Towards Tuning-Free Minimum-Volume Nonnegative Matrix Factorization

Duc Toan Nguyen[1] Eric C. Chi[2]

[1]Department of Mathematics, Texas Christian University, Fort Worth, TX

[2]Department of Statistics, Rice University, Houston, TX

siam 2024 | International Conference on Data Mining

Table of Contents

Table of Contents

Nonnegative Matrix Factorization

Nonnegative matrix

If all entries of a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ are nonnegative, we call matrix $\mathbf{X}$ is nonnegative and denote $\mathbf{X} \geq 0$.

Nonnegative Matrix Factorization

Nonnegative matrix

If all entries of a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ are nonnegative, we call matrix $\mathbf{X}$ is nonnegative and denote $\mathbf{X} \geq 0$.

Nonnegative Matrix Factorization (NMF)

Given a nonnegative matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ and a target rank r that $0 < r \ll m$. We want to find a matrix factorization model

$$\mathbf{X} \approx \mathbf{W}\mathbf{H},$$

where $\mathbf{W} \in \mathbb{R}^{m \times r}$ and $\mathbf{H} \in \mathbb{R}^{r \times n}$ are nonnegative.

NMF Applications

Applications: Hyperspectral unmixing (HU), topic modeling, representation learning.

NMF Applications

Applications: Hyperspectral unmixing (HU), topic modeling, representation learning.

Face Representation Learning:

$$\underbrace{X(:,j)}_{j\text{th facial image}} \approx \sum_{k=1}^{r} \underbrace{W(:,k)}_{\text{facial features}} \underbrace{H(k,j)}_{\substack{\text{importance of features}\\ \text{in } j\text{th image}}} = \underbrace{WH(:,j)}_{\substack{\text{approximation}\\ \text{of } j\text{th image}}}$$

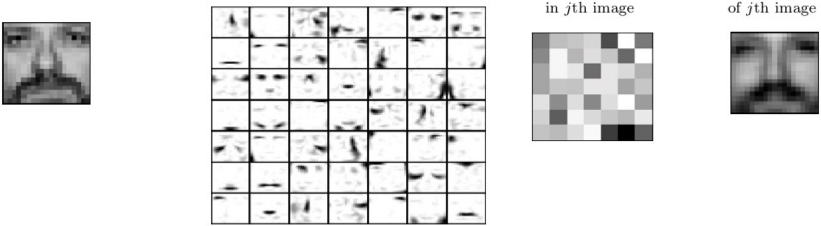


Figure: NMF for Representation Learning [3, 5]. In matrix **X**, each column represents a face image, while in matrix **W**, each column represents a facial feature (learned basis).

Minimum-Volume Rank-Deficient NMF

In [6], Leplat et al. proposed the optimization problem for NMF

MinVol NMF problem

$$\begin{aligned} \min_{\mathbf{W},\mathbf{H}} \quad & \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda \log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}) \\ \text{s.t.} \quad & (\mathbf{W}, \mathbf{H}) \in \mathcal{S} \end{aligned}$$

Notation:

- $\mathcal{S} = \{\boldsymbol{\theta} = (\mathbf{W}, \mathbf{H}) | \mathbf{W}, \mathbf{H} \geq 0$ and $\mathbf{1}^\top \mathbf{H}(:, j) \leq 1, \forall j\}$ (Constraint set)
- $\mathbf{H}(:, j)$ is the jth column of matrix $\mathbf{H}$

Minimum-Volume Rank-Deficient NMF

In [6], Leplat et al. proposed the optimization problem for NMF

MinVol NMF problem

$$\begin{aligned} \min_{\mathbf{W},\mathbf{H}} \quad & \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda \log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}) \\ \text{s.t.} \quad & (\mathbf{W}, \mathbf{H}) \in \mathcal{S} \end{aligned}$$

Notation:

- $\mathcal{S} = \{\boldsymbol{\theta} = (\mathbf{W}, \mathbf{H}) | \mathbf{W}, \mathbf{H} \geq 0 \text{ and } \mathbf{1}^\top\mathbf{H}(:, j) \leq 1, \forall j\}$ (Constraint set)
- $\mathbf{H}(:, j)$ is the jth column of matrix $\mathbf{H}$

Note: The minimum volume ($\text{vol}(\mathbf{W}) = \log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I})$) condition is necessary criteria for a "unique" solution, or the identifiability of NMF (*Craig's belief* [2]).

