Nonnegative **u**nimodal **M**atrix **F**actorization

- Andersen Ang (Dept. Combinatorics and Optimization, UWaterloo, Canada)
  Homepage: angms.science
- Coauthors: Nicolas Gillis, Arnaud Vandaele and Hans De Sterck

- Content today
  - ✓ Introduction : What is Nu?
  - ✓ Motivation: Why?
  - X Algorithm: How to solve? See my ICASSP presentation
  - ✓ Theory: What is known about Nu?

Minisymposium "Recent Advances in Matrix and Tensor Factorizations: Algorithms and Theory"
SIAM LA21, 17-May-21

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\| \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{C} \qquad\qquad \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\| + \lambda c(\mathbf{x}).$$

$\ell_2, \ell_1, \ell_0, \ell_p$ norm, sparsity, TV-norm, smoothness, nonnegativity, cone, $\ldots$

This talk: **unimodal**[1] structure.

[1]Not the **unimodular** structure in combinatorial optimization.

Nu

$$\underbrace{0 \le a_1 \le \cdots \le a_{p-1} \le}_{\text{increasing head}} \;\; a_{\color{red}p} \;\; \underbrace{\ge a_{p+1} \ge \cdots \ge a_m \ge 0}_{\text{decreasing tail}}$$

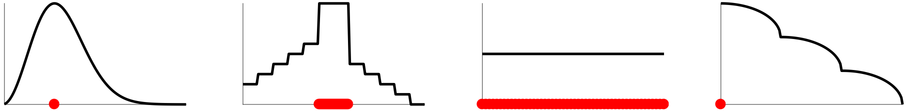
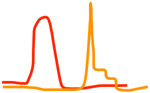


Figure: Nu vectors. Black: the sequence. Red dots: locations of p.

Characterizing the Nu set

$$\mathbf{x} \in \mathbb{R}^m \text{ is Nu} \iff \underbrace{\exists p \in [m] \text{ s.t. } 0 \leq x_1 \leq \cdots \leq x_p \geq \cdots \geq x_m \geq 0}_{\mathbf{x} \in \mathcal{U}_+^{m,p}}.$$

- Notations: $\mathbf{x} \in \mathcal{U}_+^m$ means $\mathbf{x} \in \mathbb{R}^m$ is Nu but p unknown.
- Facts
  - $\mathcal{U}_+^{m,p}$ is cvx
  - $\mathcal{U}_+^m = \bigcup_k \mathcal{U}_+^{m,k}$ is **non**cvx 
  - The set $\mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}$ is cvx.

$$\mathbf{x} \in \mathbb{R}^m \text{ is Nu} \iff \exists p \in [m] \text{ s.t. } \mathbf{x} \in \mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}$$

$$\mathbf{x} \in \mathbb{R}^m \text{ is Nu} \iff \exists p \in [m] \text{ s.t. } \mathbf{x} \in \underbrace{\mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}}_{\text{convex}}$$

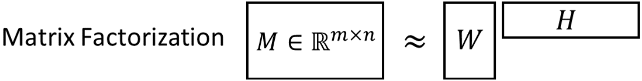
$$\iff \left\{ \begin{array}{ccc} 0 & \leq & x_1 \\ x_1 & \leq & x_2 \\ & \vdots & \\ x_{p-1} & \leq & x_p \\ x_{p+1} & \geq & x_{p+2} \\ & \vdots & \\ x_{m-1} & \geq & x_m \\ x_m & \geq & 0 \end{array} \right.$$

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

* $\mathbf{U}_p$ is full rank.

NuMF

- GIVEN $\mathbf{M} \in \mathbb{R}_{+}^{m \times n}$ and $r \in \mathbb{N}$,
  FIND $\mathbf{W} \in \mathbb{R}^{m \times r}$ and $\mathbf{H} \in \mathbb{R}^{r \times n}$ such that



  by solving

$$\min \ \frac{1}{2}\|\mathbf{M} - \mathbf{W}\mathbf{H}\|_F^2 \ \text{ s.t. } \quad \begin{aligned} &\mathbf{H} \geq \mathbf{0}, \\ &\mathbf{w}_j \in \mathcal{U}_{+}^{m} \text{ for all } j \in [r], \\ &\mathbf{w}_j^\top \mathbf{1}_m = 1 \text{ for all } j \in [r], \end{aligned}$$

- Apply Nu characterization: $\mathbf{w}_j \in \mathcal{U}_{+}^{m} \ \rightarrow \ \mathbf{U}_{p_j}\mathbf{w}_j \geq \mathbf{0}$, where integers $p_1, p_2, \ldots, p_r$ are unknown.

