Checking the Sufficiently Scattered Condition using a Global Non-Convex Optimization Software

Nicolas Gillis, *Member, IEEE*, Robert Luce

Abstract—The sufficiently scattered condition (SSC) is a key condition in the study of identifiability of various matrix factorization problems, including nonnegative, minimum-volume, symmetric, simplex-structured, and polytopic matrix factorizations. The SSC allows one to guarantee that the computed matrix factorization is unique/identifiable, up to trivial ambiguities. However, this condition is NP-hard to check in general. In this paper, we show that it can however be checked in a reasonable amount of time in realistic scenarios, when the factorization rank is not too large. This is achieved by formulating the problem as a non-convex quadratic optimization problem over a bounded set. We use the global non-convex optimization software Gurobi, and showcase the usefulness of this code on real-world hyperspectral images.

Index Terms—sufficiently scattered condition, identifiability, uniqueness, nonnegative matrix factorization, non-convex quadratic optimization

I. INTRODUCTION

$\mathbf{L}$OW-RANK matrix factorizations (LRMFs) are central techniques in numerical linear algebra, with the singular value decompositions (SVD) and principal component analysis (PCA) as the most famous examples. LRMFs are widely used in data analysis, statistics, signal processing, control, optimization, and machine learning; see, e.g., [1], [2], [3]. Given an input matrix $X \in \mathbb{R}^{m \times n}$ and a factorization rank r, LRMF aims at finding $W \in \mathcal{W} \subseteq \mathbb{R}^{m \times r}$ and $H \in \mathcal{H} \subseteq \mathbb{R}^{r \times n}$ such that $X \approx WH$. The sets $\mathcal{W}$ and $\mathcal{H}$ impose constraints on W and H, such as orthogonality in SVD and PCA, and sparsity in sparse PCA [4]. These additional constraints allow one to more easily interpret the factors, and are often motivated by the application at hand.

Among LRMFs, nonnegative matrix factorization (NMF) [5], which imposes nonnegative constraint on the factors and assumes X is nonnegative, has become a standard tool as well; see [6], [7], [8] and the references therein. Nonnegativity is motivated for example by physical considerations, e.g., in imaging, audio signal processing and chemometrics, or by probabilistic interpretations, e.g., in topic modeling. A key aspect in these applications is that the factorization is unique, a.k.a. identifiable (we will use both terms interchangeably), which allows one to recover the groundtruth factors that generated the data (such as the sources in blind source separation). A factorization, $X = WH$, is unique/identifiable if for any other factorization, $X = W'H'$,

there exists a permutation π of $\{1, 2, \ldots, r\}$ and scaling factors $\{\alpha_k\}_{k=1}^{r}$ such that for all k

$$W'(:, k) = \alpha_k W(:, \pi_k) \text{ and } H'(k, :) = \alpha_k^{-1} H(\pi_k, :). \quad (1)$$

Unfortunately, nonnegativity is typically not enough to ensure the uniqueness; see [9], [8]. A stronger condition that ensures uniqueness is the sufficiently scattered condition (SSC) which requires some degree of sparsity within the factors W and H; see Section II for a formal definition. The uniqueness of NMF under the SSC was presented in [10], and later lead to numerous identifiability results for other LRMFs, namely:

- Minimum-volume NMF [11], [12] which seeks for an NMF decomposition that minimizes the volume of the convex full of the columns of W.
- Simplex-structured matrix factorization [13], [14] which relaxes the constraints on W but imposes H to be column-wise stochastic. A bounded version, where the entries of W are bounded [15], has also been explored and shown to be identifiable under SSC-like conditions.
- Symmetric NMF with applications in topic modeling [16], [17].

The SSC condition was also generalized for polytopic matrix factorizations [18], [19] and for bounded component analysis [20] which are generalizations of NMF where the nonnegative orthant is replaced by polytopes. The SSC has also been used to provide identifiability in many other contexts where constrained matrix factorizations play a crucial role. This has been the case in particular in machine learning tasks, such as topic modeling [21], [17], crowd sourcing [22], recovering joint probability [23], label-noise learning [24], deep constrained clustering [25], dictionary learning [26], and tensor decompositions [27].

Outline and contribution. In summary, the SSC plays a critical role in checking whether a wide class of LRMFs with nonnegativity constraints are identifiable. Unfortunately, the SSC is NP-hard to check in general (see Section II for more details). To the best of our knowledge, there currently does not exist a solver that checks the SSC, even for small-size problems. In this paper, we overcome this limitation by leveraging current global non-convex quadratic optimization software, and in particular Gurobi, to check the SSC for relatively large matrices.

