NONNEGATIVE UNIMODAL MATRIX FACTORIZATION

Andersen Man Shun Ang†, Nicolas Gillis†, Arnaud Vandaele†, Hans De Sterck‡

†Department of Mathematics and Operational Research, Université de Mons, Belgium
‡Department of Applied Mathematics, University of Waterloo, Canada

ABSTRACT

We introduce a new Nonnegative Matrix Factorization (NMF) model called Nonnegative Unimodal Matrix Factorization (NuMF), which adds on top of NMF the unimodal condition on the columns of the basis matrix. NuMF finds applications for example in analytical chemistry. We propose a simple but naive brute-force heuristics strategy based on accelerated projected gradient. It is then improved by using multi-grid for which we prove that the restriction operator preserves the unimodality. We also present two preliminary results regarding the uniqueness of the solution, that is, the identifiability, of NuMF. Empirical results on synthetic and real datasets confirm the effectiveness of the algorithm and illustrate the theoretical results on NuMF.

Index Terms— Nonnegative Matrix Factorization, Unimodality, Multi-grid method, fast gradient method

1. INTRODUCTION

Nonnegative Matrix Factorization (NMF) [1] is the following problem: given a matrix $\mathbf{M} \in \mathbb{R}^{m\times n}_+$ and a factorization rank $r \in \mathbb{N}$, find $\mathbf{W} \in \mathbb{R}^{m\times r}_+$ and $\mathbf{H} \in \mathbb{R}^{r\times n}_+$ such that $\mathbf{WH} \approxeq \mathbf{M}$. In this work, we introduce a new NMF model, namely Nonnegative Unimodal Matrix Factorization (NuMF), which adds on top of NMF a condition that the columns of $\mathbf{W}$ are Nonnegative unimodal (Nu).

Definition 1 ***(Nonnegative unimodality)*** *A vector $\mathbf{x} \in \mathbb{R}^m$ is Nu if there exists an integer $p \in [m] := [1, 2, \dots, m]$ such that*

$$0 \le x_1 \le x_2 \le \cdots \le x_p \text{ and } x_p \ge x_{p+1} \ge \cdots \ge x_m \ge 0. \quad (1)$$

We let $\mathcal{U}^{m,p}_+$ be the set of vectors fulfilling (1), and let $\mathcal{U}^m_+$ be the union of all $\mathcal{U}^{m,p}_+$ for $p \in [m]$. A matrix $\mathbf{X}$ is Nu if all its columns are Nu.

Remark 1 *Note that the value of p is not necessarily unique, and p refers to the location of the change of tonicity, from increasing to decreasing. Nu generalizes the notion of log-concavity [2].*

Nu finds applications in pure mathematics [2], but in this work we focus on the applications in analytical chemistry, for examples the curve resolution problem, flow injection analysis, and gas chromatography–mass spectrometry (GCMS) [3]. See Section 4 for an example of a GCMS dataset.

NG acknowledges the support by the European Research Council (ERC starting grant No 679515), the Fonds de la Recherche Scientifique - FNRS and the Fonds Wetenschappelijk Onderzoek - Vlaanderen (FWO) under EOS project O005318F-RG47. HDS acknowledges support by NSERC of Canada.

Definition 2 ***(NuMF)*** *Given $\mathbf{M} \in \mathbb{R}^{m\times n}_+$ and $r \in \mathbb{N}$, solve*

$$\min_{\mathbf{W},\mathbf{H}} \frac{1}{2}\|\mathbf{M} - \mathbf{WH}\|^2_F \text{ subject to } \begin{array}{l} \mathbf{H} \ge 0, \\ \mathbf{w}_j \in \mathcal{U}^m_+ \text{ for all } j \in [r], \\ \mathbf{w}_j^\top \mathbf{1}_m = 1 \text{ for all } j \in [r], \end{array} \quad (2)$$

where $\mathbf{H} \ge 0$ means $\mathbf{H}$ is element-wise nonnegative, $\mathbf{w}_j$ is the jth column of $\mathbf{W}$ and $\mathbf{1}_m$ is the vector of ones in $\mathbb{R}^m$. The normalization constraint $\mathbf{w}_j^\top \mathbf{1}_m = 1$ is used to handle the scaling ambiguity of the solution $(\mathbf{W}, \mathbf{H})$; see [1].

Contributions. We propose an algorithm to solve NuMF. The algorithm is a combination of brute-force heuristics, accelerated projected gradient (APG), and a multi-grid method (MG). Theoretically, we justify the use of MG as a dimension reduction step in the algorithm by proving that a restriction operator preserves Nu (Theorem 1). We also present two preliminary results regarding the uniqueness of the solution, that is, the identifiability, of NuMF. Finally we present numerical experiments to support the effectiveness of the algorithm, and illustrate the theoretical results regarding identifiability.

