Nonnegative Unimodal Matrix Factorization

Andersen Ang, Nicolas Gillis, Arnaud Vandaele, Hans De Sterck

Contact: Andersen Ang email: ms3ang@uwaterloo.ca
Department of Combinatorics and Optimization, University of Waterloo, Canada

Abstract and contributions

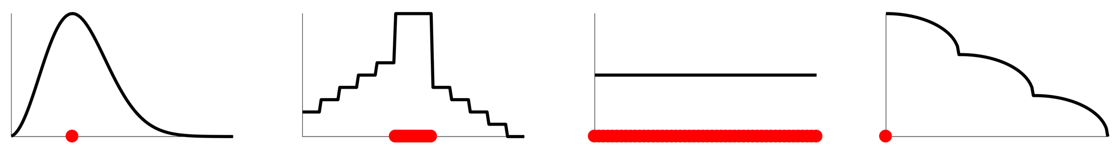
- Introduce a new model: NuMF (= NMF + unimodal).
- Propose BCD algo. to solve NuMF, the algo. = brute-force + accelerated projected gradient + Multi-Grid (MG).
- Prove restriction operator preserves the unimodality in MG.
- Provide preliminary identifiability results of NuMF on special cases.
- Empirical results confirm 1) the effectiveness of the pr posed algo. and 2) the theory on NuMF.

Nonnegative unimodality (Nu)

Def. 1. ***(Nu)*** *$\mathbf{x} \in \mathbb{R}^m$ is Nu if $\exists p \in [m] := [1, 2, \dots, m]$ such that*

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

- Notation: $\mathbf{x} \in \mathcal{U}_+^{m,p}$ and $\mathbf{x} \in \mathcal{U}_+^m$.
- p is the location of change of tonicity.
- p can be nonunique.
- p is known in $\mathcal{U}_+^{m,p}$, this set is convex.
- p is unknown in $\mathcal{U}_+^m$, this set is nonconvex.
- Examples of Nu vectors:



Nu Matrix Factorization

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

$$\begin{array}{lll}
\underset{\mathbf{W}\in\mathbb{R}^{m\times r},\mathbf{H}\in\mathbb{R}^{r\times n}}{\text{minimize}} & \frac{1}{2}\|\mathbf{M}-\mathbf{W}\mathbf{H}\|_F^2 & \textit{fitting term} \\
\textit{subject to} & \mathbf{H} \ge \mathbf{0} & \mathbf{H} \textit{ is nonnegative} \\
 & \mathbf{w}_j \in \mathcal{U}_+^m \, \forall j \in [r] & \textit{columns of } \mathbf{W} \textit{ are Nu} \\
 & \mathbf{w}_j^\top \mathbf{1}_m = 1 \, \forall j \in [r] & \textit{normalization on } \mathbf{W}
\end{array}$$

- We solve NuMF by Block Coordinate Descent (BCD).
  – HALS: we update the columns of $\mathbf{W}$ and rows of $\mathbf{H}$ one-by-one
- Subproblem on $\mathbf{H}$ is NNLS, simple.
- Subproblem on $\mathbf{W}$ is nonconvex, difficult to solve $\implies$ the main difficulty in solving NuMF

Characterizing Nu

- Fact: the union $\mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}$ is cvx
- $\mathbf{x} \in \mathbb{R}^m$ is Nu: $\mathbf{x} \in \mathcal{U}_+^m$ if $\exists p \in [m]$ s.t. $\mathbf{x} \in \mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}$, i.e.,

$$\mathbf{x} \in \mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1} \overset{(1)}{\iff} \left\{ \begin{array}{rcl} 0 & \le & x_1 \\ x_1 & \le & x_2 \\ & \vdots & \\ x_{p-1} & \le & x_p \\ x_{p+1} & \ge & x_{p+2} \\ & \vdots & \\ x_{m-1} & \ge & x_m \\ x_m & \ge & 0 \end{array} \right. \iff \mathbf{U}_p \mathbf{x} \ge 0,$$

$$\mathbf{U}_p = \begin{pmatrix} \underbrace{\begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}_{p \times p}}_{\mathbf{D}_{p\times p}} & \mathbf{0}_{p \times (m-p)} \\ \mathbf{0}_{(m-p)\times p} & \mathbf{D}^\top_{(m-p)\times(m-p)} \end{pmatrix}.$$

That is, we characteized the membership of Nu set by a matrix-vector product by introducing a integer parameter p.