The paper is organized as follows. We first define the SSC rigorously, in Section II, after having introduced useful concepts in convex geometry. Then we show in Section III how checking the SSC is equivalent to solving an non-convex quadratic program. In Section IV, we provide an alternative formulation with box constraints, which is crucial to using

global non-convex quadratic optimization software. We report numerical experiments in Section V on hyperspectral images, showing that Gurobi can solve relatively large instances, for a factorization rank r up to a few dozen, and input matrices of size up to a few thousands.

Notation. Given a vector $x \in \mathbb{R}^r$, $\|x\|_2$ denotes its ℓ_2 norm and $x^\top$ its transpose. The nonnegative orthant in dimension r is denoted $\mathbb{R}_+^r$. The vector of all ones is denoted e, and of all zeros 0. The identity matrix is denoted I, and its ith column is denoted e_i (a.k.a. the ith unit vector). The dimensions of e, 0, I and e_i will be clear from the context. For a matrix H, $H(:,j)$ and $H(i,:)$ denote its jth column and ith row, respectively. The set $\{1, 2, \ldots, r\}$ is denoted $[r]$.

II. PRELIMINARIES: CONES, THEIR DUALS AND THE SSC

Given a matrix $H \in \mathbb{R}^{r \times n}$, we define the cone generated by its columns as $\mathrm{cone}(H) = \{x \mid x = Hy, y \geq 0\}$. The dual of a cone $\mathcal{H}$ is defined as $\mathcal{H}^* = \{x \mid x^\top z \geq 0 \text{ for all } z \in \mathcal{H}\}$. In particular, the dual cone of $\mathrm{cone}(H)$ is given by

$$
\begin{aligned}
\mathrm{cone}^*(H) &= \{x \mid x^\top z \geq 0, \text{ for all } z \in \mathrm{cone}(H)\} \\
&= \{x \mid x^\top Hy = (H^\top x)^\top y \geq 0, \text{ for all } y \geq 0\} \\
&= \{x \mid H^\top x \geq 0\}.
\end{aligned}
$$

Another cone we will need is the following second-order cone:

$$
\mathcal{C} = \{x \in \mathbb{R}^r \mid e^\top x \geq \sqrt{r-1}\|x\|_2\},
$$

which is contained in the nonnegative orthant. Its dual cone is $\mathcal{C}^* = \{x \in \mathbb{R}^r \mid e^\top x \geq \|x\|_2\}$ and contains the nonnegative orthant (which is self-dual). An important and easy-to-check property of duality is that $\mathcal{H}_1 \subseteq \mathcal{H}_2$ if and only if $\mathcal{H}_2^* \subseteq \mathcal{H}_1^*$.

The SSC. We will use the following definition.

Definition 1 (SSC, [10]). *A nonnegative matrix $H \in \mathbb{R}_+^{r \times n}$ satisfies the sufficiently scattered condition (SSC) if*

$$
\mathcal{C} \quad \subseteq \quad \mathrm{cone}(H), \quad and \tag{SSC1}
$$

$$
\begin{aligned}
&q \in \mathrm{cone}^*(H) \cap \{x \mid e^\top x = \|x\|_2\} \\
&\iff q = \alpha e_i \text{ for } \alpha > 0 \text{ and } i \in [r]. \tag{SSC2}
\end{aligned}
$$

There actually exist several slight variations of the definition of the SSC. All of them include the requirement (SSC1), while the second condition is slightly modified:

- [11] requires that there does not exist any orthogonal matrix Q such that $\mathrm{cone}(H) \subseteq \mathrm{cone}(Q)$, except for permutation matrices. This is a slight relaxation of (SSC2).
- [12] requires $\mathrm{cone}(H)$ to contain a slightly larger cone than $\mathcal{C}$, namely, $\mathcal{C}_q = \{x \in \mathbb{R}_+^r \mid e^\top x \geq q\|x\|_2\}$ for any $q < \sqrt{r-1}$, which is slightly more restrictive than (SSC2).

We chose Definition 1 because it is slightly simpler to present; however, our formulations can be adapted to handle the other definitions above.

III. CHECKING THE SSC

Before checking the SSC, it will be useful to check whether a simple necessary condition holds. The cone $\mathcal{C}$ contains the points $e - e_i$ for $i \in [r]$ at its border, since $e^\top(e - e_i) = \sqrt{r-1}\|e - e_i\|_2 = r - 1$, and hence $\mathcal{T} = \mathrm{cone}(ee^\top - I) \subset \mathcal{C}$. This implies the following necessary condition [8, p.119].