2. ALGORITHM

We use block-coordinate descent to solve (2): starting with an initial pair $(\mathbf{W}_0, \mathbf{H}_0)$, we solve the optimization subproblem on $\mathbf{H}$ while fixing $\mathbf{W}$, then we solve the subproblem on $\mathbf{W}$ while fixing $\mathbf{H}$. Such alternating minimization is repeated until the sequence $\{\mathbf{W}_k, \mathbf{H}_k\}_{k\in\mathbb{N}}$ converges. In particular, we employ the HALS algorithm ([4], see also [1]) that updates the columns of $\mathbf{W}$ and rows of $\mathbf{H}$ one-by-one, which runs in $\mathcal{O}(mnr)$ operations. The subproblem on the ith row of $\mathbf{H}$, denoted as $\mathbf{h}^i$, has the following closed-form solution

$$\mathbf{h} = \Big[\mathbf{M}_i^\top \mathbf{w}_i\Big]_+ \Big/ \|\mathbf{w}_i\|^2_2, \quad (3)$$

where $\mathbf{M}_i = \mathbf{M} - \mathbf{WH} + \mathbf{w}_i\mathbf{h}^i$ and $[\,\cdot\,]_+ = \max\{\,\cdot, 0\,\}$. The main difficulty in solving NuMF comes from the subproblem on $\mathbf{W}$, which is a nonconvex problem.

2.1. Subproblem on W is nonconvex

The subproblem on $\mathbf{W}$ is nonconvex because of the Nu set. The set $\mathcal{U}^m_+ = \bigcup_i \mathcal{U}^{m,i}_+$, which is the union of m disjoint convex sets, is nonconvex for $m \ge 3$. For example, let $\mathbf{e}_i$ be the standard basis vector, both $\mathbf{e}_i$ and $\mathbf{e}_j$ are Nu, but the vector $(1-\lambda)\mathbf{e}_j + \lambda\mathbf{e}_j$ with λ in the interval [0, 1] is not Nu if $|j - i| \ge 2$. Note that the union $\mathcal{U}^{m,p}_+ \cup \mathcal{U}^{m,q}_+$ is convex if $|p - q| \le 1$, and it means that $\mathbf{x} \in \mathcal{U}^m_+$ if there exists an integer $p \in [m]$ such that $\mathbf{x} \in \mathcal{U}^{m,p}_+ \cup \mathcal{U}^{m,p+1}_+$, and the Nu membership of $\mathbf{x}$ can be characterized by an inequality

$\mathbf{U}_p\mathbf{x} \geq 0$, where $\mathbf{U}_p$ is a m-by-m matrix built by two first-order difference operators $\mathbf{D}$, as shown in the following equation

$$\underbrace{\begin{cases} 0 & \leq & x_1 \\ x_1 & \leq & x_2 \\ & \vdots & \\ x_{p-1} & \leq & x_p \\ x_{p+1} & \geq & x_{p+2} \\ & \vdots & \\ x_{m-1} & \geq & x_m \\ x_m & \geq & 0 \end{cases}}_{\mathbf{x}\in\mathcal{U}_+^{m,p}\cup\mathcal{U}_+^{m,p+1}} \iff \mathbf{U}_p\mathbf{x} \geq 0, \tag{4a}$$

where

$$\mathbf{U}_p = \left(\begin{matrix} \underbrace{\begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}_{p\times p}}_{\mathbf{D}_{p\times p}} & \mathbf{0}_{p\times(m-p)} \\ \mathbf{0}_{(m-p)\times p} & \mathbf{D}^\top_{(m-p)\times(m-p)} \end{matrix} \right). \tag{4b}$$

Based on this characterization, if the value of p of the vector $\mathbf{w}_i$ is known, the subproblem on $\mathbf{w}_i$, under the HALS framework, is a linearly constrained quadratic programming problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}_i} & \dfrac{\|\mathbf{h}^i\|_2^2}{2}\|\mathbf{w}_i\|_2^2 - \langle \mathbf{M}_i(\mathbf{h}^i)^\top, \mathbf{w}_i\rangle + c \\ \text{subject to} & \mathbf{U}_{p_i}\mathbf{w}_i \geq 0 \text{ and } \mathbf{w}_i^\top \mathbf{1}_m = 1. \end{array} \tag{5}$$

