Log-determinant Non-Negative Matrix Factorization via Successive Trace Approximation

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

Email: manshun.ang@umons.ac.be
Homepage: angms.science

May 23, 2018

Joint work with my supervisor : Nicolas Gillis (UMONS, Belgium)

The research problem (1/2)

Non-negative Matrix Factorization (NMF) : given

- Input matrix : $\mathbf{X} \in \mathbb{R}_{+}^{m \times n}$
- Factorization rank : r, positive integer

We consider

$$\min_{\substack{\mathbf{W} \geq 0 \\ \mathbf{H} \geq 0}} \frac{1}{2} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2.$$

- optimzation variables : $\mathbf{W} \in \mathbb{R}_{+}^{m \times r}, \mathbf{H} \in \mathbb{R}_{+}^{r \times n}$
- problem is : non-convex, NP-hard, ill-posed problem
- we consider low rank/complexity model $1 \leq r \leq \min\{m, n\}$.
- assumptions : (1) $\mathbf{W}$, $\mathbf{H}$ full rank, (2) r is known.

The research problem (2/2)

log-det Non-negative Matrix Factorization (NMF) : given

- Input matrix : $\mathbf{X} \in \mathbb{R}_+^{m \times n}$
- Factorization rank : r, positive integer

We consider

$$\min_{\substack{\mathbf{W} \geq 0 \\ \mathbf{H} \geq 0}} \frac{1}{2} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \frac{\lambda}{2} \log\det(\mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r).$$

- optimzation variables : $\mathbf{W} \in \mathbb{R}_+^{m \times r}, \mathbf{H} \in \mathbb{R}_+^{r \times n}$
- we consider low rank/complexity model $1 \leq r \leq \min\{m, n\}$.
- assumptions (1) $\mathbf{W}$, $\mathbf{H}$ full rank, (2) r is known.
- $\log\det(\mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r)$: volume regularizer, δ is constant
- $\lambda > 0$: regularization parameter

Solution framework : 2-Block Coordinate Descent (1/3)

Problem : given $(\mathbf{X}, r)$, solve

$$\min_{\substack{\mathbf{W}\geq 0\\ \mathbf{H}\geq 0}} \frac{1}{2}\|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \frac{\lambda}{2}\log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r).$$

Algorithm 1 BCD framework for logdet-NMF

1: **INPUT** : $\mathbf{X} \in \mathbb{R}^{m\times n}$, $r \in \mathbb{N}_+$
Initialization : $\mathbf{W} \in \mathbb{R}_+^{m\times r}$ and $\mathbf{H} \in \mathbb{R}_+^{r\times n}$
2: **OUTPUT** : $\mathbf{W} \in \mathbb{R}_+^{m\times r}$ and $\mathbf{H} \in \mathbb{R}_+^{r\times n}$
3: **for** $k = 1, 2, \dots$ **do**
4: Update($\mathbf{W}$) via $\arg\min_{\mathbf{W}\geq 0} \frac{1}{2}\|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \frac{\lambda}{2}\log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r)$.
5: Update($\mathbf{H}$) via $\arg\min_{\mathbf{H}\geq 0} \frac{1}{2}\|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2$.
6: **end for**

Solution framework : 2-Block Coordinate Descent (2/3)

- The subproblems are not symmetric.

$$\text{On } \mathbf{W}: \quad \arg\min_{\mathbf{W}\geq 0} \tfrac{1}{2}\|\mathbf{X}-\mathbf{W}\mathbf{H}\|_F^2 + \tfrac{\lambda}{2}\log\det(\mathbf{W}^\top\mathbf{W}+\delta\mathbf{I}_r)$$

$$\text{On } \mathbf{H}: \quad \arg\min_{\mathbf{H}\geq 0} \tfrac{1}{2}\|\mathbf{X}-\mathbf{W}\mathbf{H}\|_F^2$$

- Update($\mathbf{H}$) is easier, can be solved by FGM†

$$\mathbf{H} \leftarrow \text{FGM}(\mathbf{X},\mathbf{W},\mathbf{H})$$

- Update($\mathbf{W}$) is harder as $\log\det(\mathbf{W}^\top\mathbf{W}+\delta\mathbf{I}_r)$ is :
  - non-convex
  - column-coupled
  - non-proximable

† N. Gillis, "Successive Nonnegative Projection Algorithm for Robust Nonnegative Blind Source Separation", SIAM J. on Imaging Sciences 7 (2), pp. 1420-1450, 2014.

