# SEPARABLE SIMPLEX-STRUCTURED MATRIX FACTORIZATION: ROBUSTNESS OF COMBINATORIAL APPROACHES

*Nicolas Gillis*

University of Mons, Rue de Houdain 9, 7000 Mons, Belgium

**ABSTRACT**

In this paper, we consider the following low-rank matrix approximation problem, referred to as separable simplex-structured matrix factorization: given an input matrix $X$, find $W$ and $H$ such that $X \approx WH$ where the columns of $W$ are chosen among the columns of $X$ and where the entries of each column of $H$ are nonnegative and sum to at most one. This problem has been studied extensively in the literature and is a generalization of separable nonnegative matrix factorization, with applications for example in hyperspectral unmixing and document analysis. Many methods have been proposed to tackle this problem; the three main classes are greedy algorithms, convex relaxations and combinatorial approaches. For the first two classes, robustness to noise of several algorithms have been characterized precisely. As far as we know, no such result exist for combinatorial formulations. This paper fills in this gap: we provide a tight robustness analysis of an exact combinatorial formulation of the problem. Although such formulations are difficult to optimize, we show that they lead to stronger robustness to noise than greedy algorithms and convex relaxations.

***Index Terms*—** nonnegative matrix factorization, separability, sum-to-one constraint, robustness to noise

## 1. INTRODUCTION

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 $X \geq 0$. However, most algorithms for separable NMF can be used for separable SSMF as nonnegativity of $W$ is not key in their design and analysis. This problem has many applications such as hyperspectral unmixing, document analysis and time-resolved Raman spectroscopy, to cite a few. We refer the reader to the recent survey [1] for more applications and details about SSMF.

**Remark 1** (Simplex constraint)**.** *In previous works [1], SSMF refers to the problem where the columns of $H$ live on the unit simplex where the sum of the entries of each column of $H$ is equal to one, as opposed to being smaller than one as in our definition of $\Delta^r$. Our formulation is slightly more general hence more flexible. For example, in hyperspectral unmixing, it allows to take into account different intensities of light among the pixels in the image. This is the reason why we prefer this formulation.*

**Problem statement** In this paper, we will assume that the input matrix $\tilde{X}$ has the following form.

**Assumption 1.** *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) $H(:,j) \in \Delta^r$ for all $1 \leq j \leq n$, 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 $[r] = \{1, 2, \dots, r\}$. The quantity $\kappa(W)$ is related to the conditioning of the convex hull of the columns of $W$, and $\kappa(W) > 0$ if and only if no column of $W$ is contained in the convex hull of the other columns and the vector of zeros. For simplicity of the presentation, we will assume w.l.o.g. that $\max_k ||W(:,k)||_2 = 1$ (if this does not hold, $X$ can be multiplied by a positive constant). Under this assumption, $\kappa(W) \leq 1$. Given $X$, the input noisy matrix $\tilde{X}$ is given by $\tilde{X} = X + N$ with $||N(:,j)||_2 \leq \epsilon$ for all $j$.*

Finally, separable SSMF is the problem of finding the index set $\mathcal{K}$ of size $r$ that allows to recover $W$ as best as possible, given $\tilde{X}$ and $r$, that is, minimizing

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

NG acknowledges the support by the European Research Council (ERC starting grant n° 679515) and by the Fonds de la Recherche Scientifique - FNRS and the Fonds Wetenschappelijk Onderzoek - Vlanderen (FWO) under EOS Project no O005318F-RG47.

In this paper, we will be interested in characterizing the robustness to noise of different algorithms that aim to recover $W$ given $\tilde{X}$. More precisely, given that $\tilde{X}$ satisfies Asm. 1, we would like to bound the error an algorithm makes when recovering the columns of $W$, that is, provide some $\delta$ such that $q(\mathcal{K}) \leq \delta$, given that $\epsilon$ is sufficiently small, that is, given that $\epsilon \leq \gamma$ for some $\gamma$. The quantities $\gamma$ and $\delta$ depend on the algorithm and on the conditioning of the problem; see Table 1 for some examples. The condition $\epsilon \leq \gamma < \kappa$ is necessary for any algorithm, because otherwise a column of $W$ can be perturbed in such a way that it belongs the convex hull of the others, which of course makes it impossible to detect. Also, $\delta \geq \epsilon$ since the noise perturbs all columns of $X$ including the column of $W$.