2.2. Accelerated Projected Gradient (APG) for (5)

We solve (5) by accelerated projected gradient (APG) [5]. The feasible set $\{\mathbf{w}_i \mid \mathbf{U}_{p_i}\mathbf{w}_i \geq 0, \mathbf{w}_i^\top\mathbf{1} = 1\}$ in (5) is difficult to project onto, so let us reformulate (5). Note that the square matrix $\mathbf{U}_{p_i}$ in (4) is non-singular. The change of variable $\mathbf{y} = \mathbf{U}_{p_i}\mathbf{w}_i$ gives

$$\underset{\mathbf{y}}{\text{argmin}}\ \frac{1}{2}\langle \mathbf{Q}\mathbf{y}, \mathbf{y}\rangle - \langle \mathbf{p}, \mathbf{y}\rangle \ \text{ s.t. }\ \mathbf{y} \geq 0,\ \mathbf{y}^\top\mathbf{b} = 1, \tag{6}$$

where $\mathbf{Q} = \|\mathbf{h}^i\|_2^2\mathbf{U}_{p_i}^{-\top}\mathbf{U}_{p_i}^{-1}, \mathbf{p} = \mathbf{U}_{p_i}^{-1}\mathbf{M}_i(\mathbf{h}^i)^\top$ and $\mathbf{b} = \mathbf{U}_{p_i}^{-1}\mathbf{1}$. We solve (6) by APG; see Algorithm 1. Then we convert the solution $\mathbf{y}^*$ of (6) to $\mathbf{w}_i^*$ that solves (5) via $\mathbf{w}_i^* = \mathbf{U}_{p_i}^{-1}\mathbf{y}^*$.

Algorithm 1: Accelerated Projected Gradient (APG)

Input : $\mathbf{Q} \in \mathbb{R}^{m\times m}$, $\mathbf{p}$, $\mathbf{b}$
Output: Vector $\mathbf{y}$ that approximately solves (6).
Initialize $\hat{\mathbf{y}}_0 = \mathbf{y}_0 \in \mathbb{R}^m$;
For $k = 1, \ldots$ until some criteria is satisfied **do**
  $\mathbf{y}_k = P\Big(\hat{\mathbf{y}}_{k-1} - \frac{\mathbf{Q}\hat{\mathbf{y}}_{k-1}-\mathbf{p}}{\|\mathbf{Q}\|_2}\Big)$ % Projected gradient step;
  $\hat{\mathbf{y}}_k = \mathbf{y}_k + \frac{k-1}{k+2}(\mathbf{y}_k - \mathbf{y}_{k-1})$ % Extrapolation step;
End for

The key in APG is the projection P. Given a vector $\mathbf{z}$, $P(\mathbf{z})$ is defined as

$$P(\mathbf{z}) = \underset{\mathbf{y}}{\text{argmin}}\ \frac{1}{2}\|\mathbf{y} - \mathbf{z}\|_2^2 \ \text{ s.t. }\ \mathbf{y} \geq 0,\ \mathbf{y}^\top\mathbf{b} = 1. \tag{7}$$

This problem is the projection onto an irregular simplex described by the vector $\mathbf{b} = \mathbf{U}_{p_i}^{-1}\mathbf{1}$. As $\mathbf{U}_{p_i}$ in (4) is built by two first-order difference operators and is thereby block tri-diagonal, its inverse is positive block tri-diagonal and thus $\mathbf{b} > 0$. This implies that (7) satisfies the Slater's condition, that is, the feasible set has a non-empty relative interior, which guarantees strong duality. The solution to (7) can be derived from the partial Lagrangian associated with the equality constraint:

$$\mathbf{y}^* = \min_{\mathbf{y}\geq 0}\max_{\nu} L(\mathbf{y}, \nu) = \frac{1}{2}\|\mathbf{y} - \mathbf{z}\|_2^2 + \nu(\mathbf{y}^\top\mathbf{b} - 1),$$

which has a closed-form solution given by the soft-thresholding $\mathbf{y}^* = [\mathbf{z} - \nu\mathbf{b}]_+$, where the Lagrangian multiplier ν is the root of the following piece-wise linear equation

$$\sum_{i=1}^m \max\Big\{0, z_i - \nu b_i\Big\} b_i - 1 = 0, \tag{8}$$

where $\dfrac{z_i}{b_i}$ ($i \in [m]$) are the break points. Assuming there are $K \leq m$ nonzero break points, Problem (8) can be solved in $\mathcal{O}(m + K\log m)$ operations by sorting the break points. The complexity of the projection step is in between $\mathcal{O}(m)$ and $\mathcal{O}(m\log m)$, depending on K; see also the discussion in [6]. We implemented an efficient and robust MATLAB code inspired by [7] to solve this problem.

