NuMF: **N**onnegative **u**nimodal **M**atrix **F**actorization

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium
Homepage: angms.science

SeLMA Seminar
09 Nov 2020

Overview

Structural factorization

- Factorize data it into (low-rank) factors with **structural constraints**.
- Examples:
  - NMF
  - NTF
  - Tucker decomposition
- This talk: **unimodal structure.**

Unimodality

- A unimodal sequence: $-1, 0, 1, 2, 3, 2, -1$
- A nonnegative unimodal (Nu) sequence: $0, 1, 2, 3, 4$
- Def. of unimodality:

$$a_1 \leq a_2 \leq \cdots \leq a_p \geq a_{p+1} \geq \cdots \geq a_n.$$

- Def. of Nu = Def. of u + nonnegativity

$$0 \leq a_1 \leq a_2 \leq \cdots \leq a_p \geq a_{p+1} \geq \dots a_n \geq 0.$$

- A vector $\mathbf{x}$ is Nu:

$$\mathbf{x} \in \mathbb{R}^m \text{ is Nu} \iff \exists p \in [m] \text{ s.t. } 0 \leq x_1 \leq \cdots \leq x_p \geq \dots x_n \geq 0.$$

Some Nu vectors

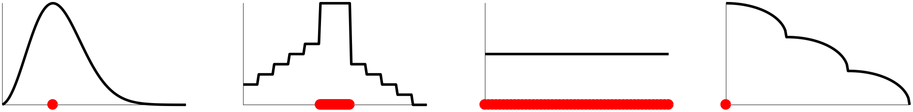


Figure: Four Nu vectors. Black curve: the plot of the sequence. Red dots: the position of p.

Note:

- p can be any integer in $\{1, 2, \ldots, m\}$.
- p can be unique or non-unique

Nonnegative unimodal factorization

- Factorize data into (low-rank) factors with **Nu constraints**.
- Examples
  - Factorize a matrix $\mathbf{M}$ into product $\mathbf{WH}$ such that the columns of $\mathbf{W}$ are Nu $+$ (other constraints).
  - Factorize a tensor $\mathcal{T}$ into product $\mathcal{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}$ such that the columns of $\mathbf{U}$ are Nu $+$ (other constraints).
- Questions
  - Why consider this problem? Motivation
  - How to formulate it and how to solve it? Algorithm
  - What is known about this model? Theory

Motivation: some data are Nu

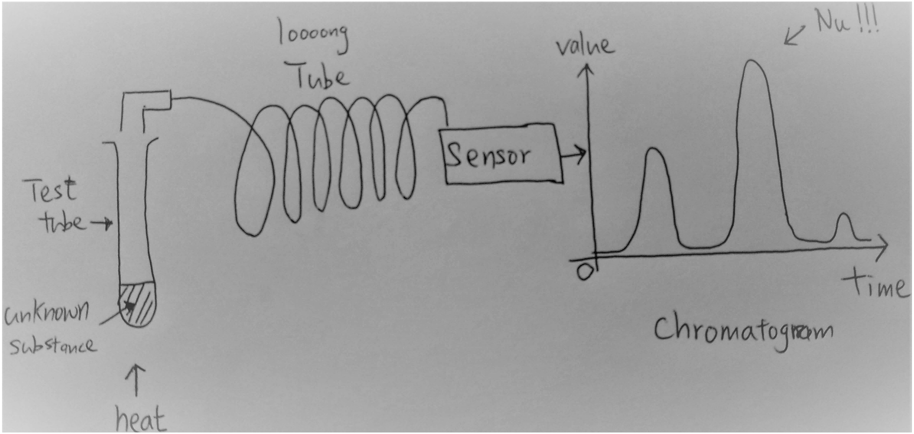


Figure: "Chromatography for monkeys".