Theme of the presentation

To handle

$$\underset{\mathbf{W}\geq 0}{\arg\min}\ \frac{1}{2}\|\mathbf{X}-\mathbf{W}\mathbf{H}\|_F^2+\frac{\lambda}{2}\log\det(\mathbf{W}^\top\mathbf{W}+\delta\mathbf{I}_r)$$

in

Algorithm 2 BCD framework for logdet-NMF

INPUT : $\mathbf{X}\in\mathbb{R}^{m\times n}$, $r\in\mathbb{N}_+$
Initialization : $\mathbf{W}\in\mathbb{R}_+^{m\times r}$ and $\mathbf{H}\in\mathbb{R}_+^{r\times n}$
OUTPUT : $\mathbf{W}\in\mathbb{R}_+^{m\times r}$ and $\mathbf{H}\in\mathbb{R}_+^{r\times n}$

for $k=1,2,...$ **do**
    Update($\mathbf{W}$) via $\underset{\mathbf{W}\geq 0}{\arg\min}\ \frac{1}{2}\|\mathbf{X}-\mathbf{W}\mathbf{H}\|_F^2+\frac{\lambda}{2}\log\det(\mathbf{W}^\top\mathbf{W}+\delta\mathbf{I}_r)$.
    $\mathbf{H}\leftarrow$ FGM($\mathbf{X},\mathbf{W},\mathbf{H}$).
end for

Motivation (1/4)

Why : Mathematician don't ask why, just want to solve it

- Nobody solve it *effectively* yet

Application from data science : for $\mathbf{X} = \mathbf{W}\mathbf{H}$

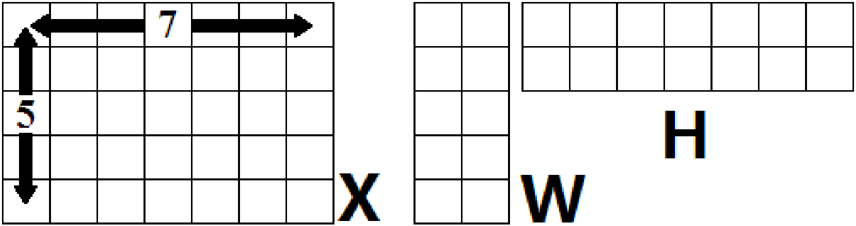
- $\mathbf{X}$ = data
- $\mathbf{W}$ = basis
- $\mathbf{H}$ = coefficient, weighting, encoding, membership
- r = model complexity
- m = dimension of data
- n = # of data points

Motivation (2/4)

Application from data science : for $\mathbf{X} = \mathbf{WH}$

- $\mathbf{X}$ = data
- $\mathbf{W}$ = basis
- $\mathbf{H}$ = coefficient, weighting, encoding, membership
- m = dimension of data
- n = # of data points
- r = model complexity

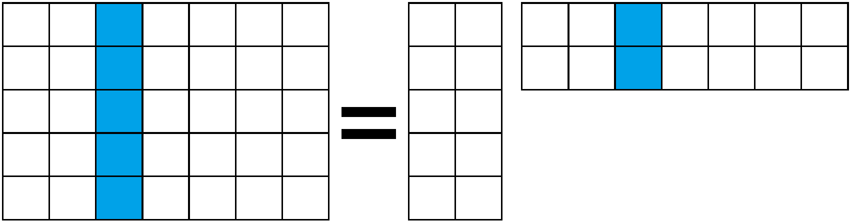
- Matrix $\mathbf{X}, \mathbf{W}$ and $\mathbf{H}$ with $m = 5, n = 7, r = 2$



Motivation (2/4)

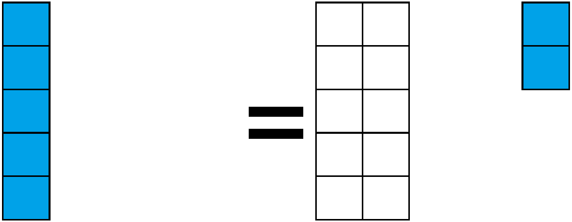
Application from data science : for $\mathbf{X} = \mathbf{WH}$

- $\mathbf{X}$ = data
- $\mathbf{W}$ = basis
- $\mathbf{H}$ = coefficient, weighting, encoding, membership
- m = dimension of data
- n = # of data points
- r = model complexity
- Consider 3^{rd} column.