Subproblem on W

- The Nu-characterization means subproblem on $\mathbf{w}_i$ is a Linearly-constrained Quadratic Program with unknown integer p_i

$$\min_{\mathbf{w}_i} \frac{\langle \mathbf{w}_i, \mathbf{B}\mathbf{w}_i \rangle}{2} - \langle \mathbf{c}, \mathbf{w}_i \rangle \;\text{ s.t. }\; \mathbf{U}_{p_i}\mathbf{w}_i \ge 0,\; \mathbf{w}_i^\top \mathbf{1}_m = 1. \qquad (*)$$

for some constant $\mathbf{B}, \mathbf{c}, c$.

- In general, p_i is unknown

$$\min_{\mathbf{w}_i, p_i} \frac{\langle \mathbf{w}_i, \mathbf{B}\mathbf{w}_i \rangle}{2} - \langle \mathbf{c}, \mathbf{w}_i \rangle \;\text{ s.t. }\; \mathbf{w}_i \in \mathcal{U}_+^m \ge 0,\; \mathbf{w}_i^\top \mathbf{1}_m = 1. \qquad (**)$$

- Solve (**) by brute force: try all p_i on (*), pick the best one as p_i^*.
- Speed up 1: solve (*) by accelerated Projected Gradient
  – Change of variable $\mathbf{y} = \mathbf{U}\mathbf{x}$ change (*) to another LCQP

$$\min_{\mathbf{y}} \frac{\langle \mathbf{y}, \mathbf{Q}\mathbf{y} \rangle}{2} - \langle \mathbf{p}, \mathbf{y} \rangle \;\text{ s.t. }\; \mathbf{y} \ge \mathbf{0},\; \mathbf{y}^\top \mathbf{b} = 1. \qquad (*')$$

  – Projection step: $\min_{\mathbf{y}} \frac{1}{2}\|\mathbf{z} - \mathbf{y}\|_2^2$ s.t. $\mathbf{y} \ge \mathbf{0}$, $\mathbf{y}^\top \mathbf{b} = 1$, which can be solved by partial Lagrangian

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

  ∗ Equality due to strong duality: $\mathbf{U}$ nonsingular $\implies \mathbf{b} > \mathbf{0} \implies$ Slater's condition.
  ∗ (#) has closed-form solution as soft-thresholding: $[\mathbf{z} - \nu^* \mathbf{b}]_+$, where μ^* is the optimal Lagrangian multiplier, which is the root of the piece-wise linear equation

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

  which can be solved with complexity $\mathcal{O}(m)$ to $\mathcal{O}(m \log m)$ by sorting the break points $\frac{z_i}{b_i}$.

- Speed up 2: we speed up brute force on p_i by dimension reduction: make the search space for p_i smaller.
  – We use multi-grid: because it preserves Nu.
  – PCA not work here: destroy Nu structure.

Multi-grid: dimension reduction

Def. 3. *Restriction operator* $\mathbf{R}$ *is defined as* $\mathbf{x} \mapsto \mathbf{R}\mathbf{x}$, *where* $\mathbf{R} \in \mathbb{R}_+^{m_1 \times m}$ *with* $m_1 < m$

$$\mathbf{R} = \begin{bmatrix} a & b & & & & \\ & b & a & b & & \\ & & \ddots & \ddots & \ddots & \\ & & & b & a & b \\ & & & & b & a \end{bmatrix}, \quad a > 0,\; b > 0,\; a + 2b = 1.$$

