}\left(\epsilon \, \kappa(W)\right)$.

IV. NUMERICAL EXPERIMENTS

In this section, we illustrate the robustness of preconditioned SPA (prec-SPA) compared to the original SPA on some synthetic data sets. We also compare it to vertex component analysis (VCA) [10], a popular endmember extraction algorithm, and XRAY from [11]. The Matlab code is available at `https://sites.google.com/site/nicolasgillis/`. Please see [6] for more details and other numerical experiments.

The near-separable matrix $\tilde{M}$ is generated as follows: We take $m = r = 20$, and $n = 210$. The matrix W is generated using the `rand(.)` function of Matlab, that is, each entry is drawn uniformly at random in the interval $[0, 1]$. The matrix $H = [I_r, H']$ is such that H' has exactly two non-zero entries in each row equal to 0.5. Hence, the 190 data points are in the middle of two different columns of W. The noise is chosen such that the columns of W (that is, the first 20 columns of M) are not perturbed, while the 190 data points are moved towards the outside of the convex hull of the columns of W: $N(:, j) = 0$ for $1 \leq j \leq 20$ and

$$N(:, j) = \delta \left(M(:, j) - \bar{w} \right) \quad 21 \leq j \leq 210,$$

where $M = WH$ and $\bar{w} = \frac{1}{r} \sum_i w_i$ is the vertex centroid of the convex hull of the columns of W. These are the same near-separable matrices as in [8]. For different noise levels δ, we generate ten such matrices and Figure 1 reports the percentage of columns of W correctly identified by the different algorithms. Clearly, prec-SPA outperforms SPA, VCA and XRAY.

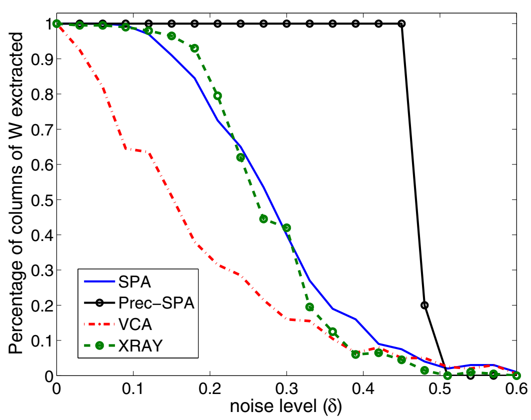

Fig. 1. Comparison of the different near-separable NMF algorithms on synthetic data sets.

V. CONCLUSIONS

We have presented the preconditioning for near-separable NMF matrices using semidefinite programming from [6]. This allows to robustify near-separable NMF algorithms. In particular, the preconditioning makes the popular successive projection algorithm (SPA) provably more robust to noise, which was illustrated on some synthetic data sets.

REFERENCES

[1] S. Vavasis, "On the complexity of nonnegative matrix factorization," *SIAM J. on Optimization*, vol. 20, no. 3, pp. 1364–1377, 2009.
[2] S. Arora, R. Ge, R. Kannan, and A. Moitra, "Computing a nonnegative matrix factorization – provably," in *STOC '12*, 2012, pp. 145–162.
[3] S. Arora, R. Ge, Y. Halpern, D. Mimno, A. Moitra, D. Sontag, Y. Wu, and M. Zhu, "A practical algorithm for topic modeling with provable guarantees," in *(ICML '13*, 2013, vol. 28, no. 2, pp. 280–288.
[4] W.-K. Ma, J. Bioucas-Dias, P. Gader, T.-H. Chan, N. Gillis, A. Plaza, A. Ambikapathi, and C.-Y. Chi, "A Signal processing perspective on hyperspectral unmixing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 67–81, 2014.
[5] T.-H. Chan, W.-K. Ma, C.-Y. Chi, and Y. Wang, "A convex analysis framework for blind separation of non-negative sources," *IEEE Trans. on Signal Processing*, vol. 56, no. 10, pp. 5120–5134, 2008.
[6] N. Gillis and S. Vavasis, "Semidefinite programming based preconditioning for more robust near-separable nonnegative matrix factorization," 2013, arXiv:1310.2273.
[7] U. Araújo, B. Saldanha, R. Galvão, T. Yoneyama, H. Chame, and V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemom. and Intell. Lab. Syst.*, vol. 57, no. 2, pp. 65–73, 2001.
[8] N. Gillis and S. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 698–714, 2014.
[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, Cambridge, 2004.
[10] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.
[11] A. Kumar, V. Sindhwani, and P. Kambadur, "Fast conical hull algorithms for near-separable non-negative matrix factorization," in *ICML '13*, 2013, vol. 28, no. 1, pp. 231–239.