Application from data science : for $\mathbf{X} = \mathbf{W}\mathbf{H}$

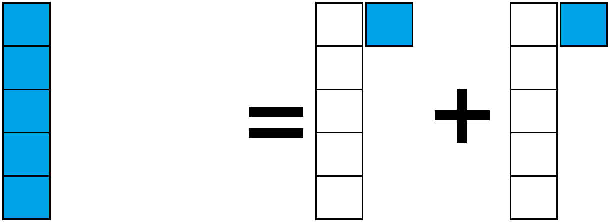
- $\mathbf{X}$ = data
- $\mathbf{W}$ = basis
- $\mathbf{H}$ = coefficient, weighting, encoding, membership
- m = dimension of data
- n = # of data points
- r = model complexity
- Consider 3^{rd} column



Motivation (2/4)

Application from data science : for $\mathbf{X} = \mathbf{WH}$

- $\mathbf{X}$ = data
- $\mathbf{W}$ = basis
- $\mathbf{H}$ = coefficient, weighting, encoding, membership
- m = dimension of data
- n = # of data points
- r = model complexity
- Linear combination :

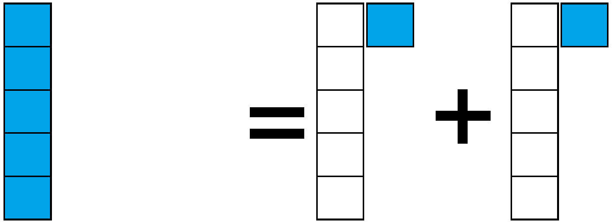


Motivation (2/4)

Application from data science : for $\mathbf{X} = \mathbf{WH}$

- $\mathbf{X}$ = data
- $\mathbf{W}$ = basis
- $\mathbf{H}$ = coefficient, weighting, encoding, membership
- m = dimension of data
- n = # of data points
- r = model complexity

- Non-negativity : conic combination

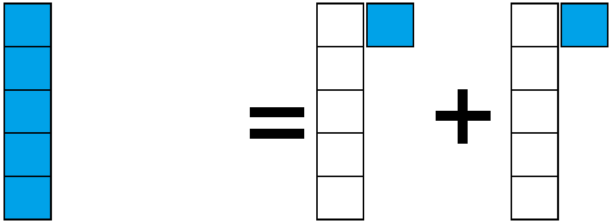


Motivation (2/4)

Application from data science : for $\mathbf{X} = \mathbf{W}\mathbf{H}$

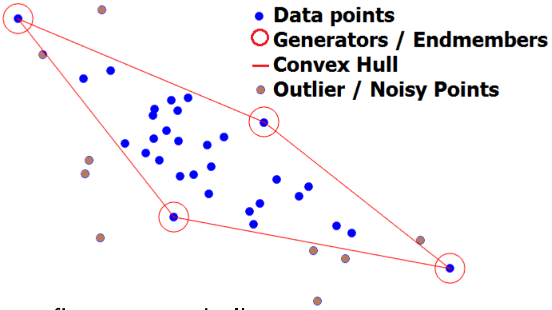
- $\mathbf{X}$ = data
- $\mathbf{W}$ = basis
- $\mathbf{H}$ = coefficient, weighting, encoding, membership
- m = dimension of data
- n = # of data points
- r = model complexity

- Non-negativity + normalization : convex combination



Application from data science : for $\mathbf{X} = \mathbf{W}\mathbf{H}$

- $\mathbf{X}$ = data
- $\mathbf{W}$ = vertex / generator
- $\mathbf{H}$ = membership
- m = dimension of data
- n = # of data points
- r = number of basis

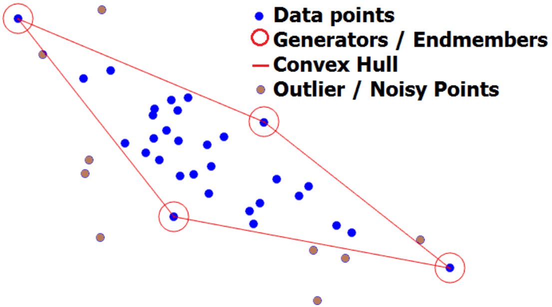


Geometrically : given data points, fit a convex hull

Motivation (3/4)