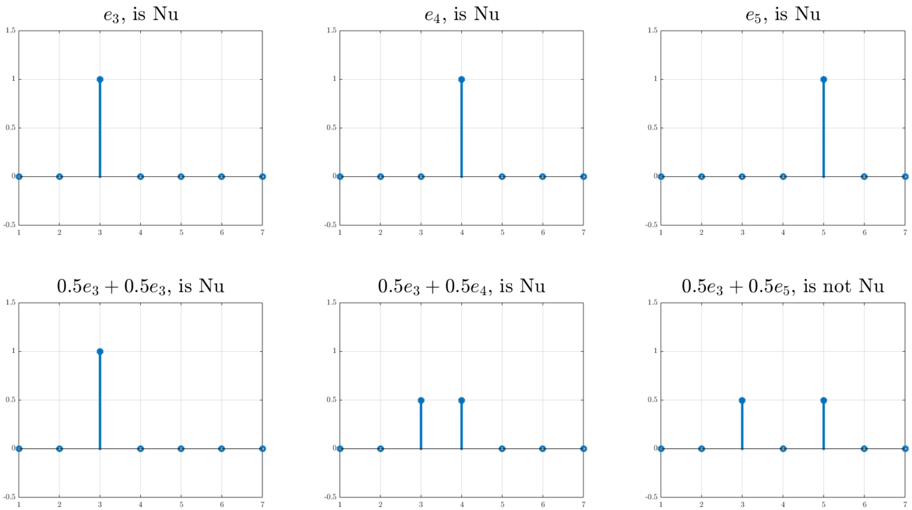
Characterization of Nu set

- A vector $\mathbf{x}$ is Nu:

$$\mathbf{x} \in \mathbb{R}^m \text{ is Nu} \iff \exists p \in [m] \text{ s.t. } 0 \leq x_1 \leq \cdots \leq x_p \geq \dots x_n \geq 0.$$

- Notations
  - $\mathbf{x} \in \mathcal{U}_+^m$ means $\mathbf{x} \in \mathbb{R}^m$ is Nu
  - $\mathbf{x} \in \mathcal{U}_+^{m,p}$ means $\mathbf{x} \in \mathbb{R}^m$ is Nu with known p
- Facts
  - $\mathcal{U}_+^{m,p}$ is a convex set.
  - $\mathcal{U}_+^m = \bigcup_k \mathcal{U}_+^{m,k}$
  - $\mathcal{U}_+^m$ is **not** convex.
    Example: $\mathbf{e}_i$ and $\mathbf{e}_j$ are Nu but $\lambda \mathbf{e}_i + (1-\lambda)\mathbf{e}_j$ is not Nu if $|i-j| \geq 2$.

$\mathbf{e}_i$ and $\mathbf{e}_j$ are Nu but $0.5\mathbf{e}_i + 0.5\mathbf{e}_j$ is not Nu if $|i-j| \geq 2$.



Characterization of Nu set

- The set $\mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}$ is convex

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

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

"Nu membership characterized by a system of monic inequalities".

$$
\underbrace{\left\{\begin{array}{ccc}
0 & \leq & x_1 \\
x_1 & \leq & x_2 \\
 & \vdots & \\
x_{p-1} & \leq & x_p \\
x_{p+1} & \geq & x_{p+2} \\
 & \vdots & \\
x_{m-1} & \geq & x_m \\
x_m & \geq & 0
\end{array}\right.}_{\mathbf{x} \in \mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}} \iff \mathbf{U}_p \mathbf{x} \geq 0
$$

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

NuMF

- Given $\mathbf{M} \in \mathbb{R}_{+}^{m \times n}$ and $r \in \mathbb{N}$, solve

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

- Apply the characterization:

$$\begin{aligned} \text{minimize } \tfrac{1}{2}\|\mathbf{M} - \mathbf{W}\mathbf{H}\|_F^2 \quad \text{subject to} \quad & \mathbf{H} \geq 0, \\ & \mathbf{U}_{p_j}\mathbf{w}_j \geq 0 \text{ for all } j \in [r], \\ & \mathbf{w}_j^\top \mathbf{1}_m = 1 \text{ for all } j \in [r], \end{aligned}$$

where $p_1, p_2, \ldots, p_r$ are unknown!

How to solve?

$$\min_{\substack{\mathbf{W},\mathbf{H}\\ p_1,\ldots,p_j}} \frac{1}{2}\|\mathbf{M}-\mathbf{W}\mathbf{H}\|_F^2 \;\text{ s.t. }\; \mathbf{H}\geq 0,\; \mathbf{U}_{p_j}\mathbf{w}_j \geq 0,\; \mathbf{w}_j^\top \mathbf{1}_m = 1,\; \forall j\in[r].$$

- Subproblem on $\mathbf{H}$ is simple.
- Subproblem on $\mathbf{W}$ involves integer variables and is nonconvex.
- The subproblem on a column of $\mathbf{W}$ (in the HALS framework) is

  $$\min_{\mathbf{w}_i,p_i} \frac{\|\mathbf{h}^i\|_2^2}{2}\|\mathbf{w}_i\|_2^2 - \langle \mathbf{M}_i {\mathbf{h}^i}^\top, \mathbf{w}_i\rangle + c \;\text{ s.t. }\; \mathbf{U}_{p_i}\mathbf{w}_i \geq 0,\; \mathbf{w}_i^\top \mathbf{1} = 1,$$

  which is a linearly-constrained quadratic program.
- Brute-force algorithm: solve the subproblem on all (even) p, and pick the best one as p_i.