Definition 2 (Necessary Condition for the SSC, NC-SSC). *The matrix $H \in \mathbb{R}_+^{r \times n}$ satisfies the NC-SSC if $\mathcal{T} = \mathrm{cone}(ee^\top - I) \subset \mathrm{cone}(H)$, that is, $e - e_i \in \mathrm{cone}(H)$ for $i \in [r]$.*

The NC-SSC can be easily checked, in polynomial time, by solving systems of linear inequalities: for all $i \in [r]$, check that there exists $y \geq 0$ such that $e - e_i = Hy$. In the remainder, we will assume that this necessary condition has been checked, that is, $\mathcal{T} \subset \mathrm{cone}(H)$, otherwise H cannot satisfy the SSC.

Note that the NC-SSC (and hence the SSC) require a certain degree of sparsity of H, since its cone must contain the vectors $\{e - e_i\}_{i=1}^r$ that contain a zero entry. In fact, one can show that a necessary condition for the SSC to hold is that H has at least $r - 1$ zeros per row [10], [9], [8].

Checking the SSC via non-convex quadratic optimization. The first condition of the SSC (SSC1) is equivalent to $\mathrm{cone}^*(H) \subseteq \mathcal{C}^*$. This condition is not satisfied if there exists $x \in \mathrm{cone}^*(H)$ while $x \notin \mathcal{C}^*$, that is, if there exists x such that

$$
H^\top x \geq 0 \quad \text{and} \quad e^\top x < \|x\|_2. \tag{2}
$$

These conditions remain valid when x is multiplied by any positive constant. Moreover, we have the following lemma.

Lemma 1. *Let $H \in \mathbb{R}_+^{r \times n}$ satisfy the NC-SSC. Then for any $x \neq 0$ satisfying $H^\top x \geq 0$, we have $e^\top x > 0$.*

Proof. Since H satisfies the NC-SSC, $\mathcal{T} \subset \mathrm{cone}(H)$, and hence $\mathrm{cone}^*(H) \subset \mathcal{T}^* = \{x \in \mathbb{R}^r \mid (ee^\top - I)x \geq 0\}$. Summing the inequalities defining $\mathcal{T}^*$, we obtain

$$
e^\top(ee^\top - I)x = (r-1)e^\top x \geq 0.
$$

(For $r = 1$, the result is trivial since $H \neq 0$.) It therefore remains to show that, for $x \in \mathcal{T}^*$, $e^\top x = 0$ if and only if $x = 0$. In fact, $e^\top x = 0$ implies that all inequalities defining $\mathcal{T}^*$ must be active at x, otherwise their sum is positive and we would get $e^\top x > 0$, a contradiction. Hence $(ee^\top - I)x = 0$ implying $x = 0$ since $ee^\top - I$ is full rank. $\square$

Now, going back to (2), we have the following corollary.

Corollary 1. *Let $H \in \mathbb{R}_+^{r \times n}$ satisfy the NC-SSC. Then* (SSC1) *is not satisfied if and only if the system*

$$
H^\top x \geq 0 \quad and \quad e^\top x = 1 < \|x\|_2 \tag{3}
$$

has a solution.

Proof. Since $x = 0$ does not satisfy $e^\top x < \|x\|_2$, we can assume w.l.o.g., by Lemma 1, that $e^\top x > 0$, and hence, using the scaling degree of freedom, that $e^\top x = 1$. $\square$

Checking whether (3) has a solution for H satisfying the NC-SSC can be done by solving the following optimization problem:

$$p^* = \max_x \|x\|_2 \text{ such that } e^\top x = 1 \text{ and } H^\top x \geq 0. \tag{4}$$

If $p^* > 1$, then H does not satisfy (SSC1). Otherwise, $p^* = 1$, and to check whether H satisfies the SSC, we need to check the second condition (SSC2), that is, whether there exists $q \neq e_i$ for all i such that $e^\top q = \|q\|_2$ and $H^\top q \geq 0$.

In summary, H satisfies the SSC if and only if (i) H satisfies the NC-SSC, and (ii) the only optimal solutions of (4) are e_i for $i \in [r]$, with optimal value $p^* = 1$. This is why it is NP-hard to check the SSC, because (4) is the maximization of a convex function over a polytope which is NP-hard in general [28], [10].

IV. SOLVING (4) WITH GLOBAL NON-CONVEX QUADRATIC OPTIMIZATION

The problem (4) cannot directly be handled by global non-convex quadratic optimization solvers, such as Gurobi, because such solvers require a bounded feasible set, with lower and upper bounds on the variables. This allows them to use the McCormick envelopes [29], and rely on branch-and-bound strategies. In a few words, the idea is as follows: for every product of two variables that appears in the objective or in the constraints, say the product of the variables x and y, a new variable is introduced, $z = xy$. Given that x and y belong to a bounded box, that is, $x \in [\underline{x}, \bar{x}]$ and $y \in [\underline{y}, \bar{y}]$, the equality $z = xy$ is approximated from above and below with linear constraints as follows:

$$z \leq \underline{y}x + \bar{x}y - \bar{x}\underline{y}, z \leq \bar{y}x + \underline{x}y - \underline{x}\bar{y},$$
$$z \geq \underline{y}x + \underline{x}y - \underline{x}\underline{y}, z \geq \bar{y}x + \bar{x}y - \bar{x}\bar{y}.$$