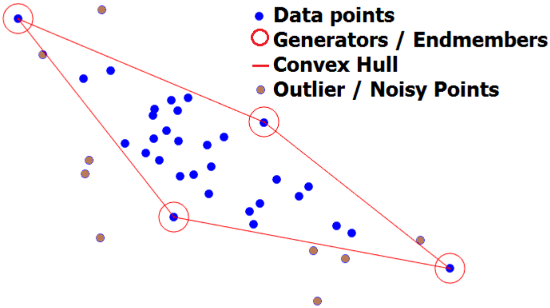
Application from data science : for $\mathbf{X} = \mathbf{W}\mathbf{H}$

- $\mathbf{X}$ = data
- $\mathbf{W}$ = vertex / generator
- $\mathbf{H}$ = membership
- m = dimension of data
- n = # of data points
- r = number of basis



Geometrically : given data points, fit a **minimum volume** convex hull

Geometry : fit a min. vol. convex hull



Optimization problem

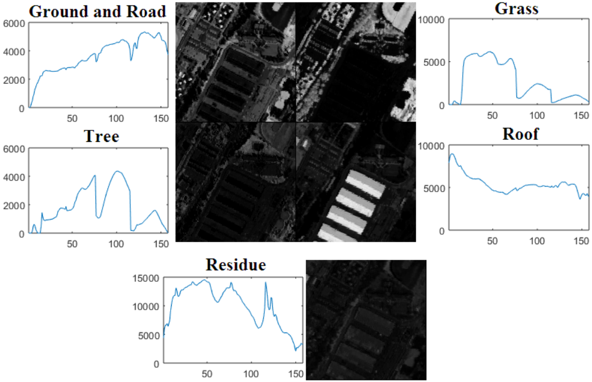
$$\min_{\substack{\mathbf{W}\geq 0\\ \mathbf{H}\geq 0\\ \mathbf{1}_r^\top \mathbf{H}\leq \mathbf{1}_n}} \frac{1}{2}\|\mathbf{X}-\mathbf{W}\mathbf{H}\|_F^2 + \frac{\lambda}{2}\log\det(\mathbf{W}^\top\mathbf{W}+\delta\mathbf{I}_r)$$

Many applications

Geoinformatic application : decomposition of hyper spectral images

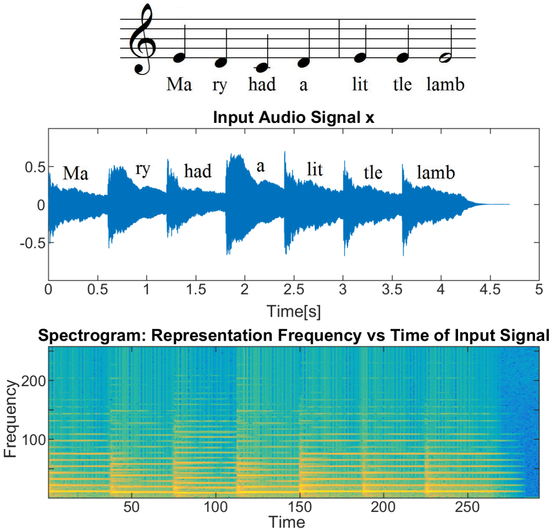
Output: $\mathbf{W}$ spectral basis, $\mathbf{H}$ weighting maps

Input : hyper-sepctral images data cube

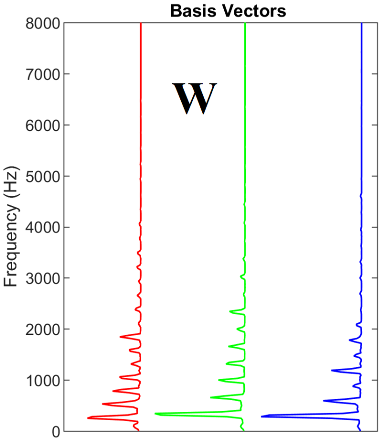


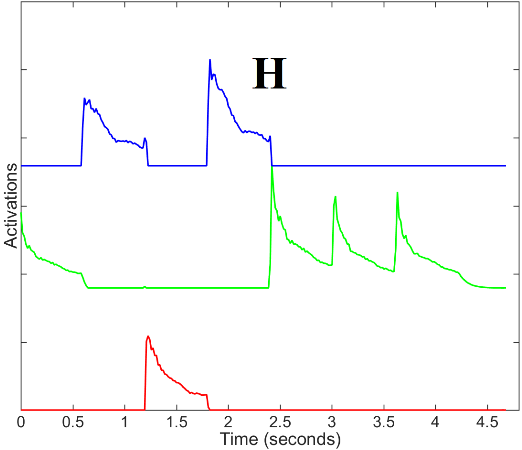
Application 2 : Source Separation (e.g. music)