Motivation

In the paper [6], Leplat et al. briefly choose the tuning parameter λ as

$$\lambda = \tilde{\lambda} \frac{\|\mathbf{X} - \mathbf{W}_0 \mathbf{H}_0\|_F^2}{\log \det(\mathbf{W}_0^T \mathbf{W}_0 + \delta \mathbf{I})}$$

Motivation

In the paper [6], Leplat et al. briefly choose the tuning parameter λ as

$$\lambda = \tilde{\lambda} \frac{\|\mathbf{X} - \mathbf{W}_0 \mathbf{H}_0\|_F^2}{\log\det(\mathbf{W}_0^\top \mathbf{W}_0 + \delta \mathbf{I})}$$

such that

- $(\mathbf{W}_0, \mathbf{H}_0)$ is the initialization for $(\mathbf{W}, \mathbf{H})$

Motivation

In the paper [6], Leplat et al. briefly choose the tuning parameter λ as

$$\lambda = \tilde{\lambda}\frac{\|\mathbf{X} - \mathbf{W}_0\mathbf{H}_0\|_F^2}{\log\det(\mathbf{W}_0^T\mathbf{W}_0 + \delta\mathbf{I})}$$

such that

- $(\mathbf{W}_0, \mathbf{H}_0)$ is the initialization for $(\mathbf{W}, \mathbf{H})$
- $\tilde{\lambda}$ is **between 1 and 10^{-3}** depends on the noise level

Motivation

In the paper [6], Leplat et al. briefly choose the tuning parameter λ as

$$\lambda = \tilde{\lambda} \frac{\|\mathbf{X} - \mathbf{W}_0 \mathbf{H}_0\|_F^2}{\log \det(\mathbf{W}_0^T \mathbf{W}_0 + \delta \mathbf{I})}$$

such that

- $(\mathbf{W}_0, \mathbf{H}_0)$ is the initialization for $(\mathbf{W}, \mathbf{H})$
- $\tilde{\lambda}$ is **between 1 and $\mathbf{10^{-3}}$** depends on the noise level

Question: What is the best $\tilde{\lambda}$ for each noise level?

Table of Contents

Experiment Set-up

Set-up:

- True $\mathbf{W}^\star = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$
- True $\mathbf{H}^\star$: A stochastic 4×500 matrix generated by *Dirichlet distribution*.
- True $\mathbf{X}^\star = \mathbf{W}^\star \mathbf{H}^\star$
- Noise level $\sigma = 10^{-i}$ for $i \in \{1, \cdots, 14\}$
- Simulated matrix $\mathbf{X} = \mathbf{X}^\star + \sigma \mathbf{N}$ ($\mathbf{N}$ is random, $n_{ij} \in [0, 1]$)

Measurements:

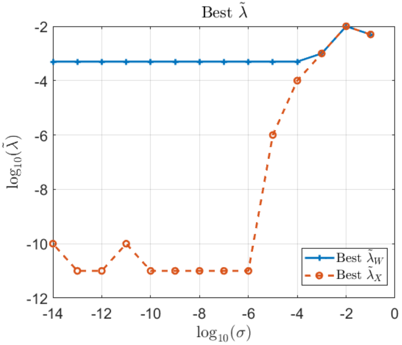
- $\texttt{rel-RMSE}(\mathbf{X}) = \|\mathbf{X}^\star - \hat{\mathbf{X}}\|_F / \|\mathbf{X}^\star\|_F$
- $\texttt{rel-RMSE}(\mathbf{W}) = \|\mathbf{W}^\star - \hat{\mathbf{W}}\|_F / \|\mathbf{W}^\star\|_F$

Experiment process

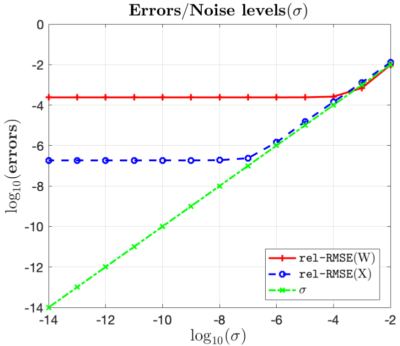
For each noise level σ

1. Run algorithm with multiple $\tilde{\lambda}$ between 1.5 and 10^{-11}
2. Choose the best results:
   - smallest $\texttt{rel-RMSE}(\mathbf{X})$
   - smallest $\texttt{rel-RMSE}(\mathbf{W})$
   - best $\tilde{\lambda}$ for $\texttt{rel-RMSE}(\mathbf{X})$ (Best $\tilde{\lambda}_X$)
   - best $\tilde{\lambda}$ for $\texttt{rel-RMSE}(\mathbf{W})$ (Best $\tilde{\lambda}_W$)

Results



(a) Best $\tilde{\lambda}$



(b) Smallest errors

Figure: Best results corresponding to different noise levels

Observation: Even with the best $\tilde{\lambda}$, the errors cannot get **under 10^{-8}**.