2.3. Brute-force algorithm for p_i

If all p_i are known, solving (5) for all $i \in [r]$ gives the solution to $\mathbf{W}$ for NuMF. In general the p_i's are unknown and should be optimized. In this sense, NuMF is a nonconvex problem with r integer variables $p_1, \ldots, p_r$. A first naive strategy is to solve it using brute force: try all the even integers in $[m]$ on p_i for solving (5), pick the one with the smallest objective function value as the solution. This requires $\mathcal{O}(m^2nr)$ operations and hence does not scale linearly with the size of the data. This brute-force strategy is ineffective for large m.

2.4. Multi-grid as the dimension reduction step

We now discuss an idea of using MG to speed up the computation. The reason why MG is used is that it preserves Nu (Theorem 1), which is not the case for other dimension reduction techniques such as PCA or sampling. Algorithm 2 shows the algorithm for solving NuMF with MG. First we use a *restriction* operator $\mathbf{R}$ on the data to form a smaller problem in a coarse grid: in the general N-level MG, the vector $\mathbf{R}_N\mathbf{R}_{N-1}\ldots\mathbf{R}_1\mathbf{w}$ has the row-dimension of $m_N \ll m$. For m_N that is sufficiently small, we can run the brute-force search to estimate p. The cost of brute-force search is now reduced from searching the even integers in $[m]$ to those in $[m_N]$. After we solve the problem on the coarse grid, we interpolate the solution back to the original fine grid by interpolation, which can be computed as the left-multiplication with the matrix $\mathbf{R}^\top$ with a scaling factor. Lastly we solve the problem on the fine grid with the information $\mathbf{p}_0$, and no further brute-force search is needed.

We now discuss the details of the restriction.

Definition 3 *Restriction operator* $\mathbf{R}$ *is defined as* $\mathbf{x} \rightarrowtail \mathbf{R}\mathbf{x}$, *where* $\mathbf{R} \in \mathbb{R}_+^{m_1\times m}$ *with* $m_1 < m$ *has the form of* (9). *The operator is defined column-wise on a matrix, i.e.,* $\mathbf{R}\mathbf{X} := [\mathbf{R}\mathbf{x}_1 \ldots \mathbf{R}\mathbf{x}_n]$.

There are many choices to build $\mathbf{R}$, for simplicity, we use

$$\mathbf{R}(a,b) = \begin{bmatrix} a & b & & & & \\ & b & a & b & & \\ & & \ddots & \ddots & \ddots & \\ & & & b & a & b \\ & & & & b & a \end{bmatrix}, \begin{array}{l} a>0, b>0, \\ a+2b=1. \end{array} \quad (9)$$

We now show that MG preserves Nu. Let us define $\mathcal{N}_+^{m,p} = \mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}$, which is a subset of the Nu vectors. We have the following result.

Theorem 1 *Let $\mathbf{x} \in \mathcal{U}_+^{m,p}$ and $\mathbf{R} \in \mathbb{R}^{m_1 \times m}$ defined in (9). Then $\mathbf{y} = \mathbf{R}\mathbf{x} \in \mathcal{N}_+^{m_1,p_y}$ with $p_y \in \{\lfloor \frac{p}{2}+1 \rfloor, \lfloor \frac{p}{2} \rfloor\}$.*

Proof *First, we assume p is even, without loss of generality, by considering the vector $[0,\mathbf{x}]$ when p is odd which does not change the unimodality and increases p by one.*

Then let us decompose $\mathbf{R}$ as $\mathbf{R} = \mathbf{A} + \mathbf{B} + \mathbf{C}$, where $\mathbf{A}$ only contains the elements a in $\mathbf{R}$, $\mathbf{B}$ only contains the elements b on the right of a in $\mathbf{R}$, and $\mathbf{C}$ only contains the elements b on the left of a in $\mathbf{R}$. Note that $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are sampling operators multiplied by a constant. These sampling operators are either picking the odd or even indices of $\mathbf{x}$, which give Nu vectors since any sub-vector of a Nu vector is Nu. It remains to show that the sum $\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{x}$ belongs to $\mathcal{N}_+^{m,p_y}$ for $p_y \in \{\lfloor \frac{p}{2}+1 \rfloor, \lfloor \frac{p}{2} \rfloor\}$. Since p is even and the matrices $\mathbf{B}$ and $\mathbf{C}$ sample the even entries of $\mathbf{x}$, the vectors $\mathbf{B}\mathbf{x}$ and $\mathbf{C}\mathbf{x}$ belong to $\mathcal{U}_+^{m_1,\frac{p}{2}}$. It is easy to see that $\mathbf{A}\mathbf{x}$ belongs to $\mathcal{U}_+^{m_1,p_y}$ with p_y either $\frac{p}{2}$ or $\frac{p}{2}+1$, which concludes the proof. □