Speed up the brute-force algorithm for large m

- Brute-force search on p among the even integers in $\{1, 2, \ldots, m\}$ is slow if m is large.
  $\implies$ if m is sufficiently small, using brute-force is not a problem.

- Speed up 1: solve the subproblem faster for each p using accelerated projected gradient

- Speed up 2: reduce the search space for p
  - By guessing the location of the p
  - By dimension reduction: multi-grid
    - It can be show multi-grid preserves Nu: a theorem with proof in 3 sentences!
    - Other dimension reduction techniques such as PCA or sampling do not work here as they destroy the Nu property

APG: Accelerated Projected Gradient

The subproblem on a column of $\mathbf{W}$ (**with p_i fix**)

$$\min_{\mathbf{w}_i} \frac{\|\mathbf{h}^i\|_2^2}{2}\|\mathbf{w}_i\|_2^2 - \langle \mathbf{M}_i {\mathbf{h}^i}^\top, \mathbf{w}_i \rangle \;\text{ s.t. }\; \mathbf{U}_{p_i}\mathbf{w}_i \geq 0,\; \mathbf{w}_i^\top \mathbf{1} = 1,$$

- The constraint $\left\{\mathbf{U}_{p_i}\mathbf{w}_i \geq 0,\; \mathbf{w}_i^\top \mathbf{1} = 1\right\}$ is hard to project.
- Transform the problem via $\mathbf{y} = \mathbf{U}\mathbf{w}$:

$$\min_{\mathbf{y}} \frac{1}{2}\left\langle \|\mathbf{h}^i\|_2^2 \mathbf{U}_{p_i}^{-\top}\mathbf{y} \,,\, \mathbf{y} \right\rangle - \left\langle \mathbf{U}_{p_i}^{-\top}\mathbf{M}_i {\mathbf{h}^i}^\top \,,\, \mathbf{y} \right\rangle \text{ s.t. } \mathbf{y} \geq 0,\, \mathbf{y}^\top \mathbf{U}_{p_i}^{-1}\mathbf{1} = 1,$$

  or equivalently

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

- Once we get $\mathbf{y}^*$, we get $\mathbf{w}_i^*$ by $\mathbf{y} = \mathbf{U}\mathbf{w}$.

APG on solving $\mathbf{y}$

- The key is the projection onto the irregular simplex: given a point $\mathbf{z}$

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

- Optimal solution given by the partial Lagrangian

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

with closed-form solution given by soft-thresholding, where the Lagrangian multiplier ν^* is the root of a piece-wise linear equation

$$\sum_{i=1}^{m} \max\left\{0, z_i - \nu b_i\right\} b_i - 1 = 0,$$

which takes $\mathcal{O}(m)$ to $\mathcal{O}(m \log m)$ to solve by sorting the break points $\frac{z_i}{b_i}$. After sorting, the magical-one-line-code that no one can read is

```
nu = max((cumsum(z.*b)-1)./(cumsum(b.*b)));
```