**Algorithms** Algorithms for separable SSMF can be classified in three classes:

1. *Greedy algorithms*. This class of algorithms identifies sequentially the columns of $W$ using different strategies. Usually, they follow a two-step strategy: (1) selection step, and (2) projection step. For example, vertex component analysis (VCA) [2] uses linear functions for the selection step and orthogonal projection for the projection step, the successive projection algorithm (SPA) [3] and related methods such as FastAnchorWords [4] use strongly convex functions for the selection step while also using orthogonal projection for the projection step. XRAY [5] and the successive nonnegative projection algorithm (SNPA) [6] take nonnegativity into account within the projection step. Preconditioned SPA (Prec-SPA) performs a preconditioning to improve robustness to noise of SPA. It is, as far as we know, the most robust greedy algorithm for separable SSMF when $W$ is full rank. Most of the greedy algorithms have low computational cost and memory requirement but are less robust to noise.

2. *Convex relaxations*. The first provably correct algorithm for separabe SSMF was proposed in [7] and is based on computing the distance of each data points from the convex hull of the other data points using linear programming [7]. Most convex relaxation are based on the following reformulation for separable SSMF: Find $Y$ with $r$ non-zero rows such that $X = XY$. Promoting row sparsity can be achieved for example by minimizing $||Y||_{1,q} = \sum_i ||Y(i,:)||_q$ [8, 9, 10] or using linear programming [11, 12, 13]. It turns out that these two types of formulations are closely related, being essetially equivalent [14], and significantly more robust than greedy algorithms; see Table 1. However, they require the introduction of the $n$-by-$n$ matrix $Y$ which is computationally heavy (for example, in hyperspectral imaging, $n$ is typically of the order of millions). Note that a possible way to overcome this issue is to selecting a subset of potential candidates among the columns of $X$ [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 $||.||$ which is a convex optimization problem. Optimizing $g$ over $\mathcal{K}$ is a difficult combinatorial problem with $\binom{n}{r}$ possible solutions. However, many heuristic have been proposed to tackle it, e.g., the famous local neighborhood approach N-FINDR [15], ant-colony optimization [16], bee-colony and genetic algorithms [17], alternating optimization [18], and approximation algorithms [19], to cite a few. Similarly as for convex relaxations, this approach can be made computationally cheaper by imposing the elements of $\mathcal{K}$ to be in a subset of all columns of $X$, e.g., columns identified by other algorithms as in [16] which reduces drastically the number of possible solutions.

Table 1 compares the different robustness recovery results for the different provably correct algorithms described in the previous paragraph.

**Outline of the paper and contribution** As far as we know, no robustness result has been provided for combinatorial approaches. This is the main contribution of this paper. Although we cannot expect to compute optimal solutions for these problems in general due to their complexity [19], it is still interesting to know whether this model leads to stronger (and tight) robustness for separable SSMF. If it was not the case, there would be no incentive to use such combinatorial approaches.

In this paper, we prove that a particular combinatorial formulation, namely (1), achieves the best possible bound, up to some constant factor. In fact, we prove in Theorem 2 that (1) achieves an error of $8\frac{\epsilon}{\kappa(W)} + 2\epsilon$ given that $\epsilon < \frac{\kappa}{4}$, while it is not possible to achieve a lower error up to some constant factor, as proved in Theorem 3. This improves the bound for the LP-based approaches either by a factor $\kappa(W)$ in the noise allowed [7] or by a factor $r$ on both the noise and error bounds for the aproaches from [11, 8, 12, 13]; see Table 1.

## 2. PROBLEM DEFINITION AND COMBINATORIAL FORMULATION

The most natural formulation that tries to recover $W$ from $\tilde{X}$ is the following:

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

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, assuming we can solve this problem, what guarantee can we provide on the recovery of $W$? Is it more robust than LP-based formulations? We answer these questions below.