Motivation

Fact: The machine precision (computer numerical noise) $\approx \mathbf{10^{-14}}$

Motivation

Fact: The machine precision (computer numerical noise) $\approx \mathbf{10^{-14}}$

Question: Is there a way to **improve** the **MinVol**?

Table of Contents

Modified problem

Original problem:

$$\begin{aligned} \min_{\mathbf{W},\mathbf{H}} \quad & \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda \log\det(\mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}) \\ \text{s.t.} \quad & (\mathbf{W}, \mathbf{H}) \in \mathcal{S} \end{aligned}$$

Modified problem

Original problem:

$$\min_{\mathbf{W},\mathbf{H}} \quad \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda \log\det(\mathbf{W}^\top \mathbf{W} + \delta \mathbf{I})$$
$$\text{s.t.} \quad (\mathbf{W}, \mathbf{H}) \in \mathcal{S}$$

New optimization problem (inspired by Square-Root Lasso)

$$\begin{aligned} \min_{\mathbf{W},\mathbf{H}} \quad & f(\mathbf{W}, \mathbf{H}) := \sqrt{\|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2} + \lambda \log\det(\mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}) \\ \text{s.t.} \quad & (\mathbf{W}, \mathbf{H}) \in \mathcal{S} \end{aligned} \tag{1}$$

Modified problem

Original problem:

$$\begin{aligned} \min_{\mathbf{W},\mathbf{H}} \quad & \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda \log\det(\mathbf{W}^\mathsf{T}\mathbf{W} + \delta\mathbf{I}) \\ \text{s.t.} \quad & (\mathbf{W}, \mathbf{H}) \in \mathcal{S} \end{aligned}$$

New optimization problem (inspired by Square-Root Lasso)

$$\begin{aligned} \min_{\mathbf{W},\mathbf{H}} \quad & f(\mathbf{W}, \mathbf{H}) := \sqrt{\|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2} + \lambda \log\det(\mathbf{W}^\mathsf{T}\mathbf{W} + \delta\mathbf{I}) \\ \text{s.t.} \quad & (\mathbf{W}, \mathbf{H}) \in \mathcal{S} \end{aligned} \tag{1}$$

Modified problem ("smoothen" square-root term)

$$\begin{aligned} \min_{\mathbf{W},\mathbf{H}} \quad & f_\varepsilon(\mathbf{W}, \mathbf{H}) := \sqrt{\|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \varepsilon} + \lambda \log\det(\mathbf{W}^\mathsf{T}\mathbf{W} + \delta\mathbf{I}) \\ \text{s.t.} \quad & (\mathbf{W}, \mathbf{H}) \in \mathcal{S} \end{aligned} \tag{2}$$

MM algorithm idea

Choose a surrogate function $g(\boldsymbol{\theta}|\boldsymbol{\theta}_k)$ for $f_\varepsilon(\boldsymbol{\theta})$ such that

$$f_\varepsilon(\boldsymbol{\theta}) \leq g(\boldsymbol{\theta}|\boldsymbol{\theta}_k), \text{ for all } \boldsymbol{\theta} \in \mathcal{S}. \tag{3}$$

$$f_\varepsilon(\boldsymbol{\theta}_k) = g(\boldsymbol{\theta}_k|\boldsymbol{\theta}_k) \tag{4}$$

MM algorithm idea

Choose a surrogate function $g(\boldsymbol{\theta}|\boldsymbol{\theta}_k)$ for $f_\varepsilon(\boldsymbol{\theta})$ such that

$$f_\varepsilon(\boldsymbol{\theta}) \leq g(\boldsymbol{\theta}|\boldsymbol{\theta}_k), \text{ for all } \boldsymbol{\theta} \in \mathcal{S}. \tag{3}$$

$$f_\varepsilon(\boldsymbol{\theta}_k) = g(\boldsymbol{\theta}_k|\boldsymbol{\theta}_k) \tag{4}$$