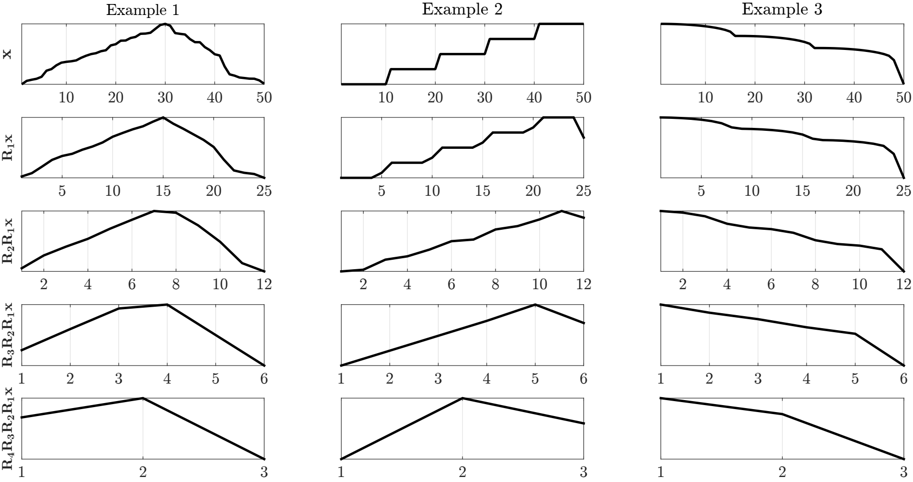
Multi-grid

- Instead of working on $\mathbf{w}$, we work on $\mathbf{R}_N \ldots \mathbf{R}_1 \mathbf{w}$ with much smaller search space of p.

- Restriction $\mathbf{R} \in \mathbb{R}_+^{m_1 \times m}$ change $\mathbf{x} \in \mathbb{R}_+^m$ to $\mathbf{R}\mathbf{x} \in \mathbb{R}_+^{m_1}$ with $m_1 < m$.

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

- **Key fact**: if $\mathbf{x}$ is NU, then $\mathbf{R}\mathbf{x}$ is Nu.


Example 1
Example 2
Example 3
x
R1x
R2R1x
R3R2R1x
R4R3R2R1x

Multi-grid preserves Nu

- Theorem: if $\mathbf{x}$ is Nu, then $\mathbf{Rx}$ is Nu.

- The 3-sentence-proof:
  1. $\mathbf{R}$ can be expressed as a sum
  $$\underbrace{\begin{bmatrix} a & b & & & \\ & b & a & b & \\ & & & b & a \end{bmatrix}}_{\mathbf{R}} = \underbrace{\begin{bmatrix} a & 0 & & & \\ & 0 & a & 0 & \\ & & & 0 & a \end{bmatrix}}_{\mathbf{A}} + \underbrace{\begin{bmatrix} 0 & b & & & \\ & & 0 & b & \\ & & & & 0 \end{bmatrix}}_{\mathbf{B}} + \underbrace{\begin{bmatrix} 0 & & & & \\ & b & 0 & & \\ & & & b & 0 \end{bmatrix}}_{\mathbf{C}}$$
  so $\mathbf{Rx} = \mathbf{Ax} + \mathbf{Bx} + \mathbf{Cx}$.
  2. $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are sampling operators picking the odd or even indices of $\mathbf{x}$, so $\mathbf{Ax}$, $\mathbf{Bx}$ and $\mathbf{Cx}$ are all Nu.
  3. The sum $\mathbf{Ax} + \mathbf{Bx} + \mathbf{Cx}$ is Nu because their p values differ at most 1.

- Theorem (formally): let $\mathbf{x} \in \mathcal{U}_+^{m,p}$ with p is even[1] and $\mathbf{R} \in \mathbb{R}^{m_1 \times m}$ defined as in page 17. Then $\mathbf{y} = \mathbf{Rx} \in \mathcal{N}_+^{m_1,p_y}$ with $\mathcal{N}_+^{m,p} = \mathcal{U}_+^{m,p} \cup \mathcal{U}_+^{m,p+1}$ and $p_y \in \{\lfloor \frac{p}{2} + 1 \rfloor, \lfloor \frac{p}{2} \rfloor\}$.

[1] If p is odd, by considering the vector $[0, \mathbf{x}]$ does not change the unimodality and increases p by one.

The whole algorithm (in words)

Goal: given $\mathbf{M}$, solve NuMF.

Steps:

1. Perform restriction $\mathbf{M}^{[N]} = \mathbf{R}_N \dots \mathbf{R}_1 \mathbf{M}$ and $\mathbf{W}_0^{[N]} = \mathbf{R}_N \dots \mathbf{R}_1 \mathbf{W}_0$
2. Solve NuMF on coarse grid:
$$(\mathbf{W}_*^{[N]}, \mathbf{H}_*, \mathbf{p}_*^{[N]}) \leftarrow \mathrm{NuMF}(\mathbf{M}^{[N]}, \mathbf{W}_0^{[N]}, \mathbf{H}_0)$$
by brute-forcing the p_i and using APG on solving subproblem on $\mathbf{W}$.
3. Interpolate: $(\mathbf{W}_0, \mathbf{p}_0) \leftarrow \mathrm{Interpolate}(\mathbf{W}_*^{[N]}, \mathbf{p}_*^{[N]})$.
4. Solve NuMF on the original fine grid:
$$(\mathbf{W}_*, \mathbf{H}_*, \mathbf{p}_*) \leftarrow \mathrm{NuMF}(\mathbf{M}, \mathbf{W}_0, \mathbf{H}_0, \mathbf{p}_0).$$
without brute-forcing the p_i.