- How to solve: BCD.
  There are a few clever ideas, but not the focus in this talk. For details, see my ICASSP presentation

Identifiability

- When does solving

$$\min \ \tfrac{1}{2}\|\mathbf{M} - \mathbf{W}\mathbf{H}\|_F^2 \ \text{ s.t. } \quad \begin{aligned} &\mathbf{H} \geq \mathbf{0}, \\ &\mathbf{w}_j \in \mathcal{U}_+^m \text{ for all } j \in [r], \\ &\mathbf{w}_j^\top \mathbf{1}_m = 1 \text{ for all } j \in [r], \end{aligned}$$

  gives unique sol.?

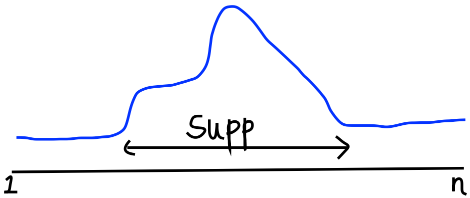
- Essentially uniqueness: permutation and scaling

$$\mathbf{W}_1\mathbf{H}_1 = \mathbf{W}_1\mathbf{Q}\mathbf{Q}^{-1}\mathbf{H}_1 = \mathbf{W}_2\mathbf{H}_2, \quad \mathbf{Q} = \mathbf{\Lambda}\mathbf{\Pi}$$

- Identifiability of NuMF is closely related to the support of Nu vectors.

Support

- Let $\mathbf{x} \in \mathbb{R}^n$ is Nu.

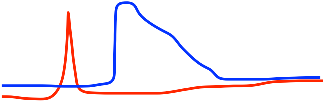


- Support of $\mathbf{x}$: $\operatorname{supp}(\mathbf{x}) = \{\, i \mid x_i \neq 0 \,\}$.

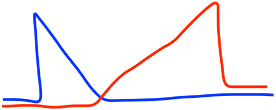
- Trivial facts
  - $\operatorname{supp}(\mathbf{x}) \subseteq [n]$
  - No middle zero: $\operatorname{supp}(\mathbf{x}) = [a, b]$ is a single interval.

Different types of support-interaction between Nu vectors

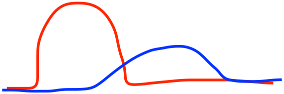
- Strictly disjoint



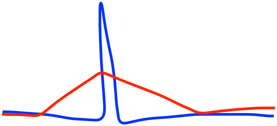
- Adjacent



- Partial overlap



- Complete overlap



When does conic combination preserves Nu?

- To study the identifiability of NuMF, we have to first understand how conic combinations of Nu vectors behave

- A feasibility problem: given two linear independent vectors $\mathbf{x}, \mathbf{y} \in \mathcal{U}_+^m$, find $\alpha \in \mathbb{R}_{++}$ and $\beta \in \mathbb{R}_{++}$ such that the vector $\mathbf{z} := \alpha\mathbf{x} + \beta\mathbf{y}$ is Nu

  OR

  Simplified as

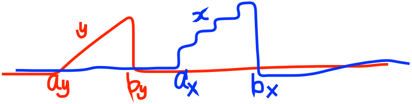
$$\begin{array}{rl} \text{Given} & \text{linear independent vectors } \mathbf{x}, \mathbf{y} \in \mathcal{U}_+^m \\ \text{find} & \text{find } \alpha \in \mathbb{R}_{++} \text{ s.t. } \mathbf{z} := \alpha\mathbf{x} + \mathbf{y} \in \mathcal{U}_+^m \end{array} \tag{5.15}$$

- Trivial result: if $\mathbf{x}, \mathbf{y}$ strictly disjoint, (5.15) has no sol.

- If $\mathbf{x}, \mathbf{y}$ not strictly disjoint, (5.15) may have sol., depends on the structure of $\mathbf{x}, \mathbf{y}$.

Identifiability of the strictly disjoint case (trivial result)

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* $a_x > b_y + 1$.