Music input



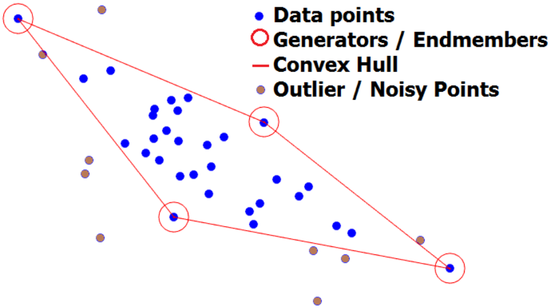
Decomposition output





On solving the problem

Geometry : fit a min. vol. convex hull



Optimization problem

$$\min_{\substack{\mathbf{W}\geq 0\\ \mathbf{H}\geq 0\\ \mathbf{1}_r^\top \mathbf{H}\leq \mathbf{1}_n}} \frac{1}{2}\|\mathbf{X}-\mathbf{W}\mathbf{H}\|_F^2 + \frac{\lambda}{2}\log\det(\mathbf{W}^\top\mathbf{W}+\delta\mathbf{I}_r)$$

The key to solve such problem is to use majorization-minimization

How to solve The key iequality : logdet-trace inequality

Given a positive definite matrix $\mathbf{A} \in \mathbb{R}^{r \times r}$, we have
$\log\det \mathbf{A} \leq \mathrm{tr}(\mathbf{A} - \mathbf{I}_r)$.
So we now have an upper bound : put $\mathbf{A} = \mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r$

$$\begin{aligned} \log\det(\mathbf{W}^t \mathbf{W} + \delta \mathbf{I}_r) &\leq \mathrm{tr}(\mathbf{W}^\top \mathbf{W} + (\delta - 1)\mathbf{I}_r) \\ \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda \log\det(\mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r) &\leq \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda\, \mathrm{tr}(\mathbf{W}^\top \mathbf{W} + (\delta - 1)\mathbf{I}_r) \end{aligned}$$

- $\log\det(\mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r)$ is not convex w.r.t. $\mathbf{W}$ but the trace is.

How to solve The key iequality : logdet-trace inequality

Given a positive definite matrix $\mathbf{A} \in \mathbb{R}^{r \times r}$, we have
$\log\det \mathbf{A} \leq \mathrm{tr}(\mathbf{A} - \mathbf{I}_r)$.
So we now have an upper bound : put $\mathbf{A} = \mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r$

$$\begin{aligned} \log\det(\mathbf{W}^t\mathbf{W} + \delta\mathbf{I}_r) &\leq \mathrm{tr}(\mathbf{W}^\top\mathbf{W} + (\delta - 1)\mathbf{I}_r) \\ \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda \log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r) &\leq \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda\,\mathrm{tr}(\mathbf{W}^\top\mathbf{W} + (\delta - 1)\mathbf{I}_r) \end{aligned}$$

- $\log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r)$ is not convex w.r.t. $\mathbf{W}$ but the trace is.
- Algorithm that minimizes this upper bound :

---

for $k = 1$ to itermax **do**
  $\mathbf{W} \leftarrow \underset{\mathbf{W} \geq 0}{\arg\min} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda\,\mathrm{tr}(\mathbf{W}^\top\mathbf{W} + (\delta - 1)\mathbf{I}_r)$.
  $\mathbf{H} \leftarrow \mathrm{FGM}(\mathbf{X}, \mathbf{W}, \mathbf{H})$.
end for

- Don't stop here, it can be better !!

A closer look on the logdet-trace inequality

- Given a positive definite matrix $\mathbf{A} \in \mathbb{R}^{r \times r}$, we have

$$\log\det\mathbf{A} \leq \mathrm{tr}(\mathbf{A} - \mathbf{I}_r).$$

A closer look on the logdet-trace inequality

- Given a positive definite matrix $\mathbf{A} \in \mathbb{R}^{r \times r}$, we have

$$\log\det \mathbf{A} \leq \operatorname{tr}(\mathbf{A} - \mathbf{I}_r).$$

- Let μ denotes eigenvalues, we have

$$\det \mathbf{A} = \prod_i \mu_i \quad \text{and} \quad \operatorname{tr} \mathbf{A} = \sum_i \mu_i.$$

- $\implies$

$$\log\det \mathbf{A} \leq \operatorname{tr}(\mathbf{A} - \mathbf{I}_r) \iff \sum_i \log \mu_i \leq \sum_i (\mu_i - 1).$$

- $\log \mu_i$ means matrix $\mathbf{A}$ has to be positive definite ($\mu_i > 0 \, \forall \, i$), which is satisfied for $\mathbf{A} = \mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r$.