Figure 1 provides an illustration of such a McCormick envelope. This McCormick linearization is then improved along
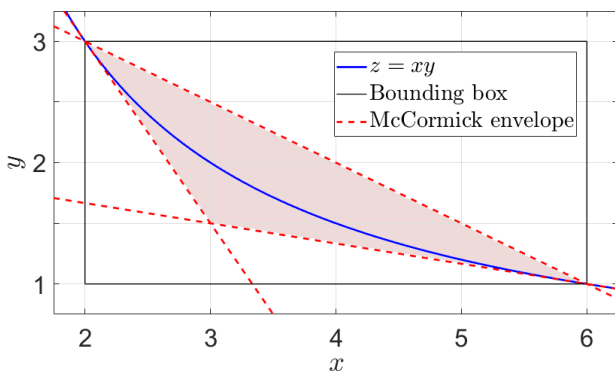

Fig. 1. Illustration of the McCormick envelope for the nonlinear constraint $xy = 6$, with $x \in [2, 6]$ and $y \in [1, 3]$. (Note that, in this example, the last two inequalities, defining the upper bound, coincide.)

the algorithm by splitting the feasible set in smaller and smaller intervals for each variable, following branch-and-bound strategies; see, e.g., [30] and the references therein.

Let us now show how any feasible solution x of (4) can be bounded.

Lemma 2. *Let $H \in \mathbb{R}_+^{r \times n}$ satisfy the NC-SSC. Then the entries of any feasible solution x of (4) satisfy $x_i \in [2 - r, 1]$ for all i.*

Proof. Since $\text{cone}(ee^\top - I) \in \text{cone}(H)$ for $i \in [r]$, we have $(ee^\top - I)x = e - x \geq 0$ and hence $x \leq e$. Since $e^\top x = 1$ and $x \leq e$, we have $x_i = 1 - \sum_{j \neq i} x_j \geq 1 - (r - 1) = 2 - r$ for all i. $\square$

Note that the lower bound in Lemma 2, $2 - r$, can be achieved: this is the case for $H = (ee^\top - I)$ for which one can check that optimal solutions of (4) are given by $e + (1 - r)e_i$ for $i \in [r]$.

Tightening the lower bound. When it comes to checking the SSC, we now show that we can actually tighten the constraint $x_i \geq 2 - r$ to $x_i \geq -1$. The reason is that we are not interested in the exact optimal value of (4), but only to know whether it is larger than one.

Lemma 3. *Let $H \in \mathbb{R}_+^{r \times n}$ satisfy the NC-SSC. Then the optimal value of (4) is equal to one, that is, $p^* = 1$, if and only of the optimal value*

$$q^* = \max_x \|x\|_2 \text{ s.t. } e^\top x = 1, H^\top x \geq 0, -1 \leq x_i \leq 1 \, i \in [r] \tag{5}$$

is equal to one.

Proof. For $r \leq 3$, Lemma 2 provides the result, since $x_i \geq 2 - r \geq -1$ for $i \in [r]$. It remains to prove the result for $r \geq 4$.

We have $p^* \geq q^*$ since (4) is a relaxation of (5), and that $q^* \geq 1$ since $x = e_i$ for all i are feasible solutions with objective equal to one[1]. It remains to show that $p^* > 1$ implies $q^* > 1$. Let x^* be an optimal solution of (4) with $p^* > 1$. By Lemma 2, $x^* \leq e$. If $x_i^* \geq -1$ for all i, we are done since x^* is feasible for (5) and $q^* = p^* > 1$. Otherwise assume $x_i^* < -1$ for some i. Let us consider the solution $y = \lambda x^* + (1 - \lambda)e_i$ for some $\lambda \in [0, 1]$ (to be chosen below). By convexity of the feasible set, y is feasible for (4). Let us take λ such that $y_i = -1$, that is,

$$\lambda = \frac{2}{1 - x_i^*} \in \left[\frac{2}{r - 1}, 1\right),$$

so that $\lambda \in (0, 1)$ since $2 - r \leq x_i^* < -1$ (Lemma 2) and $r \geq 4$. The new solution y satisfies $y_i = -1$, while the other entries of y are equal to that of x^* multiplied by $\lambda \in (0, 1)$, and hence their absolute value gets smaller. However, we have $\|y\|_2 > 1$ since $y_i = -1$ while $e^\top y = 1$. (In fact[2], we have $\|y\|_2^2 > 1 + \frac{1}{r-1}$.) Hence, we have constructed a feasible solution y of (4) such that $y_i = -1$ while $\|y\|_2 > 1$. For any other entry of y smaller than -1, we can apply the same trick as for x^*, and we will eventually get a feasible solution of (5) with at least one entry equal to -1, and hence an objective function value strictly larger than 1. $\square$

---

[1] Note that it is possible that $1 < q^* < p^*$, and we have some numerical examples.