Theorem 1. ***(Restriction preserves Nu)*** *Let $\mathbf{x} \in \mathcal{U}_+^{m,p}$ and $\mathbf{R} \in \mathbb{R}_+^{m_1 \times m}$, then $\mathbf{y} = \mathbf{R}\mathbf{x} \in \mathcal{N}_+^{m_1, p_y}$ with $p_y \in \{\lfloor \frac{p}{2} + 1 \rfloor, \lfloor \frac{p}{2} \rfloor\}$, where $\mathcal{N}_+^{m,p} = \mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}$, which is a subset of the Nu vectors.*

- Proof in 3 sentences, the core ideas:

1. $\mathbf{R}\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{x}$

$$\underbrace{\begin{bmatrix} a & b & & \\ & b & a & b \\ & & b & a \end{bmatrix}}_{\mathbf{R}} = \underbrace{\begin{bmatrix} a & 0 & & \\ & 0 & a & 0 \\ & & 0 & a \end{bmatrix}}_{\mathbf{A}} + \underbrace{\begin{bmatrix} 0 & b & & \\ & 0 & 0 & b \\ & & 0 & 0 \end{bmatrix}}_{\mathbf{B}} + \underbrace{\begin{bmatrix} 0 & 0 & & \\ & b & 0 & 0 \\ & & b & 0 \end{bmatrix}}_{\mathbf{C}}$$

2. $\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{x}, \mathbf{C}\mathbf{x}$ are Nu, because they are all subvector of a Nu vector.
3. The p values of $\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{x}, \mathbf{C}\mathbf{x}$ are at most differ by 1, so by the Nu-characterization, their sum is Nu.

The whole algorithm

Algorithm 1: Proposed algorithm for solving NuMF

Result: $\mathbf{W}, \mathbf{H}$ that solves NuMF
for $k = 1, 2, \dots$ **do**
  for $i = 1, 2, \dots, r$ **do**
    Update $\mathbf{h}^i$ by solving a NNLS (with closed-form update);
    Update $\mathbf{w}^i$ by solving (**) using brute force search on p_i;
    - Restrict the data along column dimension
    - Solve the coarse problem (∗∗):
      – Try all p_i on the coarse problem (∗). Solve it by accelerated projected gradient, pick the best sol. as p_i^*
    - Prolongate p_i to the original dimension
    - Solve the subproblem (*) on the original fine grid with the information of p_i, no brute force is needed
  end
end

Preliminary identifiability results

- Identifiability: when does solving NuMF recover the ground truth.
- Identifiability of NuMF is highly related to the support of the Nu vectors.

Def. 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 $a_x > b_y + 1$.*

Theorem 2. ***(Strictly disjoint Nu vectors)*** *Assumes $\mathbf{M} = \bar{\mathbf{W}}\bar{\mathbf{H}}$. Solving NuMF 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 $n \ge 1$, $\|\bar{\mathbf{h}}^i\|_\infty > 0$ for $i \in [r]$.