Ideal algorithm map ($\boldsymbol{\theta}_k \to \boldsymbol{\theta}_{k+1}$):

$$(\mathbf{W}_{k+1}, \mathbf{H}_{k+1}) = \underset{(\mathbf{W},\mathbf{H})\in\mathcal{S}}{\arg\min}\, g((\mathbf{W}, \mathbf{H})|(\mathbf{W}_k, \mathbf{H}_k)) \tag{5}$$

MM algorithm idea

Choose a surrogate function $g(\boldsymbol{\theta}|\boldsymbol{\theta}_k)$ for $f_\varepsilon(\boldsymbol{\theta})$ such that

$$f_\varepsilon(\boldsymbol{\theta}) \leq g(\boldsymbol{\theta}|\boldsymbol{\theta}_k), \text{ for all } \boldsymbol{\theta} \in \mathcal{S}. \tag{3}$$

$$f_\varepsilon(\boldsymbol{\theta}_k) = g(\boldsymbol{\theta}_k|\boldsymbol{\theta}_k) \tag{4}$$

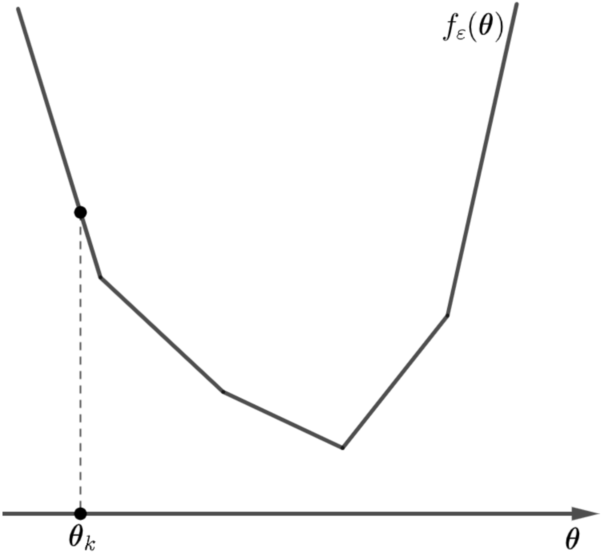
Ideal algorithm map $(\boldsymbol{\theta}_k \rightarrow \boldsymbol{\theta}_{k+1})$:

$$(\mathbf{W}_{k+1}, \mathbf{H}_{k+1}) = \operatorname*{arg\,min}_{(\mathbf{W},\mathbf{H})\in\mathcal{S}} g((\mathbf{W}, \mathbf{H})|(\mathbf{W}_k, \mathbf{H}_k)) \tag{5}$$

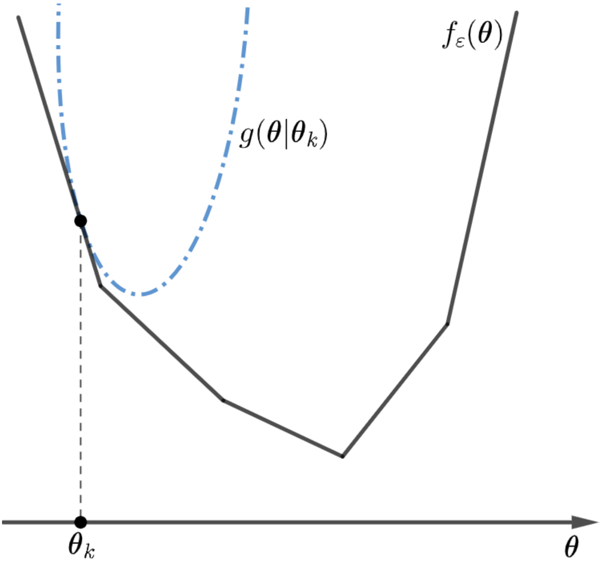
From (3), (4), and (5),

$$f_\varepsilon(\boldsymbol{\theta}_{k+1}) \underset{(3)}{\leq} g(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k) \underset{(5)}{\leq} g(\boldsymbol{\theta}_k|\boldsymbol{\theta}_k) \underset{(4)}{=} f_\varepsilon(\boldsymbol{\theta}_k).$$