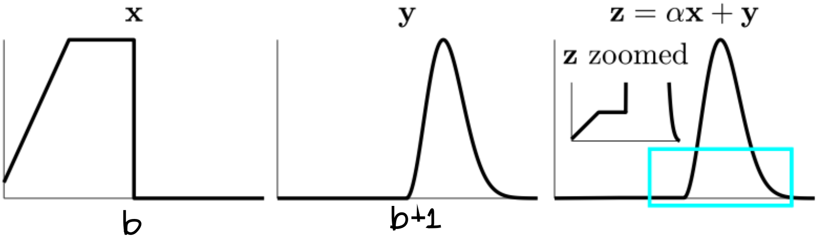
Theorem 2 *Assumes* $\mathbf{M} = \bar{\mathbf{W}}\bar{\mathbf{H}}$. *Solving* (2) *recovers* $(\bar{\mathbf{W}}, \bar{\mathbf{H}})$ *if*

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* $i \in [r]$.

Does conic combination preserves Nu on adjacent Nu vectors?

- WLOG, let $\mathbf{x}, \mathbf{y} \in \mathcal{U}_+^m$ with $\mathrm{supp}(\mathbf{x}) = [a, b]$, $\mathrm{supp}(\mathbf{y}) = [c, d]$, $a \leq b < c \leq d$ and $c = b + 1$.

- Observation: if $\mathbf{x}$ is monotonically increasing within its support, then there exists a sufficiently small $\alpha > 0$ s.t. $\mathbf{z} = \alpha\mathbf{x} + \mathbf{y}$ is Nu, regardless of the toncity of $\mathbf{y}$.



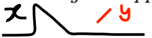
When does conic combination preserves Nu on adjacent Nu vectors?

Lemma 5.3.2. *Let* $\mathbf{x}, \mathbf{y} \in \mathcal{U}_+^m$ *with* $supp(\mathbf{x}) = [a, b]$, $supp(\mathbf{y}) = [c, d]$, $a \le b < c \le d$ *and* $c = b + 1$. *Then if* $\mathbf{x}$ *is monotonically increasing in* $supp(\mathbf{x})$, *then* $\dfrac{y(c)}{x(b)}\mathbf{x} + \mathbf{y}$ *solves problem* (5.15).

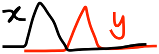
Proof. On the first case, we have $\text{supp}(\mathbf{z}) = \text{supp}(\mathbf{x}) \cup \text{supp}(\mathbf{y}) = [a, d]$ that $z(i) = \alpha x(i)$ if $i \in [a, b]$, and $z(i) = y(i)$ if $i \in [c, d]$. As $\mathbf{x}$ is monotonically increasing for $i \in [1, b]$, the sub-vector $\mathbf{z}(1 : b)$ is monotonically increasing. Regardless of the tonicity of $\mathbf{y}$, the vector $\mathbf{z}$ is Nu as long as $z(b) \le z(c)$, or equivalently $\alpha x(b) \le y(c)$. As $x(b) > 0$, we have $\alpha \le \frac{y(c)}{x(b)}$. The vector $\frac{y(c)}{x(b)}\mathbf{x} + \mathbf{y}$ solves problem (5.15). □

Lemma 5.3.3. *Let* $\mathbf{x}, \mathbf{y} \in \mathcal{U}_+^m$ *with* $supp(\mathbf{x}) = [a, b]$, $supp(\mathbf{y}) = [c, d]$, $a \le b < c \le d$ *and* $c = b + 1$. *Problem* (5.15) *has no solution*

- *if* $\mathbf{x}$ *is monotonically decreasing in* $supp(\mathbf{x})$, *and there exists* $j \in [c, d - 1]$ *such that* $y(j) < y(j + 1)$.



