Separable Simplex-Structured Matrix Factorization: Robustness of Combinatorial Approaches

Nicolas Gillis

University of Mons, Belgium

1. The Problem

The separable simplex-structured matrix factorization (separable SSMF) problem is defined as follows: Given an input matrix $X \in \mathbb{R}^{m \times n}$ and a factorization rank r, find an index set $\mathcal{K}$ of size r and a matrix $H \in \mathbb{R}^{m \times r}$ such that

$$X \approx X(:,\mathcal{K})H \quad \text{where } H(:,j) \in \Delta^r \text{ for all } j,$$

with

$$\Delta^r = \{h \in \mathbb{R}^r \mid h \geq 0, \sum_j h_j \leq 1\}.$$

Separable SSMF is a generalization of the separable nonnegative matrix factorization (NMF) problem which requires $X \geq 0$. There are many applications such as hyperspectral unmixing and document analysis [1].

2. Algorithms

Three classes for separable SSMF algorithms:

1. **Greedy algorithms** identify sequentially the columns $X(:,\mathcal{K})$ using a two-step strategy: (1) selection step, and (2) projection step. They have low computational cost and memory requirement but are less robust to noise.
Examples: vertex component analysis (VCA) [2], successive projection algorithm (SPA) [3], AnchorWords [4], successive nonnegative projection algorithm (SNPA) [6], Preconditioned SPA (Prec-SPA) [20].
2. **Convex relaxations** are usually based on the following reformulation: Find Y with r non-zero rows such that $X \approx XY$. Promoting row sparsity can be achieved by ℓ_1 minimization. These formulations are significantly more robust than greedy algorithms but computationally demanding.
Examples: $\ell_{1,q}$ relaxations [8,9,10], Hottopixx [11], LP-based relaxations [12,13,14].
3. **Combinatorial approaches**. Given an index set $\mathcal{K}$, one can compute the error on fitting X, namely, $g(\mathcal{K}) = \min_{H(:,j) \in \Delta^r \, \forall j} ||X - X(:,\mathcal{K})H||$ for some norm $||.||$. Optimizing g over $\mathcal{K}$ is a difficult combinatorial problem with $\binom{n}{r}$ possible solutions.
Examples: N-FINDR [15], ant-colony optimization [16], bee-colony and genetic algorithms [17], alternating optimization [18], approximation algorithms [19].

3. Assumptions

The noiseless separable matrix $X \in \mathbb{R}^{m \times n}$ is given by

$$X = X(:,\mathcal{K}^*)H = WH,$$

where (i) $|\mathcal{K}^*| = r$, (ii) $H(:,j) \in \Delta^r \ \forall\, 1 \leq j \leq n$, and (iii) the matrix $W = X(:,\mathcal{K}^*) \in \mathbb{R}^{m \times r}$ satisfies $\kappa(W) > 0$, where

$$\kappa(W) = \min_{1 \leq k \leq r} \min_{h \in \Delta^{r-1}} ||W(:,k) - W(:,[r] \backslash \{k\})h||_2,$$

with $[r] = \{1, 2, \dots, r\}$. Given X, the input noisy matrix $\tilde{X}$ is given by $\tilde{X} = X + N$ with $||N(:,j)||_2 \leq \epsilon$ for all j.

4. Provably robust algorithms

Given an algorithm computing the solution $\mathcal{K}$ whose goal is to have $\tilde{X}(:,\mathcal{K}) \approx X(:,\mathcal{K}^*) = W$, provide some δ such that

$$q(\mathcal{K}) = \max_{1 \leq k \leq r} \min_{j \in \mathcal{K}} ||W(:,k) - \tilde{X}(:,j)||_2 \ \leq \ \delta,$$

given that ϵ is sufficiently small (that is, $\epsilon \leq \gamma$ for some γ).

5. Goal of this paper

For greedy algorithms and LP-based relaxations, robustness results already exist (see Table 1). However, there was a gap in the literature: no such result exists for combinatorial approaches. The most natural formulation that tries to recover W from $\tilde{X}$ is the following:

$$\min_{\mathcal{K}} f(\mathcal{K}), \quad \text{with } f(\mathcal{K}) = \max_{1 \leq j \leq n} \min_{z \in \Delta} ||\tilde{x}_j - \tilde{X}(:,\mathcal{K})z||_2. \quad (1)$$

This formulation tries to find the index set $\mathcal{K}$ so that all data points are well approximated by a linear combinations of the columns of $\tilde{X}(:,\mathcal{K})$. The problem (1) is a difficult combinatorial problem with $\binom{n}{r}$ possible solutions. However, what guarantee can we provide on the recovery of W? Is it more robust than LP-based formulations?

6. Summary of robustness results

	noise level γ ($\epsilon \leq \gamma$)	error δ ($q(\mathcal{K}) \leq \delta$)
SPA [3]	$\mathcal{O}\left(\frac{\sigma_{\min}(W)}{\sqrt{r}\,\text{cond}(W)^2}\right)$	$\mathcal{O}\left(\epsilon\, \text{cond}(W)^2\right)$
AnchorWords [4]	$\mathcal{O}\left(\frac{\sigma_{\min}(W)}{\sqrt{r}\,\text{cond}(W)^2}\right)$	$\mathcal{O}\left(\epsilon\, \text{cond}(W)\right)$
SNPA [6]	$\mathcal{O}\left(\beta(W)^4\right)$	$\mathcal{O}\left(\frac{\epsilon}{\beta(W)^3}\right)$
Prec-SPA [20]	$\mathcal{O}\left(\frac{\sigma_{\min}(W)}{r\sqrt{r}}\right)$	$\mathcal{O}\left(\epsilon\, \text{cond}(W)\right)$
LPs [7]	$\mathcal{O}\left(\kappa(W)^2\right)$	$\mathcal{O}\left(\frac{\epsilon}{\kappa(W)}\right)$
Hottopixx [11, 12], LP [13] $\equiv \ell_{1,q}$ [14]	$\mathcal{O}\left(\frac{\kappa(W) \min_{i \neq j} \lVert W(:,i) - W(,j) \rVert}{r}\right)$	$\mathcal{O}\left(\frac{r\epsilon}{\kappa(W)}\right)$
This paper, model (1)	$\frac{\kappa(W)}{4}$	$8\frac{\epsilon}{\kappa(W)} + \epsilon$

Table 1. Comparison of robust algorithms for separable SSMF. We have $\text{cond}(W) = \frac{\sigma_{\min}(W)}{\sigma_{\max}(W)}$, and $\beta(W)$ is another closely related quantity (taking nonnegativity into account).

6. Conclusion

We provided a tight robustness analysis for a combinatorial formulation of separable SSMF, showing it is more robust than greedy and LP-based approaches.

Implication: it is worth investigating combinatorial approaches in practice. In fact, as shown, e.g., in [18], heuristics for combinatorial approaches can improve the solutions provided by Greedy and LP-based algorithms:

Algorithm	Urban	Terrain
VCA	13.07% → 4.66%	18.61% → 3.29%
SPA	9.58% → 4.57%	5.89% → 3.37%
SNPA	9.63% → 4.91%	5.76% → 3.78%
LP	5.58% → 4.47%	3.34% → 3.01%

Table 2. Improvements in relative error for two hyperspectral images (Urban and Terrain) when applying a combinatorial local search heuristic from [18] on the solutions generated by different algorithms.