A closer look on the logdet-trace inequality

- Given a positive definite matrix $\mathbf{A} \in \mathbb{R}^{r \times r}$, we have

$$\log\det \mathbf{A} \leq \operatorname{tr}(\mathbf{A} - \mathbf{I}_r).$$

- Let μ denotes eigenvalues, we have

$$\det \mathbf{A} = \prod_i \mu_i \quad \text{and} \quad \operatorname{tr} \mathbf{A} = \sum_i \mu_i.$$

- $\implies$

$$\log\det \mathbf{A} \leq \operatorname{tr}(\mathbf{A} - \mathbf{I}_r) \iff \sum_i \log \mu_i \leq \sum_i (\mu_i - 1).$$

- $\log \mu_i$ means matrix $\mathbf{A}$ has to be positive definite ($\mu_i > 0 \, \forall \, i$), which is satisfied for $\mathbf{A} = \mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r$.
- $\sum_i \log \mu_i \leq \sum_i (\mu_i - 1) \;\Leftarrow\; \log \mu_i \leq \mu_i - 1 \, \forall \, i$. We can focuse on the inequality $\log \mu_i \leq \mu_i - 1$ with $\mu_i \geq 0$

On $\log x \leq x - 1$, $x \geq 0$

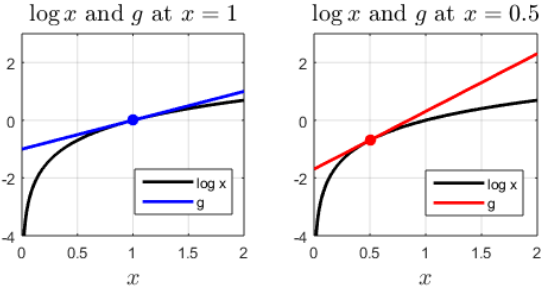
- $\log x$ is concave.
- $x - 1$ is the first order Taylor approximation of $\log x$ at $x = 1$.
- $x - 1$ is the only **convex-tight** upper bound of $\log x$.†
- Tight : $x - 1$ touch $\log x$ at the point $x = 1$.

† Higher order Taylor approximation of $\log x$ is tight, more accurate but not convex.

On $\log x \leq x - 1$, $x \geq 0$

- $\log x$ is concave.
- $x - 1$ is the first order Taylor approximation of $\log x$ at $x = 1$.
- $x - 1$ is the only **convex-tight** upper bound of $\log x$.†
- Tight : $x - 1$ touch $\log x$ at the point $x = 1$.
- Generalize to point x_0 : $\log x \leq g(x|x_0) = a_1(x_0)x + a_0(x_0)$ is

$$\log x \leq \frac{1}{x_0}x + \log x_0 - 1.$$



† Higher order Taylor approximation of $\log x$ is tight, more accurate but not convex.

A parametric trace upper bound for $\log\det \mathbf{A}$

$$
\begin{aligned}
\log\det \mathbf{A} &= \sum \log \mu_i \\
&\leq \sum \frac{1}{\mu_i^-}\mu_i + \log \mu_i^- - 1 \\
&\leq \sum \frac{1}{\mu_{\min}^-}\mu_i + \log \mu_i^- - 1 \\
&= \operatorname{tr}(\mathbf{D}^1\mathbf{A} + \mathbf{D}^0)
\end{aligned}
$$

$\mathbf{D}^1 = \frac{1}{\mu_{\min}^-}\mathbf{I}_r$, $\mathbf{D}^0 = \text{Diag}(\log \mu_i^- - 1)$, μ_i^- is μ_i of the previous step

Put $\mathbf{A} = \mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r$, we have

$$
\begin{aligned}
\log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r) &\leq \operatorname{tr}(\mathbf{D}^1\mathbf{W}^\top\mathbf{W} + \delta\mathbf{D}^1 + \mathbf{D}^0) \\
\text{(ignore constants)} &= \operatorname{tr}\mathbf{D}^1\mathbf{W}^\top\mathbf{W}
\end{aligned}
$$