## 3. ROBUSTNESS TO NOISE OF MODEL (1)

In this section, we prove robustness of using (1) to solve the near-separable NMF problem. The first Lemma shows that the solution $\mathcal{K}^*$ achieves an error $f(\mathcal{K}^*)$ of at most $2\epsilon$.

| | noise level $\gamma$ ($\epsilon \leq \gamma$) | error $\delta$ ($q(\mathcal{K}) \leq \delta$) |
|---|---|---|
| SPA [3] | $\mathcal{O}\left(\frac{\sigma_{\min}(W)}{\sqrt{r}\,\mathrm{cond}(W)^2}\right)$ | $\mathcal{O}\left(\epsilon\,\mathrm{cond}(W)^2\right)$ |
| AnchorWords [4] | $\mathcal{O}\left(\frac{\sigma_{\min}(W)}{\sqrt{r}\,\mathrm{cond}(W)^2}\right)$ | $\mathcal{O}\left(\epsilon\,\mathrm{cond}(W)\right)$ |
| SNPA [6] | $\mathcal{O}\left(\beta(W)^4\right)$ | $\mathcal{O}\left(\frac{\epsilon}{\beta(W)^3}\right)$ |
| Prec-SPA [20] | $\mathcal{O}\left(\frac{\sigma_{\min}(W)}{r\sqrt{r}}\right)$ | $\mathcal{O}\left(\epsilon\,\mathrm{cond}(W)\right)$ |
| 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] | $\mathcal{O}\left(\frac{\kappa(W)\min_{i\neq j}\|\|W(:,i)-W(,j)\|\|}{r}\right)$ | $\mathcal{O}\left(\frac{r\epsilon}{\kappa(W)}\right)$ |
| This paper, model (1) | $\frac{\kappa(W)}{4}$ | $8\frac{\epsilon}{\kappa(W)}+\epsilon$ |

**Table 1**. Comparison of robust algorithms for separable SSMF applied on a matrix satisfying Asm. 1. The condition number of $W$ is denoted $\mathrm{cond}(W) = \frac{\sigma_{\min}(W)}{\sigma_{\max}(W)}$, while $\beta(W)$ is the minimum between the norms of the residuals of the projections of the columns of $W$ onto the convex hull of the other columns of W, and the distances between these residuals [6].

**Lemma 1.** *For a matrix $\tilde{X}$ satisfying Asm. 1, we have $f(\mathcal{K}^*) \leq 2\epsilon$.*

*Proof.* We have $\tilde{x}_j = x_j + n_j = X(:,\mathcal{K}^*)h_j + n_j$ and $\tilde{W} = \tilde{X}(:,\mathcal{K}^*) = W + N(:,\mathcal{K}^*)$. Therefore, we have for all $j$,

$$\begin{aligned}\min_y ||\tilde{x}_j - \tilde{W}y||_2 &\leq \min_y ||x_j - Wy||_2 + ||Ny||_2 + ||n_j||_2 \\ &\leq \min_y ||x_j - Wy||_2 + \max_y ||Ny||_2 + \epsilon = 2\epsilon,\end{aligned}$$

since $\max_{y\in\Delta} ||Ny||_2 = \max_j ||N(:,j)||_2 \leq \epsilon$ (Asm. 1). □

The second Lemma provides a lower bound on the error for any solution $\mathcal{K}$.

**Lemma 2.** *Let $\tilde{X} = X + N$ satisfy Asm. 1. Let $\mathcal{K}$ be an index set of size $r$ and let $B = H(:,\mathcal{K}) \in \mathbb{R}^{r\times r}$ so that $X(:,\mathcal{K}) = WB$. Let $\alpha = \min_j \max_k B(j,k) \leq 1$. We have $f(\mathcal{K}) \geq (1-\alpha)\kappa - 2\epsilon$.*

*Proof.* Clearly, this holds for $\alpha = 1$ since $f(\mathcal{K}) \geq 0$ (in that case, $B$ is actually the identity matrix, up to permutation). Otherwise, $\alpha < 1$ and let $j$ be such that $\alpha = \max_k B(j,k)$. We have, using a similar derivation as in Lemma 1,

$$\min_h ||\tilde{W}(:,j) - \tilde{X}(:,\mathcal{K})h||_2 \geq \min_h ||W(:,j) - X(:,\mathcal{K})h||_2 - 2\epsilon.$$