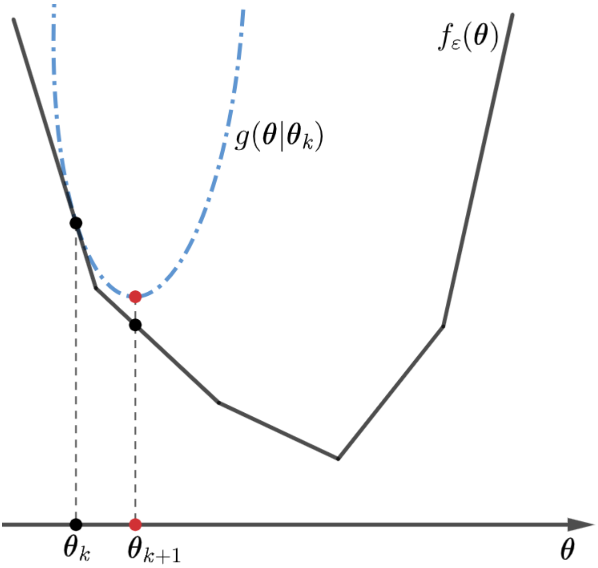
MM idea



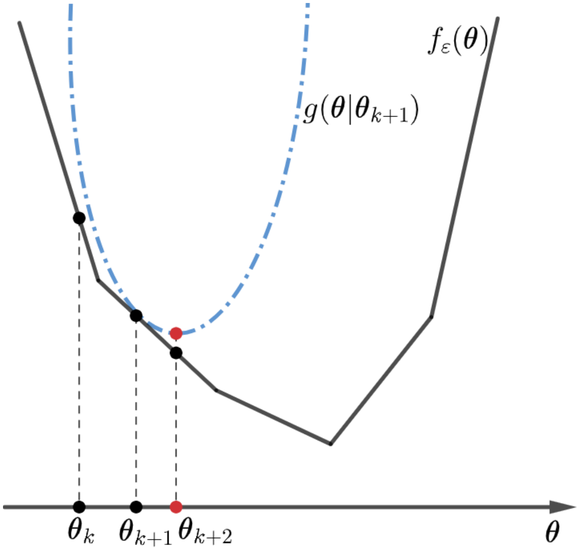
MM idea



MM idea

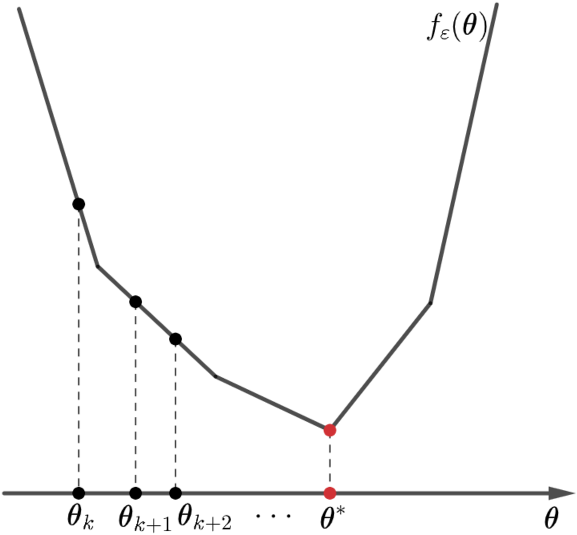


MM idea



$f_\varepsilon(\theta)$
$g(\theta|\theta_{k+1})$
θ_k
θ_{k+1}
θ_{k+2}
θ

MM idea



$f_\varepsilon(\theta)$
θ_k
θ_{k+1}
θ_{k+2}
$\cdots$
θ^*
θ

Surrogate function

Surrogate function: (Use first-order Taylor approximation)

$$g((\mathbf{W},\mathbf{H})|(\mathbf{W}_k,\mathbf{H}_k)) = \sqrt{r_k} + \frac{1}{2\sqrt{r_k}}(\|\mathbf{X}-\mathbf{W}\mathbf{H}\|_{\mathrm{F}}^2 + \varepsilon - r_k)$$
$$+ \lambda(\log\det(\mathbf{Q}_k) + \mathrm{Tr}(\mathbf{Q}_k^{-1}((\mathbf{W}^{\mathsf{T}}\mathbf{W}+\delta\mathbf{I}) - \mathbf{Q}_k)))$$

for

- $r_k = ||\mathbf{X} - \mathbf{W}_k\mathbf{H}_k||_{\mathrm{F}}^2 + \varepsilon$
- $\mathbf{Q}_k = \mathbf{W}_k^{\mathsf{T}}\mathbf{W}_k + \delta\mathbf{I}$.