[2] This follows from the inequalities $\sqrt{r - 1}\sqrt{\sum_{j \neq i}|y_j|^2} \geq \sum_{j \neq i}|y_j| \geq \sum_{j \neq i} y_j > 1$ which implies $\sum_{j \neq i}|y_j|^2 > \frac{1}{r-1}$.

In summary, we have the following theorem.

Theorem 1. *The matrix $H \in \mathbb{R}_+^{r \times n}$ satisfies the SSC if and only if the following two conditions are satisfied*
1) *The matrix H satisfies the NC-SSC, that is, $e - e_i \in \text{cone}(H)$ for all $i \in [r]$.*
2) *The optimal value of (5) is equal to $q^* = 1$, and the set of optimal solutions is $\{e_i\}_{i=1}^r$.*

Proof. This follows from Corollary 1 and Lemma 3. □

Gurobi: more than one solution, and early stopping. To check the SSC, we therefore have to check the NC-SSC, and then solve (5). If the optimal value is equal to one, we need to check whether there exist optimal solutions different from the r solutions which are the unit vectors, $\{e_i\}_{i=1}^r$. To solve (5), we rely on the global non-convex optimization software Gurobi[3]. It allows one to generate more than one solution, and we require Gurobi to provide at least $r + 1$ solutions[4] (to have at least one solution different from $\{e_i\}_{i=1}^r$).

Moreover, since we only care about checking whether the optimal value of (5) is larger than one, we can stop Gurobi as soon as the best found solution has objective strictly larger than 1, and we use the parameter: `params.BestObjStop = 1.0001`. This stopping criterion is key, as it allows Gurobi to stop very early for matrices far from satisfying the SSC.

Finally, we do not know in advance how long it will take for Gurobi to solve (5), and hence it is useful to use a timelimit (e.g., 5 minutes). If Gurobi cannot finish within 5 minutes, it means it was not able to find a solution with objective larger than 1. Hence, on top of the NC-SSC being satisfied (this is the first step of the algorithm), there has been additional necessary conditions satisfied, that is, all nodes explored within the branch-and-bound strategy do not violate the system (3). Hence although we cannot guarantee the SSC, the chances for the SSC to be satisfied are higher.

Finally, our algorithm to check the SSC, based on Theorem 1, is the first to check the SSC, up to machine precision.

V. NUMERICAL EXPERIMENTS

In [9], authors use a heuristic to solve (4), while, in [8, Chapter 4.2.3.6], the author relied on the necessary condition that the vectors $\{e - e_i\}_{i=1}^r$ belong to the relative interior of $\text{cone}(H)$. For the first time, we will provide results where the SSC is checked exactly on real hyperspectral images factorized with minimum-volume NMF.

The supplementary material [31] contains a details analysis for synthetic data. In a nutshell, the SSC can be checked very fast for $r \leq 10$ and $n = 10r$, in less than 6.3 seconds in all cases. When r increases, the computational time increases. For example, the time limit of 5 minutes is often reached when $r = 20$, $n = 10r$ and H is close to the phase transition between satisfying the SSC and not satisfying it.

All experiments are performed with a 12th Gen Intel(R) Core(TM) i9-12900H 2.50 GHz, 32GB RAM, on MATLAB R2019b. The code and data sets are available on https://gitlab.com/ngillis/check-ssc, also available in Python and Julia.

Given a matrix $X = W^\# H^\#$, where $H^\#$ satisfies the SSC, $(W^\#, H^\#)$ can be identified by solving the following minimum-volume (min-vol) NMF problem [32]:

$$\min_{W, H \geq 0} \det(W^\top W) \text{ s.t. } X = WH, \text{ and } W^\top e = e. \quad (6)$$

In practice, one has to balance the data fitting term, that is, $\|X - WH\|_F^2$, and the volume regularization[5], $\text{logdet}(W^\top W)$, by solving, for some penalty parameter $\lambda > 0$,

$$\min_{W \geq 0, W^\top e = e, H \geq 0} \|X - WH\|_F^2 + \lambda \, \text{logdet}(W^\top W). \quad (7)$$