* step 1-4 can be repeated several times: V-cycle, W-cycle, blablabla.

Identifiability: when does solving NuMF give a unique sol?

- Definition: for $\mathbf{x} \in \mathbb{R}^m_+$, $\mathrm{supp}(\mathbf{x}) := \{i \in [m] \mid x_i \neq 0\}$.
- $\forall$ Nu vectors, supp is a closed-interval $[a, b]$ $\because$ no "internal zeros".
- Interactions between two Nu vectors $\mathbf{x}, \mathbf{y}$:
  let $\mathrm{supp}(\mathbf{x}) = [a_x, b_x]$ and $\mathrm{supp}(\mathbf{y}) = [a_y, b_y]$,
  - Strictly disjoint: $a_x > b_y + 1$.
  - Adjacent: $a_x = b_y + 1$.
  - Disjoint = strictly disjoint $\cup$ adjacent
  - Overlap: not disjoint
  - Partial overlap: supports overlap but $\mathrm{supp}(\mathbf{x}) \underset{\supsetneq}{\overset{\subsetneq}{}} \mathrm{supp}(\mathbf{y})$
  - Complete overlap: $\mathrm{supp}(\mathbf{x}) \subseteq \mathrm{supp}(\mathbf{y})$
- Current research status: identifiability for the first two cases.

Identifiability of the strictly disjoint case

Theorem

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* $n \geq 1$, $\|\bar{\mathbf{h}}^i\|_\infty > 0$ *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.*

Note: this theorem holds for $r \geq n$. You can have a $r = 1000$ factorization with $n = 1$.

Demixing two non-fully overlapping Nu vectors

- Given non-zero partially overlap vectors $\mathbf{x}, \mathbf{y}$ in $\mathcal{U}_+^m$, 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 can only have either $\mathbf{u} = \mathbf{x}$, $\mathbf{v} = \mathbf{y}$ or $\mathbf{u} = \mathbf{y}$, $\mathbf{v} = \mathbf{x}$.

- Let $\mathbf{X} = \mathbf{U}\mathbf{Q}$, where $\mathbf{X} := [\mathbf{x}, \mathbf{y}]$, $\mathbf{U} := [\mathbf{u}, \mathbf{v}]$ and $\mathbf{Q} := \begin{bmatrix} a & c \\ b & d \end{bmatrix} \geq 0$.

  What we show: $\mathbf{Q}$ is a permutation matrix.

Sketch of the proof

- $\mathbf{x}, \mathbf{y}$ are Nu with partial-overlap supports imply
  - $\mathbf{u}, \mathbf{v}$ are linearly independent: $\mathbf{U}$ is rank-2
  - $\mathbf{x} \neq 0$, $\mathbf{y} \neq 0$ and $\mathbf{x}, \mathbf{y}$ are linearly independent: $\mathbf{X}$ is rank-2
  - non-zero indices

$$\begin{aligned} \operatorname{supp}(\mathbf{x}) \nsubseteq \operatorname{supp}(\mathbf{y}) &\implies \exists i^* \in [m] \text{ s.t. } x_{i^*} > 0, y_{i^*} = 0, \\ \operatorname{supp}(\mathbf{y}) \nsubseteq \operatorname{supp}(\mathbf{x}) &\implies \exists j^* \in [m] \text{ s.t. } y_{j^*} > 0, x_{j^*} = 0. \end{aligned} \tag{1}$$

- $\mathbf{X}, \mathbf{U}$ are rank 2 imply $\mathbf{Q}$ is 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 $\mathbf{Q}^{-1} \geq 0$.

- $\mathbf{Q} \geq 0$ and $\mathbf{Q}^{-1} \geq 0$ imply $\mathbf{Q}$ is the permutation of a diagonal matrix with positive diagonal, where the diagonal matrix here is $\mathbf{I}$.