Note that for any $h \in \Delta$, $(Bh)_j = \sum_k B(j,k)h_k \leq \max_k B(j,k) = \alpha$. Then,

$$\begin{aligned}\min_h ||W(:,j) - X(:,\mathcal{K})h||_2 &= \min_h ||W(:,j) - WBh||_2 \\ &= \min_h ||(1-(Bh)_j)W(:,j) - W(:,[r]\backslash\{k\})(Bh)_{[r]\backslash\{k\}}||_2 \\ &\geq \min_{h\in\delta}(1-(Bh)_j)||W(:,j) - W(:,[r]\backslash\{j\})\frac{(Bh)_{[r]\backslash\{j\}}}{(1-(Bh)_j)}||_2 \\ &\geq (1-\alpha)\kappa.\end{aligned}$$

In fact, $Bh \in \Delta^r$ hence $\frac{(Bh)_{[r]\backslash\{j\}}}{(1-(Bh)_j)} \in \Delta^{r-1}$ for $(Bh)_j \leq \alpha < 1$. □

**Without duplicates and near-duplicates** Under Asm. 1, the matrix $X$ can be written as

$$X = W\,[I_r, H']\,\Pi, \tag{2}$$

for some permutation matrix $\Pi$. If the entries of $H'$ are strictly smaller than one, there are no duplicated columns of $W$ in the data set. Under this condition, we can prove the following robustness of recovering $\mathcal{K}^*$ using model (1).

**Theorem 1.** *Let $\tilde{X} = X + N$ satisfy Asm. 1 where $X$ has the form* (2). *Let us assume that $\epsilon < \frac{(1-\beta)\kappa}{4}$ where $\beta = \max_{i,j} H'(i,j) < 1$. Then the optimal solution of* (1) *is $\mathcal{K}^*$ hence $\max_{1\leq k\leq r}\min_{j\in\mathcal{K}} ||W(:,k) - \tilde{X}(:,j)||_2 \leq \epsilon$.*

*Proof.* Since any solution other than $\mathcal{K}^*$ will have error at least $(1-\beta)\kappa - 2\epsilon$ (Lemma 2) and $\mathcal{K}^*$ has error at most $2\epsilon$ (Lemma 1), $\mathcal{K}^*$ is the unique optimal solution as long as $(1-\beta)\kappa - 2\epsilon > 2\epsilon$ while $\mathcal{K}^*$ leads to an error on $W$ of less than $\epsilon$ (Asm. 1). □

The bound of the theorem above is tight since we cannot have error smaller than $\epsilon$ by construction. In this particular setting (no duplicated nor near duplicated columns of $W$), LP-based formulations provide error smaller than $\epsilon$ when $\epsilon < \frac{(1-\beta)\kappa}{10}$ [13], hence provide as good solutions, up to a factor $\frac{5}{2}$. Unfortunately, in many practical settings, there are duplicated and near-duplicated columns of $W$ in the data set. For example, in hyperspectral images satisfying the pure-pixel assumption, there are usually more than one pure pixel per endmember, and many pixels contain mostly one material.

**With duplicates and near-duplicates** In the presence of duplicated and near-duplicated columns of $W$, the robustness result is the following.

**Theorem 2.** *Let $\tilde{X}$ satisfy Assumption 1 and let us assume that $\epsilon < \frac{\kappa}{4}$. Then any optimal solution $\mathcal{K}$ of* (1) *satisfies*

$$\max_{1 \leq k \leq r} \min_{j \in \mathcal{K}} ||W(:,k) - \tilde{X}(:,j)||_2 \leq \frac{8\epsilon}{\kappa} + \epsilon.$$