We applied this model on 5 widely-used hyperspectral images[6]: the (i, j)th entry of matrix $X \in \mathbb{R}^{m \times n}$ contains the reflectance of the jth pixel at the ith wavelength. Each column of X is the spectral signature of a pixel, and each row a vectorized image at a given wavelength. Performing NMF on such a matrix allows one to extract the spectral signatures of the pure materials (called endmembers) as the columns of W, and their abundances in each pixel as the rows of H; see, e.g., [35]. We solve (7) with[7] the function `minvolNMF.m`, with the same parameters as in [8]. Table I reports the results. The supplementary material contains a link to the abundance maps of these solutions, illustrating the meaningful results obtained by min-vol NMF. In all cases, the SSC of the matrix

TABLE I
GIVEN A MIN-VOL NMF (W, H) OBTAINED BY SOLVING (7), THIS TABLE REPORTS THE SPARSITY OF H (THAT IS, PERCENTAGE OF ZERO ENTRIES, SPAR(H)), WHETHER THE NC-SSC (DENOTED NC) AND THE SSC ARE SATISFIED, AND THE TIME THAT IT TOOK TO CHECK THE SSC WITH GUROBI.

Data set	m	n	r	spar(H)	NC	SSC	time (s.)
Samson	156	95×95	3	40%	no	no	0.0
Terrain	166	500×307	4	40%	yes	no	2
Jasper	188	100×100	4	45%	yes	yes	0.5
Urban	162	307×307	6	50%	yes	yes	9
San Diego	158	400×400	7	53%	yes	yes	14

H could be checked within 15 s. The Terrain data set is interesting because it satisfies the NC-SSC, but not the SSC.

To the best of our knowledge, this is the first time the identifiability of the NMF of real-world hyperspectral images is guaranteed. More precisely, the solution (W^*, H^*) obtained by solving (7) on the noisy X is the unique solution of (6) when factorizing $X^* = W^* H^*$. Further research on noisy minimum-volume NMF is necessary to guarantee directly the uniqueness of the solution of (7).

VI. CONCLUSION

In this paper, we have provided a formulation, (5), suitable to check the SSC with non-convex quadratic optimization solvers (see Theorem 1), and we used Gurobi. This allows one to check, a posteriori, whether the solutions to various matrix factorization problems with nonnegativity constraints are essentially unique. We illustrated the use of our algorithm on the uniqueness of min-vol NMF for hyperspectral images.

---

[3]https://www.gurobi.com/solutions/gurobi-optimizer/
[4]`params.PoolSearchMode=2; params.PoolSolutions=r+1.`

[5]The regularizer $\text{logdet}(W^\top W)$ has been shown to perform better than $\det(W^\top W)$ in practice [33] .
[6]See http://lesun.weebly.com/hyperspectral-data-set.html and [34] from which we chose the values of r.
[7]Available from https://gitlab.com/ngillis/nmfbook/

References

[1] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications.* Springer Science & Business Media, 2011.

[2] M. Udell, C. Horn, R. Zadeh, and S. Boyd, "Generalized low rank models," *Foundations and Trends in Machine Learning*, vol. 9, no. 1, pp. 1–118, 2016.

[3] N. Kishore Kumar and J. Schneider, "Literature survey on low rank approximation of matrices," *Linear and Multilinear Algebra*, vol. 65, no. 11, pp. 2212–2244, 2017.

[4] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. G. Lanckriet, "A direct formulation for sparse PCA using semidefinite programming," *SIAM Review*, vol. 49, no. 3, pp. 434–448, 2007.

[5] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[6] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation.* John Wiley & Sons, 2009.

[7] X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 36, no. 2, pp. 59–80, 2019.

[8] N. Gillis, *Nonnegative matrix factorization.* SIAM, Philadelphia, 2020.

[9] X. Fu, K. Huang, and N. D. Sidiropoulos, "On identifiability of non-negative matrix factorization," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 328–332, 2018.

[10] K. Huang, N. D. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 211–224, 2013.

[11] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, "Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2306–2320, 2015.

[12] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 10, pp. 5530–5546, 2015.

[13] C.-H. Lin, R. Wu, W.-K. Ma, C.-Y. Chi, and Y. Wang, "Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 1651–1679, 2018.

[14] M. Abdolali and N. Gillis, "Simplex-structured matrix factorization: Sparsity-based identifiability and provably correct algorithms," *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 2, pp. 593–623, 2021.

[15] O. Vu Thanh, N. Gillis, and F. Lecron, "Bounded simplex-structured matrix factorization: Algorithms, identifiability and applications," *IEEE Transactions on Signal Processing*, vol. 71, pp. 2434–2447, 2023.

[16] X. Fu, K. Huang, B. Yang, W.-K. Ma, and N. D. Sidiropoulos, "Robust volume minimization-based matrix factorization for remote sensing and document clustering," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6254–6268, 2016.