Algorithm 2: MG-BFS for solving NuMF

Input : $\mathbf{M} \in \mathbb{R}_+^{m \times n}$, $r \in \mathbb{N}$, $\mathbf{W}_0, \mathbf{H}_0$
Output: $\mathbf{W}, \mathbf{H}$ that approximately solve (2)

1. Perform restriction: $\mathbf{M}^{[N]} = \mathbf{R}_N \dots \mathbf{R}_1 \mathbf{M}$, $\mathbf{W}_0^{[N]} = \mathbf{R}_N \dots \mathbf{R}_1 \mathbf{W}_0$.;
2. Solve NuMF on the coarse grid: $$[\mathbf{W}^{[N]}, \mathbf{H}', \mathbf{p}^{[N]}] = \text{HALS}(\mathbf{M}^{[N]}, \mathbf{W}_0^{[N]}, \mathbf{H}_0),$$ with $\mathbf{h}^i$ updated using (3), and $\mathbf{w}_i$ updated using APG discussed in section 2.2.;
3. Interpolation: $[\mathbf{W}_0, \mathbf{p}_0] = \text{Interpolate}(\mathbf{W}^{[N]}, \mathbf{p}^{[N]})$.;
4. Solve the problem in the original grid $$[\mathbf{W}, \mathbf{H}] = \text{HALS}(\mathbf{M}, \mathbf{W}_0, \mathbf{H}', \mathbf{p}_0).$$

* Steps 2-4 can be repeated several time.

Lastly, we give a few remarks concerning Algorithm 2. In practice, due to errors introduced by the restriction and interpolation processes, the vector $\mathbf{p}_0$ obtained from $\mathbf{p}^{[N]}$ may not be precisely accurate for solving NuMF in the original dimension. As a safeguard, we still perform a brute-force search for p values of NuMF in the original dimension, but we only search it in a small neighborhood of $\mathbf{p}_0$, depending on the grid size. Say, for $\mathbf{R}$ defined in (9), we search ± 5 around $\mathbf{p}_0$.

The restriction can be computed at once by forming directly the operator $\mathbf{R} = \mathbf{R}_N \dots \mathbf{R}_1 \in \mathbb{R}^{m_N \times m}$ for a total computational cost of $\mathcal{O}(m_N^2 nr) + \mathcal{O}(mnr)$ operations for Algorithm 2. A natural choice for m_N is therefore $m_N = \mathcal{O}(\sqrt{m})$ so that the computational cost of Algorithm 2 would be asymptotically equivalent to HALS.

Remark 2 (Peak finding) *Another possible heuristic is to use peak detection, for example using `findpeaks` in MATLAB, to preselect a small number of candidate values for the p_i's. This strategy is explored in [8] and will be investigated in future work.*

3. IDENTIFIABILITY

Now we present some results on the identifiability of NuMF, i.e., when does solving NuMF give a unique solution. We first define the support of a vector $\mathbf{x} \in \mathcal{U}_+^m$ as $\text{supp}(\mathbf{x}) := \{i \in [m] \mid x_i \neq 0\} = [a,b]$, where the second equality is due to the fact that the support of all Nu vector contains only a single close interval. Then, we define the notion of strictly disjoint between supports as follows.

Definition 4 ***(Strictly disjoint)*** *Given two vectors $\mathbf{x}, \mathbf{y} \in \mathcal{U}_+^m$ with supp$(\mathbf{x}) = [a_x, b_x]$ and supp$(\mathbf{y}) = [a_y, b_y]$. The two vectors are called strictly disjoint if $a_x > b_y + 1$.*

Remark 3 *Other related concepts to strictly disjoint are adjacent and overlap [8, Section 5.3], which we do not discuss in this paper.*

3.1. Preliminary results on identifiability of NuMF

Here we give the identifiability result of NuMF for two special cases.