Comparing upper bounds for $\log\det\mathbf{A}$

The original function $F(\mathbf{W}) = \log\det(\mathbf{W}^\top\mathbf{W} + \delta\mathbf{I}_r)$ is upper bounded by :

$$\begin{aligned} \text{Eigen bound} \quad (B_1): \quad & \operatorname{tr}\mathbf{D}^1\mathbf{W}^\top\mathbf{W} + \text{constants.} \\ \text{Taylor bound} \quad (B_2): \quad & \operatorname{tr}\mathbf{D}^{\text{Taylor}}\mathbf{W}^\top\mathbf{W} + \text{constants.} \end{aligned}$$

- Constants $\mathbf{D}^1$, $\mathbf{D}^0$ are defined as before, and constant $\mathbf{D}^{\text{Taylor}} = (\mathbf{W}_{-1}^\top\mathbf{W}_{-1} + \delta\mathbf{I}_r)^{-1}$.
- Both bounds are trace functional with an **relaxation gap** :
  - (B_1) has eigen gap $\mu_i \geq \mu_{\min}$
  - (B_2) has convexification gap
  - $\mathbf{D}^1$ is **diagonal** but $\mathbf{D}^{\text{Taylor}}$ is not (it is dense) $\implies$ column-wise decomposition is possible

Successive Trace Approximation (STA)

Algorithm 3 Successive Trace Approximation

```
INPUT: X ∈ R_+^{m×n}, r ∈ N_+, λ > 0, δ > 0.
OUTPUT: W ∈ R_+^{m×r} and H ∈ R_+^{r×n}.
INITIALIZATION : W ∈ R_+^{m×r} , H ∈ R_+^{r×n}, D^1 = I_r
for K = 1 to itermax do
    for k = 1 to itermax do
        W ← argmin_{W≥0} ||X − WH||_F^2 + λ tr D^1 W^T W.
        H ← FGM(X, W, H).
    end for
    μ_i ← svd(W^T W + δI_r), D^1 = Diag(μ_min^{-1})
end for
```

INPUT: $\mathbf{X} \in \mathbb{R}_+^{m \times n}$, $r \in \mathbb{N}_+$, $\lambda > 0$, $\delta > 0$.

OUTPUT: $\mathbf{W} \in \mathbb{R}_+^{m \times r}$ and $\mathbf{H} \in \mathbb{R}_+^{r \times n}$.

INITIALIZATION : $\mathbf{W} \in \mathbb{R}_+^{m \times r}$, $\mathbf{H} \in \mathbb{R}_+^{r \times n}$, $\mathbf{D}^1 = \mathbf{I}_r$

for $K = 1$ to itermax **do**

- **for** $k = 1$ to itermax **do**
  - $\mathbf{W} \leftarrow \underset{\mathbf{W} \geq 0}{\arg\min} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 + \lambda \operatorname{tr} \mathbf{D}^1 \mathbf{W}^\top \mathbf{W}$.
  - $\mathbf{H} \leftarrow$ FGM$(\mathbf{X}, \mathbf{W}, \mathbf{H})$.
- **end for**
- $\mu_i \leftarrow$svd$(\mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r)$, $\mathbf{D}^1 =$Diag$(\mu_{\min}^{-1})$

end for

Improving STA

The STA algorithm with decomposed $f(w_i)$

Algorithm 4 Successive Trace Approximation

INPUT: $\mathbf{X} \in \mathbb{R}_+^{m\times n}$, $r \in \mathbb{N}_+$, $\lambda > 0$ and $\delta > 0$
OUTPUT: $\mathbf{W} \in \mathbb{R}_+^{m\times r}$ and $\mathbf{H} \in \mathbb{R}_+^{r\times n}$
INITIALIZATION : $\mathbf{W} \in \mathbb{R}_+^{m\times r}$, $\mathbf{H} \in \mathbb{R}_+^{r\times n}$ and $\mathbf{D}^1 = I_r$, $\gamma = 10^{-6}$
for $K = 1$ to itermax **do**
  for $k = 1$ to itermax **do**
    for $i = 1$ to r **do**
      $w_i = \arg\min_{w_i \geq 0} f(w_i) = \|\mathbf{X}_i - w_i h_i\|_F^2 + \lambda \mathbf{D}_{ii}^1 \|w_i\|_2^2 + \frac{\gamma}{2}\|w_i - w_i^-\|_2^2$
      $\mathbf{H} \leftarrow$ FGM$(\mathbf{X}, \mathbf{W}, \mathbf{H})$.
    end for
  end for
  $\mu_i \leftarrow$ svd$(\mathbf{W}^\top \mathbf{W} + \delta I)$ and $\mathbf{D}^1 =$Diag$(\mu_{\min}^{-1})$
end for