[17] X. Fu, K. Huang, N. D. Sidiropoulos, Q. Shi, and M. Hong, "Anchor-free correlated topic modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 5, pp. 1056–1071, 2018.

[18] G. Tatli and A. T. Erdogan, "Polytopic matrix factorization: Determinant maximization based criterion and identifiability," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5431–5447, 2021.

[19] O. V. Thanh and N. Gillis, "Identifiability of polytopic matrix factorization," in *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE, 2023, pp. 1290–1294.

[20] J. Hu and K. Huang, "Identifiable bounded component analysis via minimum volume enclosing parallelotope," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[21] K. Huang, X. Fu, and N. D. Sidiropoulos, "Anchor-free correlated topic modeling: Identifiability and algorithm," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[22] S. Ibrahim, X. Fu, N. Kargas, and K. Huang, "Crowdsourcing via pairwise co-occurrences: Identifiability and algorithms," in *Advances in Neural Information Processing Systems*, 2019, pp. 7845–7855.

[23] S. Ibrahim and X. Fu, "Recovering joint probability of discrete random variables from pairwise marginals," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4116–4131, 2021.

[24] X. Li, T. Liu, B. Han, G. Niu, and M. Sugiyama, "Provably end-to-end label-noise learning without anchor points," in *International Conference on Machine Learning*, 2021, pp. 6403–6413.

[25] T. Nguyen, S. Ibrahim, and X. Fu, "Deep clustering with incomplete noisy pairwise annotations: A geometric regularization approach," in *International Conference on Machine learning (ICML)*, 2023.

[26] J. Hu and K. Huang, "Global identifiability of ℓ_1-based dictionary learning via matrix volume optimization," in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[27] Y. Sun and K. Huang, "Volume-regularized nonnegative tucker decomposition with identifiability guarantees," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[28] R. M. Freund and J. B. Orlin, "On the complexity of four polyhedral set containment problems," *Mathematical Programming*, vol. 33, no. 2, pp. 139–145, 1985.

[29] G. P. McCormick, "Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems," *Mathematical programming*, vol. 10, no. 1, pp. 147–175, 1976.

[30] J. E. Mitchell, J.-S. Pang, and B. Yu, "Convex quadratic relaxations of nonconvex quadratically constrained quadratic programs," *Optimization Methods and Software*, vol. 29, no. 1, pp. 120–136, 2014.

[31] N. Gillis and R. Luce, "Supplementary material for the paper "Checking the Sufficiently Scattered Condition using a Global Non-Convex Optimization Software"," 2024, see also https://arxiv.org/abs/2402.06019.

[32] V. Leplat, N. Gillis, and M. S. Ang, "Blind audio source separation with minimum-volume beta-divergence NMF," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3400–3410, 2020.

[33] A. M. S. Ang and N. Gillis, "Algorithms and comparisons of nonnegative matrix factorizations with volume regularization for hyperspectral unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 4843–4853, 2019.

[34] F. Zhu, "Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey," *arXiv preprint arXiv:1708.05125*, 2017.

[35] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 67–81, 2014.

Supplementary Material: Checking the SSC Synthetic Data Examples

Nicolas Gillis, *Member, IEEE*, Robert Luce

As explained in Section II of the paper, for a matrix to satisfy the SSC, it requires a certain degree of sparsity. Let us generate matrices $H \in \mathbb{R}_+^{r \times n}$ whose columns are k-sparse, that is, they have k non-zero entries for $1 \leq k \leq r - 1$. The position of the k non-zero entries are picked uniformly at random, while the k non-zero values are picked using the uniform distribution in the probability simplex of dimension k, via the Dirichlet distribution with all parameters equal to one. We will use two values of n: $5r$ and $10r$. Figure 1 displays the number of times, over 20 runs, the SSC was satisfied for k-sparse r-by-n matrices. As expected, we observe that, for n
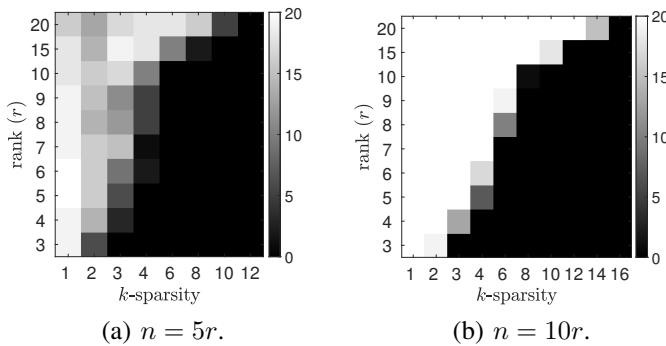

(a) $n = 5r$.   (b) $n = 10r$.