Theorem 2 *Assumes $\mathbf{M} = \bar{\mathbf{W}}\bar{\mathbf{H}}$. Solving (2) recovers $(\bar{\mathbf{W}}, \bar{\mathbf{H}})$ if*

1. *$\bar{\mathbf{W}}$ is Nu and all the columns have strictly disjoint support.*
2. *$\bar{\mathbf{H}} \in \mathbb{R}_+^{r \times n}$ has $n \geq 1$, $\|\bar{\mathbf{h}}^i\|_\infty > 0$ for $i \in [r]$.*

Proof *Assume there is another solution $(\mathbf{W}^*, \mathbf{H}^*)$ that solves the NuMF. The columns $\bar{\mathbf{w}}_j$ contribute in $\mathbf{M}$ a series of disjoint unimodal components. For the solution $\mathbf{W}^*\mathbf{H}^*$ to fit $\mathbf{M}$, each $\mathbf{w}_i^*$ has to fit each of these disjoint component in $\mathbf{M}$, and hence $\mathbf{W}^*$ recovers $\bar{\mathbf{W}}$ up to permutation. There is no scaling ambiguity here because of the normalization constraints $\mathbf{w}_i^\top \mathbf{1} = 1$. Moreover, $\mathbf{W}^*$ and $\bar{\mathbf{W}}$ have rank r, since their columns have disjoint support, and hence $\mathbf{H}^*$ and $\bar{\mathbf{H}}$ are uniquely determined (namely, using the left inverses of $\mathbf{W}^*$ and $\bar{\mathbf{W}}$), up to permutation.*

The assumptions in Theorem 2 are strong, but are satisfied in many GCMS data. Furthermore, the theorem holds for $r \geq n$ which is uncommon for most NMF models [1].

We now present another result on general identifiability for NuMF with r limited to 2. This result is general in the sense that it includes vectors with overlapping supports, which is not addressed in the previous theorem. First we have the following lemma on demixing two Nu vectors with non-fully overlapping supports:

Lemma 1 *(On demixing two non-fully overlapping Nu vectors) Given two non-zero vectors $\mathbf{x}, \mathbf{y}$ in $\mathcal{U}_+^m$ with supp$(\mathbf{x}) \not\subseteq$ supp$(\mathbf{y})$ and supp$(\mathbf{x}) \not\supseteq$ supp$(\mathbf{y})$. If $\mathbf{x}, \mathbf{y}$ are generated by two non-zero Nu vectors $\mathbf{u}, \mathbf{v}$ as $\mathbf{x} = a\mathbf{u} + b\mathbf{v}$ and $\mathbf{y} = c\mathbf{u} + d\mathbf{v}$ with nonnegative coefficients a, b, c, d, then we have either $\mathbf{u} = \mathbf{x}$, $\mathbf{v} = \mathbf{y}$ or $\mathbf{u} = \mathbf{y}$, $\mathbf{v} = \mathbf{x}$.*

Proof *Since $\mathbf{x}$ an $\mathbf{y}$ have non-overlapping supports, we cannot have $\mathbf{u} = \alpha\mathbf{v}$ for some $\alpha > 0$, hence $\mathbf{u}$ and $\mathbf{v}$ are linearly independent. Let $\mathbf{X} = \mathbf{U}\mathbf{Q}$, where $\mathbf{X} := [\mathbf{x}, \mathbf{y}]$, $\mathbf{U} := [\mathbf{u}, \mathbf{v}]$ and $\mathbf{Q} := \begin{bmatrix} a & c \\ b & d \end{bmatrix} \geq 0$. The conditions that $\mathbf{x}, \mathbf{y}$ are Nu with supp$(\mathbf{x}) \not\subseteq$ supp$(\mathbf{y})$ and supp$(\mathbf{x}) \not\supseteq$ supp$(\mathbf{y})$ imply $\mathbf{x} \neq 0$, $\mathbf{y} \neq 0$, $\mathbf{x} \neq \mathbf{y}$ and*

$$\begin{aligned} \text{supp}(\mathbf{x}) \not\subseteq \text{supp}(\mathbf{y}) &\implies \exists i^* \in [m] \text{ s.t. } x_{i^*} > 0, y_{i^*} = 0, \\ \text{supp}(\mathbf{y}) \not\subseteq \text{supp}(\mathbf{x}) &\implies \exists j^* \in [m] \text{ s.t. } y_{j^*} > 0, x_{j^*} = 0. \end{aligned} \quad (10)$$

Then $\mathbf{x} \neq \mathbf{y}$ *and* $\mathbf{u} \neq \alpha\mathbf{v}$ *imply* $\mathbf{X}, \mathbf{U}, \mathbf{Q}$ *are all rank-2, hence*

$$\mathbf{U} = \mathbf{X}\mathbf{Q}^{-1} = \mathbf{X}\begin{bmatrix} d & -c \\ -b & a \end{bmatrix}\frac{1}{ad - bc}, \quad ad - bc \neq 0. \qquad (11)$$

Put i^*, j^* *from* (10) *into* (11)*, together with the fact that* $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v}$ *are nonnegative give* $\mathbf{Q}^{-1} \geq 0$*. Lastly* $\mathbf{Q} \geq 0$ *and* $\mathbf{Q}^{-1} \geq 0$ *imply* $\mathbf{Q}$ *is the permutation of a diagonal matrix with positive diagonal [9], where here the diagonal matrix is the identity.* □

Now we can present the general identifiability of NuMF for $r = 2$.