*Proof.* First, note that $||W(:,j) - \tilde{X}(:,j)||_2 \leq ||W(:,j) - X(:,j)||_2 + \epsilon$. Then, let $X(:,\mathcal{K}) = WB$ with $B = H(:,\mathcal{K})$. Let $\alpha = \min_j \max_k B(j,k) \leq 1$. By optimality of $\mathcal{K}$ and by Lemmas 1 and 2, we have

$$2\epsilon \geq f(\mathcal{K}^*) \geq f(\mathcal{K}) \geq \kappa(1-\alpha) - 2\epsilon.$$

This implies that $\kappa(1-\alpha) \leq 4\epsilon \iff \alpha \geq 1 - \frac{4\epsilon}{\kappa}$. By definition $\alpha \leq \max_k B(j,k)$, hence for each row of $B$ there is at least one entry with value $1 - \frac{4\epsilon}{\kappa}$, that is, for each $j$ there exists $k_j$ such that $B(j,k_j) \geq 1 - \frac{4\epsilon}{\kappa}$. Hence we have that for all $j$, there exists $k_j$ such that

$$\begin{aligned}
&||W(:,j) - WB(:,k_j)||_2 \\
&= ||W(:,j) - W(:,j)B(j,k_j) - W_{:,[r]\backslash\{j\}}B([r]\backslash\{j\},j_k)||_2 \\
&= ||(1 - B(j,k_j))W(:,j) - W_{:,[r]\backslash\{j\}}B([r]\backslash\{j\},j_k)||_2 \\
&\leq \frac{4\epsilon}{\kappa}||W(:,j)||_2 + \frac{4\epsilon}{\kappa} \max_{h \in \Delta^{r-1}} ||W_{:,[r]\backslash\{j\}}h||_2 \\
&\leq \frac{8\epsilon}{\kappa} \max_k ||W(:,k)||_2 = \frac{8\epsilon}{\kappa},
\end{aligned}$$

since $B(:,k_j) \in \Delta^r$ and $B(j,k_j) \geq 1 - \frac{4\epsilon}{\kappa}$ implies that $||B([r]\backslash\{j\},j_k)||_1 \leq \frac{4\epsilon}{\kappa}$ and $B(j,k_j) \leq \frac{4\epsilon}{\kappa}$, while we have $\max_k ||W(:,k)||_2 = 1$ by Assumption 1. □

**Oracle and optimal bound** In the following, we prove a lower bound on the best possible accuracy achievable by solving (1). This proves that the bounds in Theorem 2 are tight up to a (small) multiplicative factor.

**Theorem 3.** *There exists a class of near-separable matrices $\tilde{X}$ satisfying Assumption 1 and with $\epsilon < \frac{\kappa}{4}$ such that the optimal solution of* (1) *satisfies*

$$\max_{1 \leq k \leq r} \min_{j \in \mathcal{K}} ||W(:,k) - \tilde{X}(:,j)||_2 > \sqrt{2}\frac{\epsilon}{\kappa} + \sqrt{2}\epsilon.$$

*Proof.* Let $W = \begin{pmatrix} 1 & 0 & 1/2 + \frac{\kappa}{\sqrt{2}} \\ 0 & 1 & 1/2 + \frac{\kappa}{\sqrt{2}} \end{pmatrix}$ and let $\kappa \leq 1 - \sqrt{2}/2$ so that $\max_k ||W(:,k)||_2 = 1$ and $\kappa(W) = \kappa$. Let also $H = [I, h]$ where $h = (1-\lambda, 0, \lambda)$ for some $\lambda \in [0,1]$ to be defined later. This implies that the fourth column of $X = WH$ is a linear combination of the first and third column of $W$, that is, $X(:,4) = (1-\lambda)W(:,1) + \lambda W(:,3) = (1 - \lambda/2 + \kappa\lambda/\sqrt{2}, \lambda/2 + \kappa\lambda/\sqrt{2})$. We only add noise to the third column of $X$ using $N(:,3) = -(\epsilon,\epsilon)/\sqrt{2}$ with $||N(:,3)||_2 = \epsilon$; see Figure 1 for an illustration.

In the following, we show that, for a suitable choice of $\lambda$, the optimal solution of (1) is $\{1,2,4\}$ and that $||W(:,3) - \tilde{X}(:,4)||_2 > \sqrt{2}\frac{\epsilon}{\kappa} + \sqrt{2}\epsilon$ which will conclude the proof.