Fig. 1. Number of times, over 20 trials, the SSC for r-by-n matrices whose columns are k-sparse satisfied the SSC.

larger ($n = 10r$), more matrices satisfy the SSC. In fact, for $n = 5r$, many matrices do not satisfy the SSC, even when k is small, because n is not sufficiently large. In summary, as n increases, the phase transition, that is, the largest value of k that allows the SSC to be satisfied becomes larger.

Tables I and II report, for $n = 5r$ and $n = 10r$ respectively, the average computational time, the number of instances that reached the 5-minute time limit, and the number of times the NC-SSC was satisfied but not the SSC. We observe that:

- It is harder for Gurobi to check the SSC close to the phase transition; this explains why the computational cost increases and then decreased as k increases.
- For $r \leq 10$, the time is at most 6.3 seconds.
- For $n = 5r$, the time limit is often reached for $r = 20$ and $4 \leq k \leq 10$. For $n = 10r$, it happens for $r = 15$ and $k = 8, 10$, and for $r = 20$ and $4 \leq k \leq 14$.
- Close to the phase transition, it happens more often that the SSC is not satisfied while the NC-SSC is. This happens for example 7 times out of 20 for $n = 5r$,

NG is with the University of Mons. Email: nicolas.gillis@umons.ac.be.
RL is with Gurobi, email: luce@gurobi.com.

TABLE I
CHECKING THE SSC FOR r-BY-$5r$ MATRICES WHOSE COLUMNS ARE k-SPARSE. THE TABLE REPORTS, OVER 20 TRIALS, THE AVERAGE COMPUTATIONAL TIME IN SECONDS, AND, IF THEY ARE NON-ZEROS, THE NUMBER OF TIMES THE 5-MINUTE TIME LIMIT WAS REACHED, AND THE NUMBER OF TIMES THE NC-SSC WAS SATISFIED BUT NOT THE SSC.

r/k	1	2	3	4	6	8
3	0.0	0.0, 0, 2	/	/	/	/
4	0.0	0.0, 0, 1	0.0, 0, 1	/	/	/
5	0.0	0.0, 0, 1	0.1, 0, 4	0.0	/	/
6	0.0	0.1	0.1, 0, 4	0.0, 0, 3	/	/
7	0.0	0.1	0.3, 0, 1	0.0, 0, 2	0.0	/
8	0.1	0.1, 0, 2	0.3, 0, 4	0.3, 0, 7	0.0	/
9	0.0	0.2, 0, 3	0.4, 0, 2	0.4, 0, 7	0.0	0.0
10	0.1	0.2, 0, 2	0.9, 0, 3	1.1, 0, 4	0.0, 0, 1	0.0
15	0.1	1.1, 0, 3	7.1, 0, 1	280	117, 0, 4	30, 2, 2
20	0.2	3.0, 0, 1	72, 0, 1	271, 18	271, 18, 2	240, 16, 4

TABLE II
CHECKING THE SSC FOR k-SPARSE r-BY-$10r$ MATRICES.

r/k	1	2	3	4	6	8	10
3	0.0	0.1	/	/	/	/	/
4	0.0	0.1	0.1, 0, 1	/	/	/	/
5	0.0	0.2	0.2	0.1, 0, 2	/	/	/
6	0.0	0.3	0.4	0.5, 0, 1	/	/	/
7	0.0	0.3	0.5	0.7	0.0	/	/
8	0.1	0.4	0.7	1.1	1.2	/	/
9	0.0	0.4	1.0	1.7	3.5	0.0	/
10	0.1	0.6	1.6	2.5	6.3	0.7, 0, 1	/
15	0.1	1.7	5.8	26	127	301, 20	271, 18
20	0.2	3.2	31	293, 16	301, 20	300, 20	300, 20

$r = 8, 9$ and $k = 4$. However, this does not happen often when $n = 10r$: only 5 times over all the cases.

One may be a bit disappointed by the fact that Gurobi reaches the time limit on these medium-scale problems. However, in practice, typically r and k are small. For example,

- in hyperspecral imaging, n is the number of pixels in the images (typically larger than 10000), r is the number of materials present in the image (typically smaller than 10), and k is the number of materials present in the pixels (typically smaller than 3).
- in topic modeling, n is the number of documents (typically larger than 1000), r is the number of topics discussed in these documents (typically smaller than 30), and k is the number of topics discussed by the documents present each pixel (a few).

In these real-world scenarios, because r and k are small, it is likely that Gurobi can check the SSC within a reasonable amount of time. This is illustrated in the paper on hyperspectral images; see https://arxiv.org/abs/2402.06019 for more details with the abundance maps for the hyperspectral images.