Theorem 3 *Assumes* $\mathbf{M} = \bar{\mathbf{W}}\bar{\mathbf{H}}$*. If* $r = 2$*, solving* (2) *recovers* $(\bar{\mathbf{W}}, \bar{\mathbf{H}})$ *if the columns of* $\bar{\mathbf{W}}$ *satisfy the conditions of Lemma 1 and* $\bar{\mathbf{H}} \in \mathbb{R}^{r \times n}_+$ *is full rank.*
Proof *It follows directly from Lemma 1.*

Theorems 2 and 3 address the identifiability of NuMF from two angles: the number of columns in $\mathbf{W}$ and how the supports of $\mathbf{w}_i$ interact. Neither of the theorems is complete. Generalizing these theorems to all possible interactions between supports of $\mathbf{w}_i$ for $r \geq 3$ is a topic of further research.

4. EXPERIMENTS

We now present experiments on NuMF. The code is available from `https://angms.science/`.

Toy example on MG performance. A Nu matrix $\mathbf{W}$ and a nonnegative matrix $\mathbf{H}$ are constructed with $(m, n, r) = (100, 6, 3)$; see Fig. 1. The NuMF problem is solved with 0, 1 and 2 layers of MG. Fig. 1 shows that MG significantly speeds up the convergence: more than 50% run time reduction for 1 layer MG, and more than 75% time reduction for 2 layers. This result shows that our method is far superior in terms of efficiency, as other existing approaches such as those in [3, 10] that have a similar complexity to Algorithm 2 without MG.

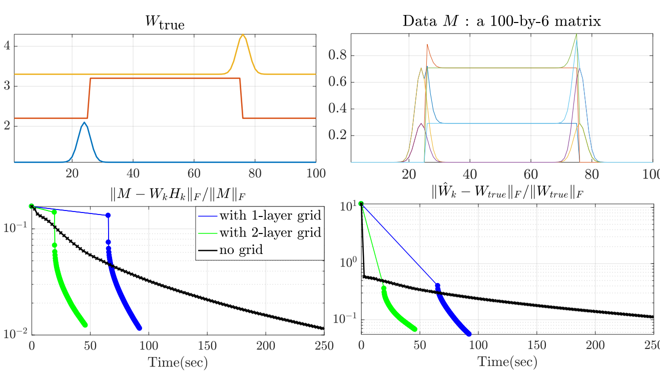

Fig. 1. Experiment on a toy example. **Top**: The ground truth $\mathbf{W}$ matrix and the data $\mathbf{M}$. **Bottom**: The curve plotted against time. All algorithms run 100 iterations and are initialized with SNPA [11]. For algorithms with MG, the computational time taken on the coarse grid are also taken into account, as reflected by the time gap between time 0 and the first dot in the curves.

On GCMS data of Belgian beers. We now demonstrate the regularizing power of the unimodality constraint in the factorization, using a beer dataset [12]. Here $\mathbf{M} \in \mathbb{R}^{518 \times 947}_+$ where each column is a GCMS spectrum. With $r = 7$, three methods: NuMF, NMF [1] and separable NMF (SNMF) [13] are used to decompose the data, and Fig. 2 shows the results. As expected, only NuMF can decompose the data into individual Nu components, while for the other two models, some components are highly mixed with multiple peaks. Note however that the relative error $\|\mathbf{M} - \mathbf{W}\mathbf{H}\|_F / \|\mathbf{M}\|_F$ is similar for the three methods, and around 10%.

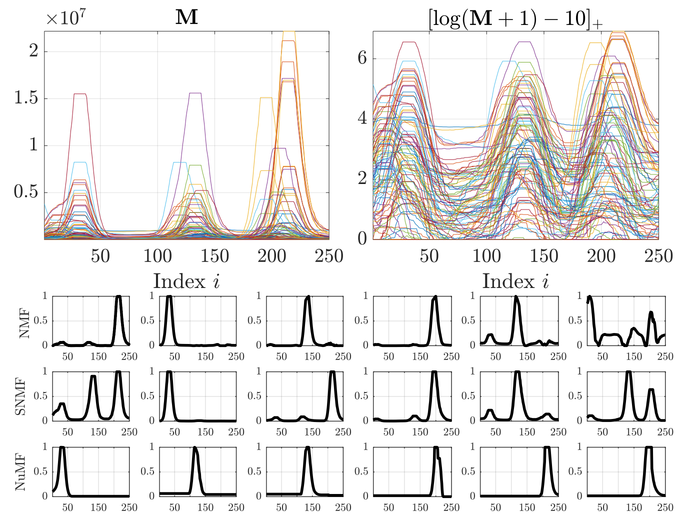

Fig. 2. Experiment on beer data. The bottom plots show the $\mathbf{W}$'s obtained by the three methods.