- *if* $\mathbf{y}$ *is monotonically increasing in* $supp(\mathbf{y})$, *and there exists* $i \in [a+1, b]$ *such that* $x(i-1) > x(i)$.
- *if both* $\mathbf{x}$, $\mathbf{y}$ *are not monotonic in their support.*



Proof. The three cases means that there exists $i \in [a+1, b]$ such that $x(i-1) > x(i)$ and $j \in [c, d-1]$ such that $y(j) < y(j+1)$. For all $\alpha > 0$ we have $z(i-1) > z(i)$ and $z(j) < z(j+1)$ for $j > i$. This violates Definition 5.1.2, so $\mathbf{z}$ is not Nu. □

Identifiability: the adjacent case

Theorem 5.3.2. ***(Identifiability of NuMF: adjacent case)*** *Let* $(\bar{\mathbf{W}}, \bar{\mathbf{H}})$ *be the matrices generating the data. Solving* (5.3) *recovers* $(\bar{\mathbf{W}}, \bar{\mathbf{H}})$ *if the followings are satisfied:*

1. $\bar{\mathbf{W}}$ *is Nu and adjacent.*

2. $\bar{\mathbf{H}} \in \mathbb{R}_+^{r \times n}$ *has* $n \geq 1$, $\|\bar{\mathbf{h}}^i\|_\infty > 0$ *for* $i \in [r]$.

3. *Independent sensing: for each index pair* (j_1, j_2), $j_1, j_2 \in [r]$, $j_1 \neq j_2$ *such that the vectors* $\bar{\mathbf{w}}_{j_1}, \bar{\mathbf{w}}_{j_2}$ *satisfy the condition of the vectors* $\mathbf{x}, \mathbf{y}$ *in Lemma 5.3.2, the* j_1, j_2 *rows of* $\mathbf{H}$ *contains a positive diagonal block* $\bar{\mathbf{D}}$.

Proof. By assumption 3 in the theorem, those $\bar{\mathbf{w}}_{j_1}, \bar{\mathbf{w}}_{j_2}$ satisfying Lemma 5.3.2 will have to appear separately in the data. So for $\mathbf{W}^*$ will have to recover them to for fitting the data $\mathbf{M}$. Thereby, we have $\mathbf{H}^*$ recovers $\bar{\mathbf{H}}$. Note that the fitting is subject to permutation and scaling, which does not change anything. $\square$

Remark 5. *We give a toy example to illustrate the importance of assumption 3 in Theorem 5.3.2. Let $m = 6$ and $r = 3$ and we have $\bar{\mathbf{W}}$ as follows with a few additional matrices:*

$$\bar{\mathbf{W}} = \begin{bmatrix} 1 & & \\ 1 & & \\ & 1 & \\ & 1 & \\ & & 1 \\ & & 1 \end{bmatrix}, \quad \mathbf{E}_1 = 1.5 \begin{bmatrix} 1 & \\ 1 & \\ & 1 \\ & 1 \\ 1 & \\ 1 & \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 1 & & \\ 1 & & \\ 1 & & \\ & 1 & \\ & 1 & \\ & & 1 \end{bmatrix},$$

$$\mathbf{C}_1 = \begin{bmatrix} 1.5 \\ 1.5 \\ 1.5 \end{bmatrix}, \qquad \mathbf{C}_2 = \begin{bmatrix} 1.5 & 1.5 \\ 1.5 & 1.5 \\ 1.5 & 1.5 \end{bmatrix}, \quad \mathbf{C}_3 = \begin{bmatrix} 1.5 & \\ & 1.5 \\ 1.5 & \end{bmatrix},$$

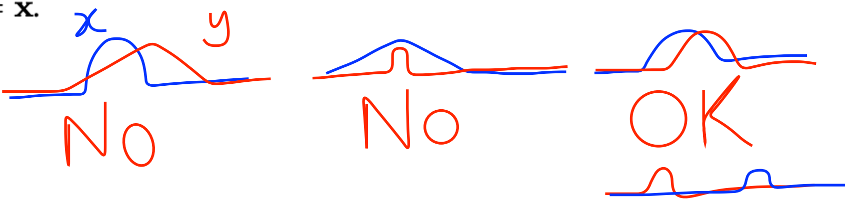
where an empty slot represents zero.

Here the pairs $(\bar{\mathbf{w}}_1, \bar{\mathbf{w}}_2)$, $(\bar{\mathbf{w}}_2, \bar{\mathbf{w}}_3)$ satisfy Lemma 5.3.2. If assumption 3 is not satisfied, say $\bar{\mathbf{H}}$ has the form of $\mathbf{C}_1$ or $\mathbf{C}_2$, then $\mathbf{M}$ contains copies of the vector $[1\,1\,1\,1\,1\,1]^\top$ (multiplied by 1.5). In this case, the problem is not identifiable as $\mathbf{B}_1$ is a feasible solution for $\mathbf{W}^*$. Note that there does not exists a full rank matrix $\mathbf{Q} \in \mathbb{R}^{3\times 3}$ such that $\bar{\mathbf{W}} = \mathbf{B}_1\mathbf{Q}$.