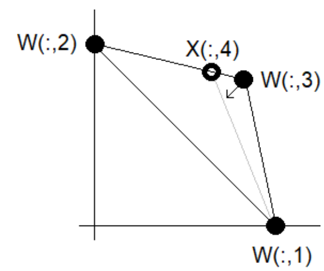


**Fig. 1**. Illustration of the construction from Theorem 3.

Let us first compute $||W(:,3) - X(:,4)||_2 = ||W(:,3) - (1-\lambda)W(:,1) - \lambda W(:,3)||_2 = (1-\lambda)||W(:,3) - W(:,1)||_2 > \frac{1}{\sqrt{2}}(1-\lambda)$, since $||W(:,3) - W(:,1)||_2 = ||(1/2 - \frac{\kappa}{\sqrt{2}}, 1/2 + \frac{\kappa}{\sqrt{2}})||_2 > ||(1/2,1/2)||_2 = \frac{1}{\sqrt{2}}$.

Let us define the vector $v = \mu X(:,4) + (1-\mu)W(:,2)$ with $\mu = \frac{1}{2-\lambda}$ so that $v = (v_1, v_2)$ with $v_1 = v_2 = \frac{1}{2} + \kappa\frac{\lambda}{\sqrt{2}(2-\lambda)}$. The vector $v$ will approximate $\tilde{X}(:,3)$ using $\tilde{X}(:,2)$ and $\tilde{X}(:,4)$. We have

$$||W(:,3) - v||_2 = \frac{\kappa}{\sqrt{2}}\left(1 - \frac{\lambda}{2-\lambda}\right) = \sqrt{2}\kappa\frac{1-\lambda}{2-\lambda}.$$

Let us choose $\lambda$ such that $||W(:,3) - v||_2 < 2\epsilon$ (note that $N(:,3)$ makes $W(:,3)$ goes toward $v$ hence this will imply that $||\tilde{X}(:,3) - v||_2 < \epsilon$). Denoting $\delta_\lambda = 1 - \lambda$, this requires $\sqrt{2}\kappa\frac{\delta_\lambda}{1+\delta_\lambda} < 2\epsilon \iff \delta_\lambda < \frac{2\epsilon}{\sqrt{2}\kappa - 2\epsilon} \iff \lambda > 1 - \frac{2\epsilon}{\sqrt{2}\kappa - 2\epsilon}$. Assuming $\epsilon < \kappa/4$, this condition is implied by $\delta_\lambda < \frac{2\epsilon}{\sqrt{2}\kappa - \kappa/2} = \frac{2}{\sqrt{2} - 1/2}\frac{\epsilon}{\kappa}$ hence also by $\delta_\lambda < 2\frac{\epsilon}{\kappa}$ since $\sqrt{2} - 1/2 = 0.9142$. Finally, choosing $\lambda = 1 - 2\frac{\epsilon}{\kappa}$ leads to a problem where the solution $\mathcal{K} = \{1,2,4\}$ leads to a lower error than $\mathcal{K}^* = \{1,2,3\}$ (namely, $||\tilde{X}(:,3) - v||_2 < \epsilon$ vs. $\epsilon = ||\tilde{X}(:,3) - W(:,2)||_2$ –all other columns of $\tilde{X}$ are exactly reconstructed since no noise is added to them and they are extracted by $\mathcal{K}$) while $||X(:,4) - W(:,3)||_2 = (1 - \lambda)\sqrt{1+\kappa^2} \geq \frac{2\epsilon}{\kappa}\frac{1+\kappa}{\sqrt{2}} = \sqrt{2}\frac{\epsilon}{\kappa} + \sqrt{2}\epsilon$, since $\sqrt{1+\kappa^2} \geq \sqrt{2} + \frac{\kappa-1}{\sqrt{2}} = \frac{1+\kappa}{\sqrt{2}}$ (first-order Taylor approximation around $\kappa = 1$ of the convex function $\sqrt{1+\kappa^2}$). □

## 4. CONCLUSION

In this paper, we have proved, for the first time, tight robustness bounds for the combinatorial model (1) that tackles separable SSMF. Although (1) is difficult to tackle, it leads to more robust solutions than other approaches hence it makes sense in practice to devise (heuristic) algorithms for this type of models. In particular, this justifies the use of these algorithms initialized by solutions obtained by greedy methods and convex relaxations to improve these solutions further; as done for example in [16, 18]. Further research include extensive numerical comparison between Greedy algorithms, convex relaxations and combinatorial approaches to asses their robustness in practical scenarios.