On data with $r > n$**.** We now consider NuMF on the case $r > n$ (more sources than samples), which is not possible for most other NMF models. Here a GCMS data vector in $\mathbb{R}^{947}_+$ is used. With $r = 8 > n = 1$, we decompose this vector into r unimodal components. NuMF provides a meaningful decomposition; see Fig. 3. Note that the first two peaks in the data satisfy Theorem 2 and hence NuNMF identifies them perfectly. For the other peaks, their supports overlap, and hence the decomposition is not unique. Investigating the identifiability of NuMF on data with overlapping supports is a direction of future research.

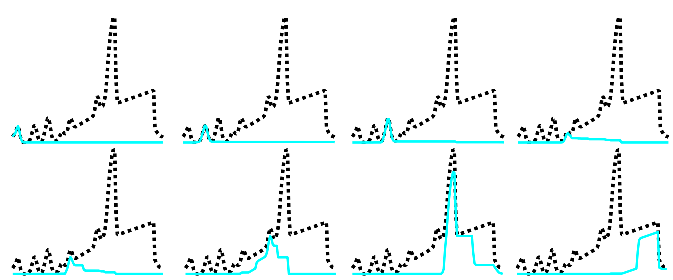

Fig. 3. On data $\mathbf{M}$ (dotted black curve) with $r = 8 > 1 = n$. The cyan curves are the components $\mathbf{w}_i h_i$. Relative error $\|\mathbf{M} - \mathbf{W}\mathbf{H}\|_F / \|\mathbf{M}\|_F = 10^{-8}$.

5. CONCLUSION

We introduced NuMF and proposed to solve it by combining APG and MG. We showed that the restriction operator in MG preserves Nu, and we present two preliminary identifiability results. Numerical experiments support the effectiveness of the proposed algorithm. Future works will be to study the general identifiability of NuMF, and to further improve the algorithm using for example peak finding algorithms; see Remark 2.

6. REFERENCES

[1] Nicolas Gillis, "The why and how of nonnegative matrix factorization," *Regularization, optimization, kernels, and support vector machines*, pp. 257–291, 2014.

[2] Richard Stanley, "Log-concave and unimodal sequences in algebra, combinatorics, and geometry," *Annals of the New York Academy of Sciences*, vol. 576, no. 1, pp. 500–535, 1989.

[3] Rasmus Bro and Nicholaos Sidiropoulos, "Least squares algorithms under unimodality and non-negativity constraints," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 12, no. 4, pp. 223–247, 1998.

[4] Andrzej Cichocki, Rafal Zdunek, and Shun-ichi Amari, "Hierarchical ALS algorithms for nonnegative matrix and 3d tensor factorization," in *International Conference on Independent Component Analysis and Signal Separation*. Springer, 2007, pp. 169–176.

[5] Yurii E Nesterov, "A method for solving the convex programming problem with convergence rate o (1/kˆ 2)," in *Dokl. akad. nauk Sssr*, 1983, vol. 269, pp. 543–547.

[6] Laurent Condat, "Fast projection onto the simplex and the l1 ball," *Mathematical Programming*, vol. 158, no. 1-2, pp. 575–585, 2016.

[7] Laurent Condat, "Matlab code to project onto the simplex or the l1 ball," `https://lcondat.github.io/software.html`, 2015.

[8] Man Shun Ang, *Nonnegative Matrix and Tensor Factorizations: Models, Algorithms and Applications*, Ph.D. thesis, University of Mons, 2020.

[9] Abraham Berman and Robert J Plemmons, *Nonnegative matrices in the mathematical sciences*, SIAM, 1994.

[10] Junting Chen and Urbashi Mitra, "Unimodality-constrained matrix factorization for non-parametric source localization," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2371–2386, 2019.

[11] 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.

[12] Christophe Vanderaa, "Development of a state-of-the-art pipeline for high throughput analysis of gas chromatography - mass spectrometry data," Master thesis, 2018.

[13] Nicolas Gillis and Stephen Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 698–714, 2013.