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

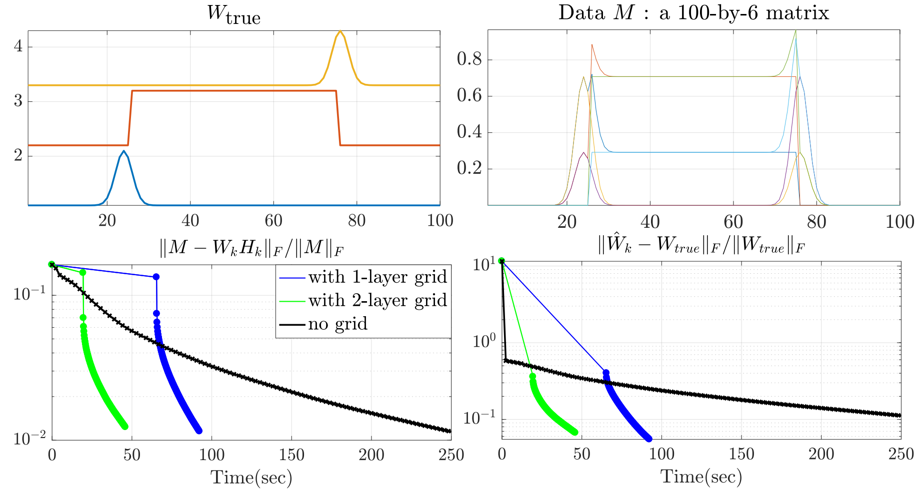
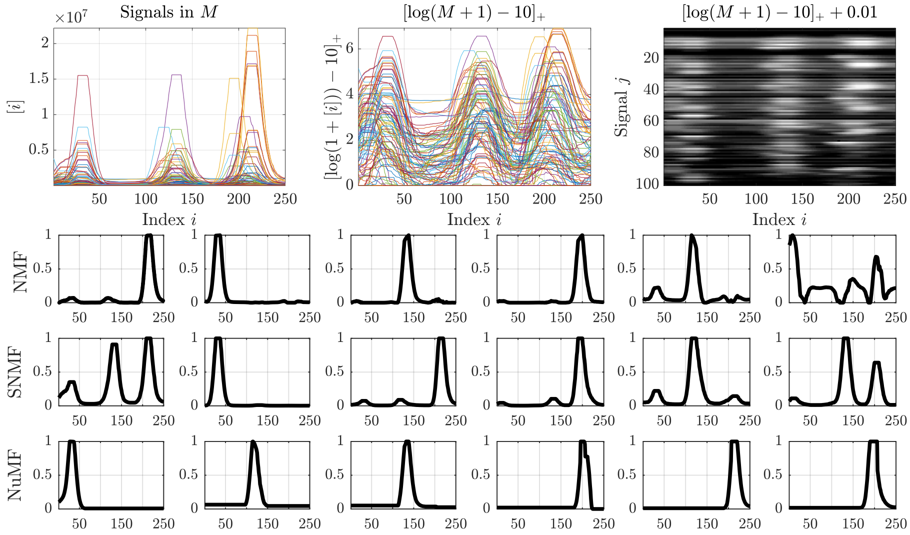
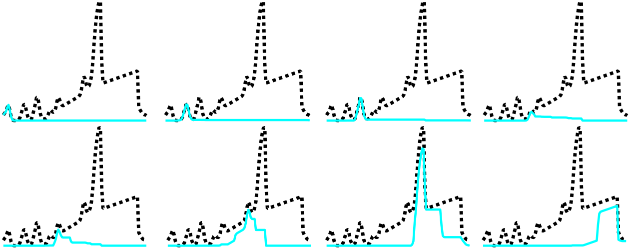


Figure: Experiment on a toy example. All algorithms run 100 iterations with same initialization. For algorithms with MG, the computational time taken on the coarse grid are also taken into account, as reflected by the time gap between time 0 and the first dot in the curves.

Fancy picture: on Belgian beers



Fancy picture: on $r > n$



- On a GCMS data vector in $\mathbb{R}_+^{947}$ (dotted 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 Theorem 1 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.

Last page - summary

- NuMF: motivation, modeling, algorithm, identifiability
- Not discussed
  - The log-concavity
  - Guessing location of p by peak detection
  - Non-uniform adaptive multi-grid
  - Identifiability of NuMF for Nu vectors with adjacent support.
  - The traditional approach used in analytical chemistry other than NuMF
  - Minimum-volume NuMF?
- References
  - Chapter 5 of my thesis "Nonnegative Matrix and Tensor Factorizations: Models, Algorithms and Applications".
  - **A**, Gillis, Vandaele and De Sterck, "Nonnegative Unimodal Matrix Factorization", submitted to ICASSP21.
- Slide, paper, thesis available at https://angms.science/research.html

End of document