## 5. REFERENCES


[1] Xiao Fu, Kejun Huang, Nicholas D Sidiropoulos, and Wing-Kin Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications," *IEEE Signal Processing Magazine*, 2019, to appear.

[2] José M. P. Nascimento and José M. 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.

[3] Nicolas Gillis and Stephen A Vavasis, "Fast and robust recursive algorithmsfor separable nonnegative matrix factorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 698–714, 2014.

[4] Sanjeev Arora, Rong Ge, Yonatan Halpern, David Mimno, Ankur Moitra, David Sontag, Yichen Wu, and Michael Zhu, "A practical algorithm for topic modeling with provable guarantees," in *International Conference on Machine Learning*, 2013, pp. 280–288.

[5] Abhishek Kumar, Vikas Sindhwani, and Prabhanjan Kambadur, "Fast conical hull algorithms for near-separable non-negative matrix factorization," in *International Conference on Machine Learning*, 2013, pp. 231–239.

[6] Nicolas Gillis, "Successive nonnegative projection algorithm for robust nonnegative blind source separation," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1420–1450, 2014.

[7] Sanjeev Arora, Rong Ge, Ravindran Kannan, and Ankur Moitra, "Computing a nonnegative matrix factorization–provably," in *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. ACM, 2012, pp. 145–162.

[8] Ernie Esser, Michael Moller, Stanley Osher, Guillermo Sapiro, and Jack Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3239–3252, 2012.

[9] René Vidal, Guillermo Sapiro, and E Elhamifar, "See all by looking at a few: Sparse modeling for finding representative objects," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[10] Xiao Fu and Wing-Kin Ma, "Robustness analysis of structured matrix factorization via self-dictionary mixed-norm optimization," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 60–64, 2016.

[11] Ben Recht, Christopher Re, Joel Tropp, and Victor Bittorf, "Factoring nonnegative matrices with linear programs," in *Advances in Neural Information Processing Systems*, 2012, pp. 1214–1222.

[12] Nicolas Gillis, "Robustness analysis of hottopixx, a linear programming model for factoring nonnegative matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 1189–1212, 2013.

[13] Nicolas Gillis and Robert Luce, "Robust near-separable nonnegative matrix factorization using linear optimization.," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1249–1280, 2014.

[14] Nicolas Gillis and Robert Luce, "A fast gradient method for nonnegative sparse regression with self dictionary," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 24–37, 2018.

[15] Michael E Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Imaging Spectrometry V*. International Society for Optics and Photonics, 1999, vol. 3753, pp. 266–276.

[16] Lianru Gao, Jianwei Gao, Jun Li, Antonio Plaza, Lina Zhuang, Xu Sun, and Bing Zhang, "Multiple algorithm integration based on ant colony optimization for endmember extraction from hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2569–2582, 2015.

[17] Zheng Fu, Chi-Man Pun, Hao Gao, and Huimin Lu, "Endmember extraction of hyperspectral remote sensing images based on an improved discrete artificial bee colony algorithm and genetic algorithm," *Mobile Networks and Applications*, Sep 2018.

[18] Jérémy E. Cohen and Nicolas Gillis, "Dictionary-based tensor canonical polyadic decomposition," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1876–1889, 2018.

[19] Greg Van Buskirk, Benjamin Raichel, and Nicholas Ruozzi, "Sparse approximate conic hulls," in *Advances in Neural Information Processing Systems*, 2017, pp. 2534–2544.

[20] Nicolas Gillis and Stephen A Vavasis, "Semidefinite programming based preconditioning for more robust near-separable nonnegative matrix factorization," *SIAM Journal on Optimization*, vol. 25, no. 1, pp. 677–698, 2015.