Surrogate function

Surrogate function: (Use first-order Taylor approximation)

$$g((\mathbf{W},\mathbf{H})|(\mathbf{W}_k,\mathbf{H}_k)) = \sqrt{r_k} + \frac{1}{2\sqrt{r_k}}(\|\mathbf{X}-\mathbf{W}\mathbf{H}\|_{\mathrm{F}}^2 + \varepsilon - r_k) + \lambda(\log\det(\mathbf{Q}_k) + \mathrm{Tr}(\mathbf{Q}_k^{-1}((\mathbf{W}^{\mathsf{T}}\mathbf{W}+\delta\mathbf{I}) - \mathbf{Q}_k)))$$

for

- $r_k = ||\mathbf{X}-\mathbf{W}_k\mathbf{H}_k||_{\mathrm{F}}^2 + \varepsilon$
- $\mathbf{Q}_k = \mathbf{W}_k^{\mathsf{T}}\mathbf{W}_k + \delta\mathbf{I}$.

Simplified algorithm map

$$\begin{aligned}(\mathbf{W}_{k+1},\mathbf{H}_{k+1}) &= \underset{(\mathbf{W},\mathbf{H})\in\mathcal{S}}{\arg\min}\, g((\mathbf{W},\mathbf{H})|(\mathbf{W}_k,\mathbf{H}_k)) \\ &= \underset{(\mathbf{W},\mathbf{H})\in\mathcal{S}}{\arg\min}\, ||\mathbf{X}-\mathbf{W}\mathbf{H}||_{\mathrm{F}}^2 + \lambda_k \log\det(\mathbf{W}^{\mathsf{T}}\mathbf{W}+\delta\mathbf{I})(*)\end{aligned}$$

for $\lambda_k = 2\sqrt{r_k}\lambda$

Note

Simplified algorithm map:

$$(\mathbf{W}_{k+1}, \mathbf{H}_{k+1}) = \underset{(\mathbf{W},\mathbf{H})\in\mathcal{S}}{\arg\min} ||\mathbf{X} - \mathbf{WH}||_F^2 + \lambda_k \log\det(\mathbf{W}^\mathsf{T}\mathbf{W} + \delta\mathbf{I})(*)$$

$\rightarrow$ Use the **MinVol** algorithm (Leplat et al. [6]) to solve $(*)$.

Note

Simplified algorithm map:

$$(\mathbf{W}_{k+1}, \mathbf{H}_{k+1}) = \underset{(\mathbf{W},\mathbf{H})\in\mathcal{S}}{\arg\min} ||\mathbf{X} - \mathbf{W}\mathbf{H}||_F^2 + \lambda_k \log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}) (*)$$

$\rightarrow$ Use the **MinVol** algorithm (Leplat et al. [6]) to solve $(*)$.

Also,

- $r_k := ||\mathbf{X} - \mathbf{W}_k\mathbf{H}_k||_F^2 + \varepsilon$ approximates noise level at iteration k [1]
- The update $\lambda_k = 2\sqrt{r_k}\lambda$ follows the noise level.

Algorithm Square-Root Min-Vol NMF

Input: $\mathbf{X} \in \mathbb{R}_{+}^{m \times n}$, target rank r, λ, δ, ε.
Output: $\mathbf{W} \in \mathbb{R}_{+}^{m \times r}, \mathbf{H} \in \mathbb{R}_{+}^{r \times n}$ in $\mathcal{S}$

1: $(\mathbf{W}_1, \mathbf{H}_1) = \textbf{SNPA}(\mathbf{X}, r)$
2: $\lambda_1 = \lambda$
3: **for** $k = 1, \ldots$ **do**
4: $\quad (\mathbf{W}_{k+1}, \mathbf{H}_{k+1}) = \textbf{MinVol}(\mathbf{X}, r, [\mathbf{W}_k, \mathbf{H}_k, \lambda_k, \delta])$
5: $\quad \lambda_{k+1} \leftarrow (2\sqrt{\|\mathbf{X} - \mathbf{W}_{k+1}\mathbf{H}_{k+1}\|_F^2 + \varepsilon})\lambda$
6: **end for**

MinVol: minimum-volume algorithm (Leplat et al.) [6]
SNPA: successive nonnegative projection algorithm [4]

Experiment Set-up

Set-up: (Similar to last experiment)