If assumption 3 is satisfied, say $\bar{\mathbf{H}}$ has the form $\mathbf{C}_3$, then the data $\mathbf{M}$ has the form of $\mathbf{E}_1$. In this case, the problem is identifiable as $\mathbf{W}^*$ can only have the form of $\bar{\mathbf{W}}$ (subject to permutation and scaling).

De-mixing two non-fully overlapped Nu vectors

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* $\mathbf{u} = \mathbf{y}$, $\mathbf{v} = \mathbf{x}$.



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.*

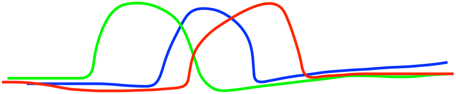
Lemma 1 *(On demixing two non-fully overlapping Nu vectors) Given two non-zero vectors* $\mathbf{x}, \mathbf{y}$ *in* $\mathcal{U}_+^m$ *with supp*$(\mathbf{x}) \not\subseteq$ *supp*$(\mathbf{y})$ *and supp*$(\mathbf{x}) \not\supseteq$ *supp*$(\mathbf{y})$. *If* $\mathbf{x}, \mathbf{y}$ *are generated 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

- $\because$ $\mathbf{x}$ an $\mathbf{y}$ have non-overlapping supports, we cannot have $\mathbf{u} = \alpha\mathbf{v}$ for some $\alpha > 0$, so $\mathbf{u}$ 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 $\mathrm{supp}(\mathbf{x}) \not\subseteq \mathrm{supp}(\mathbf{y})$ and $\mathrm{supp}(\mathbf{x}) \not\supseteq \mathrm{supp}(\mathbf{y})$ imply $\mathbf{x} \neq 0$, $\mathbf{y} \neq 0$, $\mathbf{x} \neq \mathbf{y}$ and
$$\begin{aligned} \mathrm{supp}(\mathbf{x}) \not\subseteq \mathrm{supp}(\mathbf{y}) &\implies \exists i^* \in [m] \text{ s.t. } x_{i^*} > 0, y_{i^*} = 0, \\ \mathrm{supp}(\mathbf{y}) \not\subseteq \mathrm{supp}(\mathbf{x}) &\implies \exists j^* \in [m] \text{ s.t. } y_{j^*} > 0, x_{j^*} = 0. \end{aligned} \tag{1}$$
- 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
$$\mathbf{U} = \mathbf{X}\mathbf{Q}^{-1} = \mathbf{X}\begin{bmatrix} d & -c \\ -b & a \end{bmatrix}\frac{1}{ad - bc}, \quad ad - bc \neq 0. \tag{2}$$
- Put i^*, j^* from (1) into (2), 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, where here the diagonal matrix is the identity. $\square$

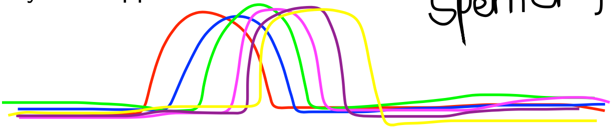
What about other cases: open problems

- De-mixing 3/more non-fully overlapped Nu vectors

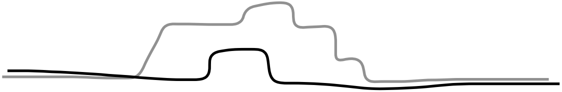


- Partially overlapped Nu vectors

Sperner family?



- Completely overlapped Nu vectors.