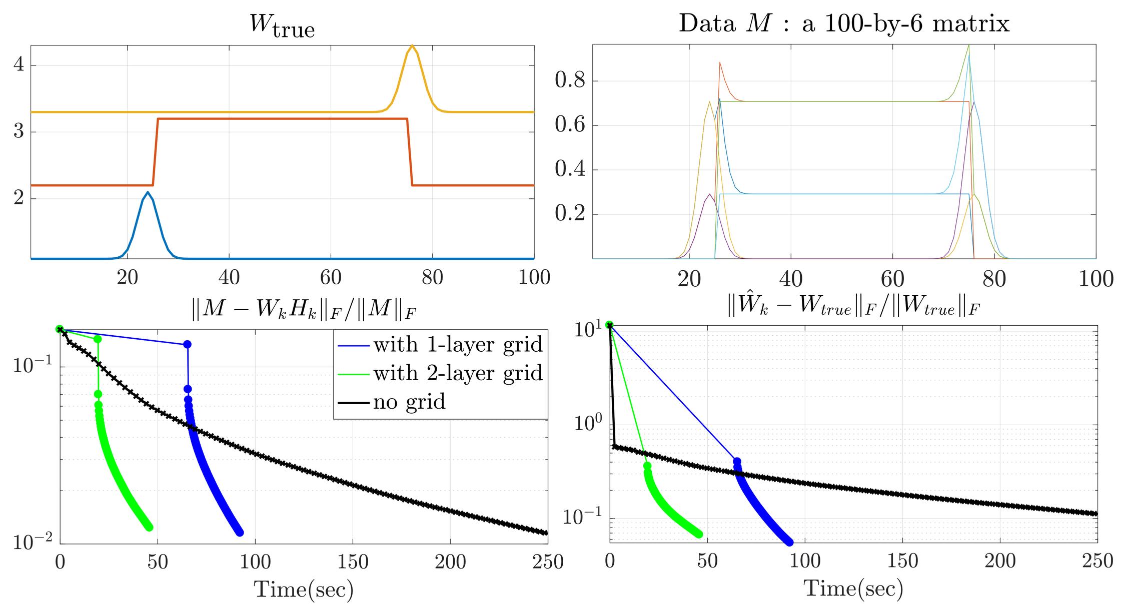
- Strong assumptions, but they are satisfied in chemistry datasets.
- The theorem holds for $r \ge n$ which is uncommon for most NMF models.

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$(\mathbf{x}) \nsubseteq$ supp$(\mathbf{y})$ and supp$(\mathbf{x}) \nsupseteq$ 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 non-negative coefficients a, b, c, d, then we have either $\mathbf{u} = \mathbf{x}$, $\mathbf{v} = \mathbf{y}$ or $\mathbf{u} = \mathbf{y}$, $\mathbf{v} = \mathbf{x}$.*