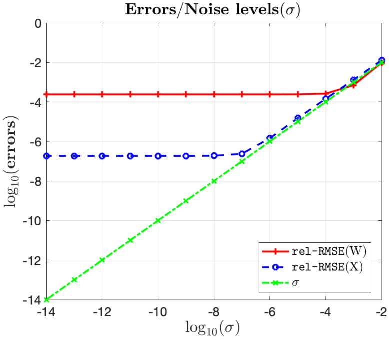
- True $\mathbf{W}^\star = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$
- True $\mathbf{H}^\star$: A stochastic 4×500 matrix generated by *Dirichlet distribution*.
- True $\mathbf{X}^\star = \mathbf{W}^\star \mathbf{H}^\star$
- Noise level $\sigma = 10^{-i}$ for $i \in \{1, \cdots, 14\}$
- Simulated matrix $\mathbf{X} = \mathbf{X}^\star + \sigma \mathbf{N}$ ($\mathbf{N}$ is random, $n_{ij} \in [0, 1]$)
- $\varepsilon = 10^{-8}$

Measurements:

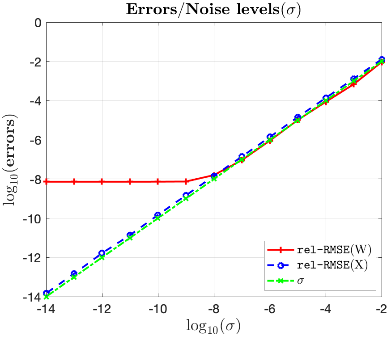
- $\texttt{rel-RMSE}(\mathbf{X}) = \|\mathbf{X}^\star - \hat{\mathbf{X}}\|_F / \|\mathbf{X}^\star\|_F$
- $\texttt{rel-RMSE}(\mathbf{W}) = \|\mathbf{W}^\star - \hat{\mathbf{W}}\|_F / \|\mathbf{W}^\star\|_F$

Result 1

- Use the same list of initial λ's as last experiment



(a) Errors by **MinVol**

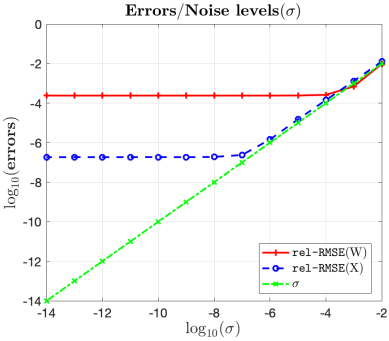


(b) Errors by **Square-Root Min-Vol**

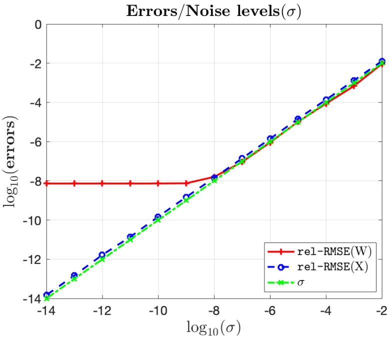
Figure: Smallest errors corresponding to different noise levels by two algorithms

Result 1

- Use the same list of initial λ's as last experiment



(a) Errors by **MinVol**



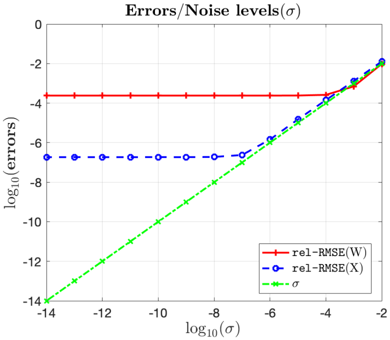
(b) Errors by **Square-Root Min-Vol**

Figure: Smallest errors corresponding to different noise levels by two algorithms

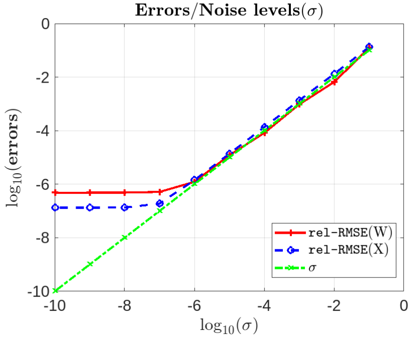
Observation: For **Square-Root Min-Vol**, both errors are much smaller, and $\texttt{rel-RMSE}(\mathbf{X}) = \mathbf{10^{-14}}$!