Fancy picture: multi-grid saves 75% time with 2-layer

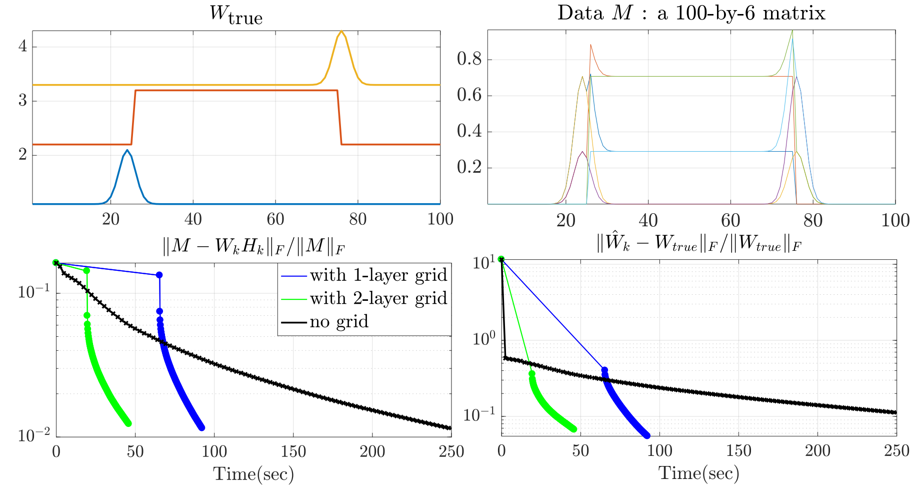
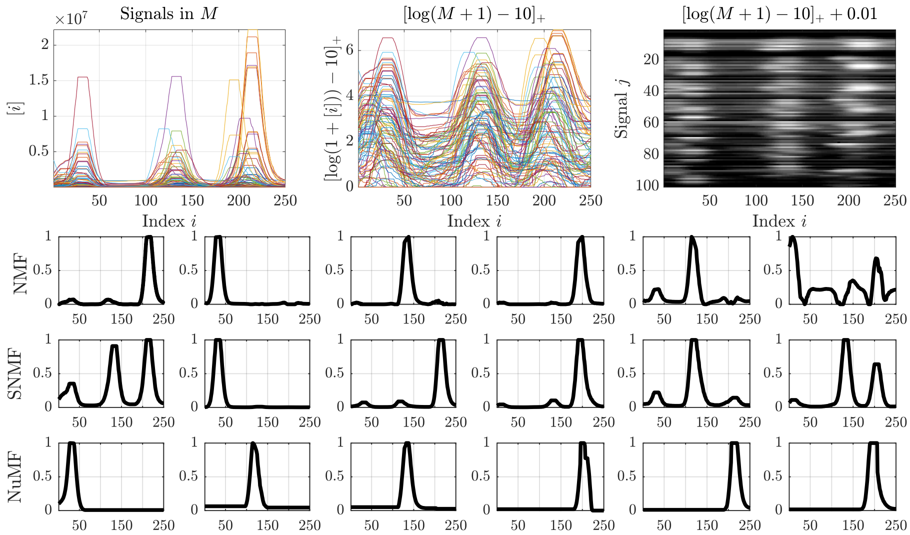
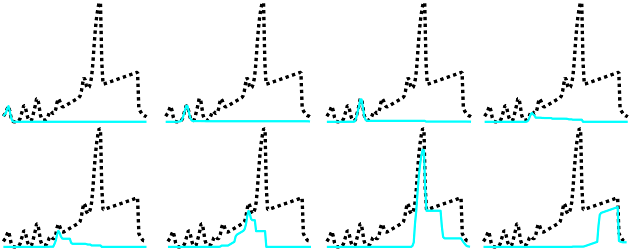


Figure: Experiment on a toy example. All algo. run 100 iterations with same initialization. For algo. with MG, the computational time taken on the coarse grid are also taken into account, as reflected by the time gap between time 0 and the first dot in the curves.

Fancy picture: on Belgian beers



Fancy picture: on $r > n$



- On a data vector in $\mathbb{R}^{947}_{+}$ (black curve) with $r = 8 > 1 = n$.
- Cyan curves are the components $\mathbf{w}_i h_i$.
- Relative error $\|\mathbf{M} - \mathbf{W}\mathbf{H}\|_F / \|\mathbf{M}\|_F = 10^{-8}$.
- The first two peaks in the data satisfy an identifiability Theorem, NuNMF identifies them perfectly.
- For the other peaks: supports overlap, decomposition not unique.

Last page - summary

- NuMF problem: nonconvex and block-nonconvex.
- Identifiability of NuMF (Not discussed in-depth)
- Not discussed: how to actually solve it.
- References
  - **A**, Gillis, Vandaele and De Sterck, “Nonnegative Unimodal Matrix Factorization”, ICASSP21, June 6-11, 2021
  - Chapter 5 of my thesis “Nonnegative Matrix and Tensor Factorizations: Models, Algorithms and Applications”.
- Slide, paper, code at angms.science