Improving STA

Finally we have

Algorithm 5 STA, Final form

INPUT: $\mathbf{X} \in \mathbb{R}_+^{m \times n}$, $r \in \mathbb{N}_+$, $\lambda > 0$ and $\delta > 0$

OUTPUT: $\mathbf{W} \in \mathbb{R}_+^{m \times r}$ and $\mathbf{H} \in \mathbb{R}_+^{r \times n}$

INITIALIZATION : $\mathbf{W} \in \mathbb{R}_+^{m \times r}$, $\mathbf{H} \in \mathbb{R}_+^{r \times n}$ and $\mathbf{D}^1 = \mathbf{I}_r$, $\gamma = 10^{-6}$

for $K = 1$ to itermax **do**

  for $k = 1$ to itermax **do**

    $P = \mathbf{X}\mathbf{H}^\top$ and $Q = \mathbf{H}\mathbf{H}^\top$.

    for $i = 1$ to r **do**

$$w_i = \frac{\left[P_i - \sum_{j=1}^{i-1} w_j Q_{ji} - \sum_{j=i+1}^{r} w_j^- Q_{ji} + \gamma w_i^-\right]_+}{Q_{ii} + \lambda \mathbf{D}_{ii}^1 + \frac{\gamma}{2}}$$

      $\mu_i \leftarrow \text{svd}(\mathbf{W}^\top \mathbf{W} + \delta \mathbf{I}_r)$ and $D^1 = \text{Diag}(\mu_{\min}^{-1})$

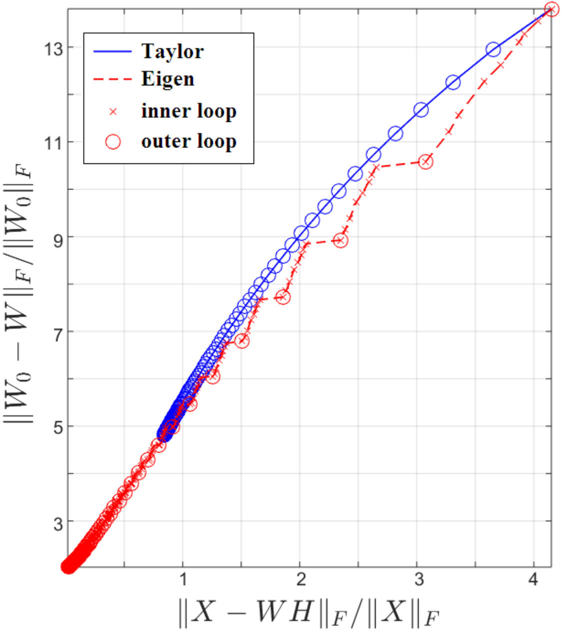
    end for

    $\mathbf{H} \leftarrow \text{FGM}(\mathbf{X}, \mathbf{W}, \mathbf{H})$.

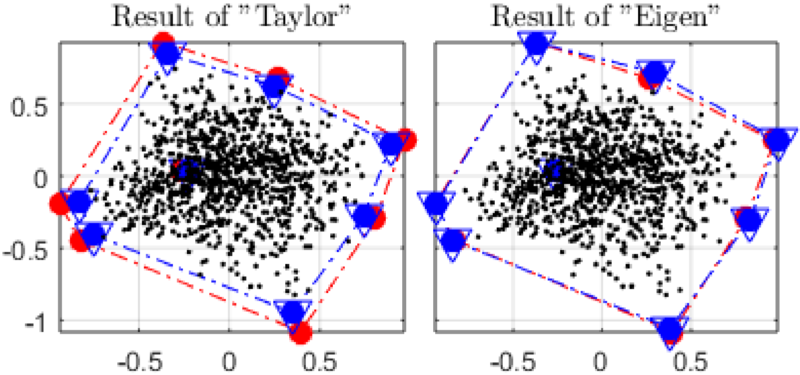
  end for

end for

Experimental result : comparing the two logdet inequality



Experimental result : comparing the two logdet inequality



Last page - summary

- Nonegative Matrix Factorization with logdet regularizer
- Algorithmic development of solving the logdet NMF

Slides (and code) avaliable at *angms.science*

– END OF PRESENTATION –