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

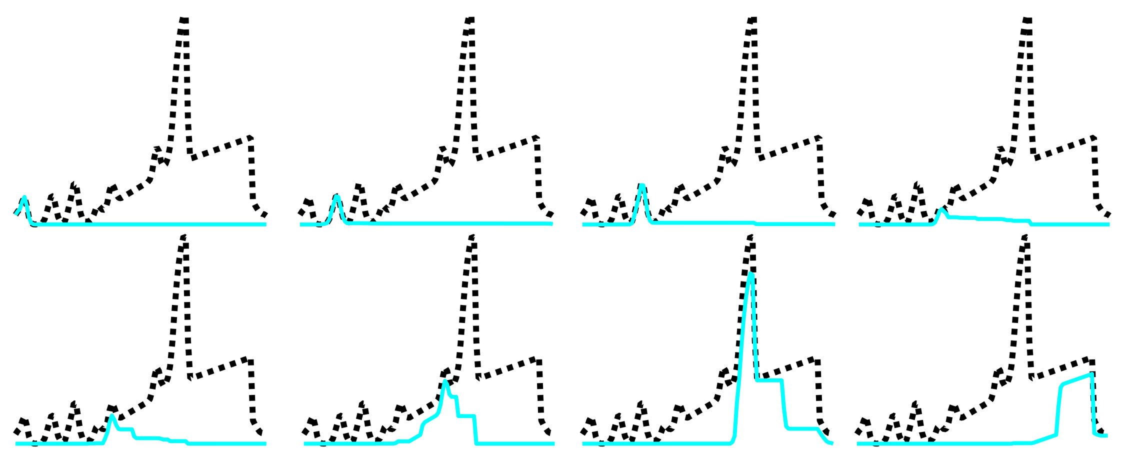
Experiments

Toy example on MG performance



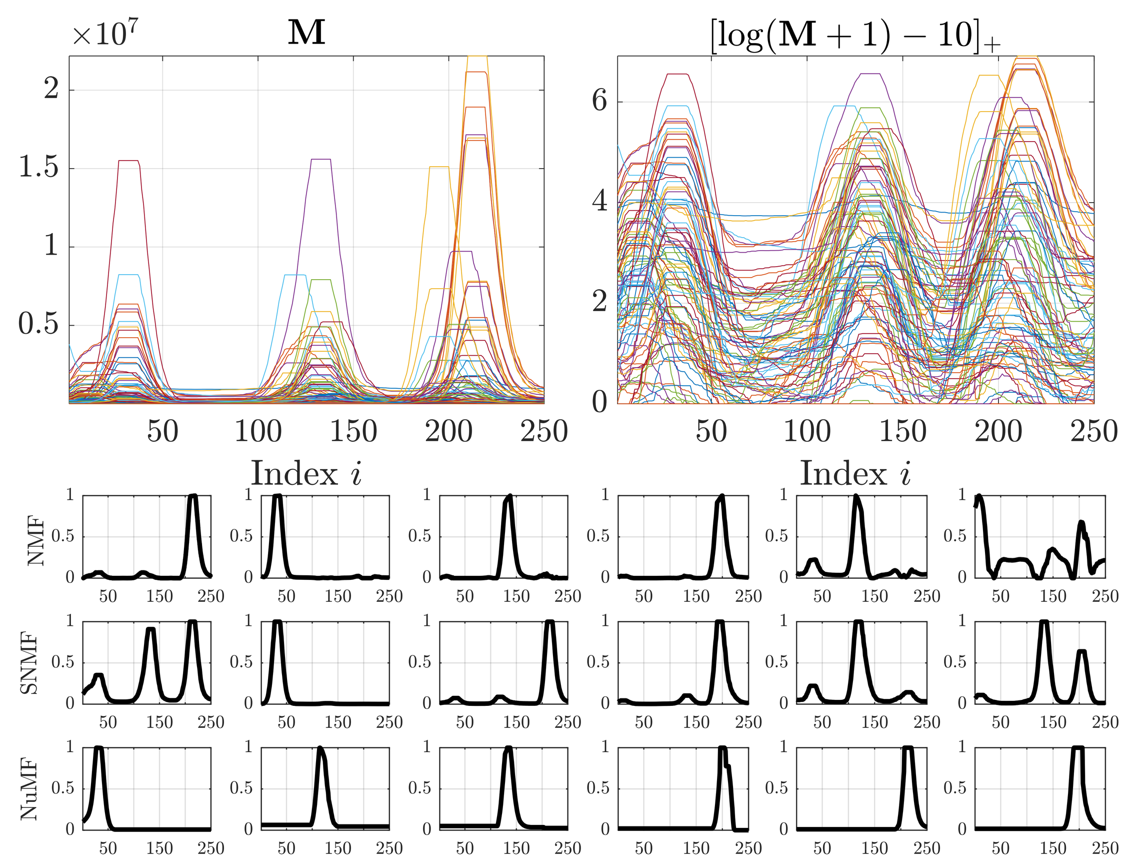
- Special-made synthetic dataset to test MG and theory.
- MG save 75% time for 2-layer. Faster than existing approaches.
- NuMF is ncvx, but we have convergence to global minima because the dataset satisfy the identifiability assumptions.

On data with $r > n$ ($r = 8, n = 1$)



- Factorization rank > data dimension, not possible for NMF.
- First three peaks in the data satisfy Theorem 2 so they are perfectly recovered. Other peaks have overlapped support and their decomposition is non-unique.
- Identifiability of Nu vectors with overlapped support remains open.

On GCMS data of Belgian beers



- Application of Nu in analytical chemistry.
- Other NMF models produce mixed results.

Conclusion and future works

NuMF: characterization, algorithm, theory, experimental verifications. Future works: General identifiability theory of NuMF. NuMF with rows of $\mathbf{H}$ also Nu. Log-concavity of vectors. Applications.

Funding acknowledgements

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) underEOS project O005318F-RG47. HDS acknowledges support by NSERC of Canada.

References

[1] Andersen Man Shun Ang, Nicolas Gillis, Arnaud Vandaele, and Hans De Sterck. Nonnegative unimodal matrix factorization.

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