Result 2

- For **Square-Root Min-Vol**, if use only $\lambda = 0.5$ for all noise levels



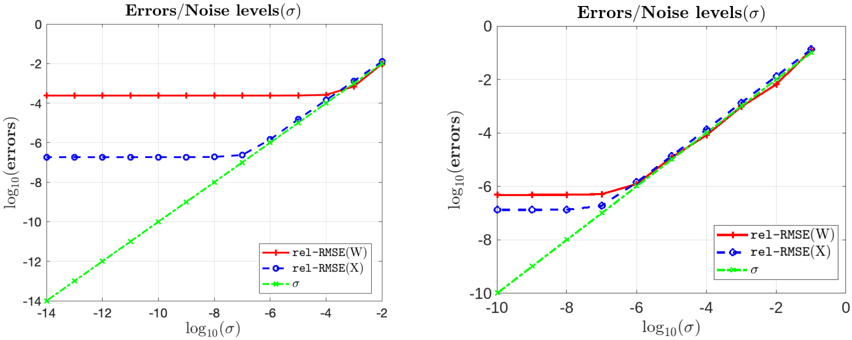
(a) Errors by **MinVol**



(b) Errors by **Square-Root Min-Vol**

Figure: Smallest errors corresponding to different noise levels by two algorithms

- For **Square-Root Min-Vol**, if use only $\lambda = 0.5$ for all noise levels



(a) Errors by **MinVol**

(b) Errors by **Square-Root Min-Vol**

Figure: Smallest errors corresponding to different noise levels by two algorithms

Observation: For **Square-root Min-Vol**, `rel-RMSE`(**W**) is much improved to $\mathbf{10^{-6}}$!

Convergence Theory

Proposition 3.2. (Convergence Theory of Square-Root Min-Vol NMF)

The limit points of the iterate sequence produced by **Square-Root Min-Vol NMF** are first-order stationary points of the problem (2).

Convergence Theory

Proposition 3.2. (Convergence Theory of Square-Root Min-Vol NMF)

The limit points of the iterate sequence produced by **Square-Root Min-Vol NMF** are first-order stationary points of the problem (2).

Process:

1. Show that all limits points of **Square-Root Min-Vol NMF** algorithm are fixed points of the algorithm map based on *Meyer's Monotone Convergence Theorem*.

Convergence Theory

Proposition 3.2. (Convergence Theory of Square-Root Min-Vol NMF)

The limit points of the iterate sequence produced by **Square-Root Min-Vol NMF** are first-order stationary points of the problem (2).

Process:

1. Show that all limits points of **Square-Root Min-Vol NMF** algorithm are fixed points of the algorithm map based on *Meyer's Monotone Convergence Theorem*.
2. Show that all fixed points of the algorithm map are first-order stationary points of the optimization problem (2).

Table of Contents

Conclusions

Conclusions:

- The efficiency of **MinVol** algorithm depends on the initial λ choice
- The **MinVol** algorithm can be improved by the **Square-Root Min-Vol NMF**

Conclusions

Conclusions:

- The efficiency of **MinVol** algorithm depends on the initial λ choice
- The **MinVol** algorithm can be improved by the **Square-Root Min-Vol NMF**

Open questions:

- Under what conditions is the square-root min-vol NMF provably guaranteed to be tuning-free?
- Can we design faster algorithms for solving the square-root min-vol NMF problem?

Table of Contents

Acknowledgement

Acknowledgement:

- This work was conducted as part of the 2023 REU STAT-DATASCI program hosted by the Department of Statistics at Rice University.
- This work was also partially funded by a grant from the National Institute of General Medical Sciences (R01GM135928: EC).

Reference

Alexandre Belloni, Victor Chernozhukov, and Lie Wang.
Square-root lasso: pivotal recovery of sparse signals via conic programming.
Biometrika, 98(4):791–806, 2011.

Xiao Fu, Kejun Huang, and Nicholas D. Sidiropoulos.
On identifiability of nonnegative matrix factorization.
IEEE Signal Processing Letters, 25(3):328–332, 2018.

Xiao Fu, Kejun Huang, Nicholas D Sidiropoulos, and Wing-Kin Ma.
Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications.
IEEE Signal Process. Mag., 36(2):59–80, 2019.

Nicolas Gillis.
Successive nonnegative projection algorithm for robust nonnegative blind source separation.
SIAM Journal on Imaging Sciences, 7(2):1420–1450, 2014.

Ending

THANK YOU!