Inter-University Research Institute Corporation / Research Organization of Information and Systems
National Institute of Informatics

Fast Rank-1 NMF for Missing Data with KL Divergence

Kazu Ghalamkari[1,2], Mahito Sugiyama[1,2]

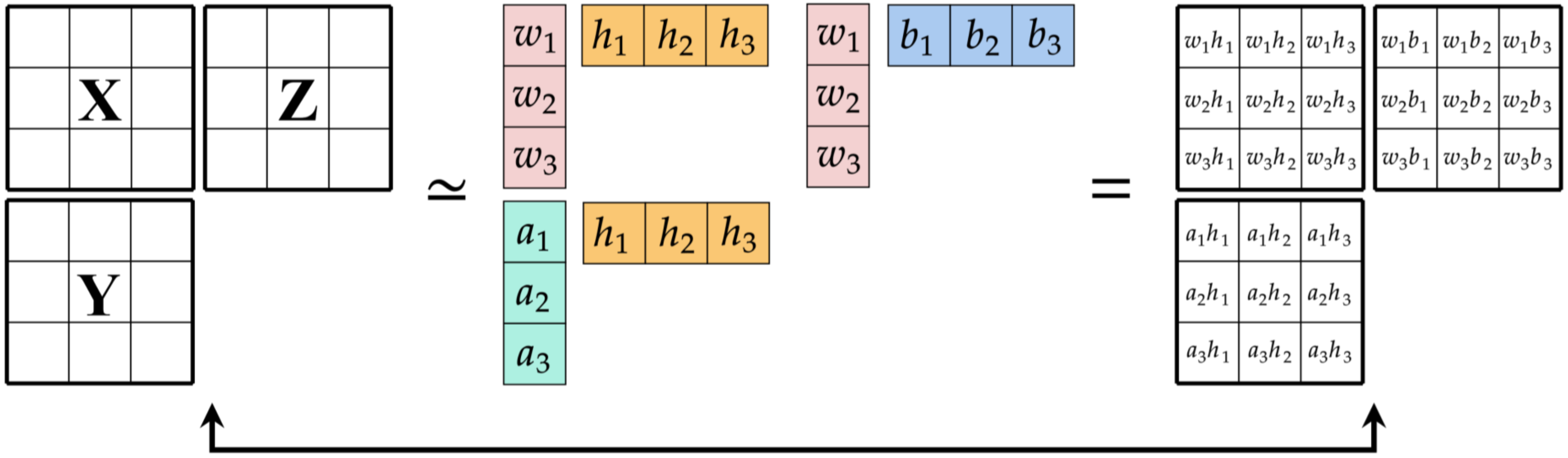
1 : The Graduate University for Advanced Studies, SOKENDAI
2 : National Institute of Informatics

github.com/gkazunii/A1GM

The 25th International Conference on Artificial Intelligence and Statistics (AISTATS 2022)

Contributions

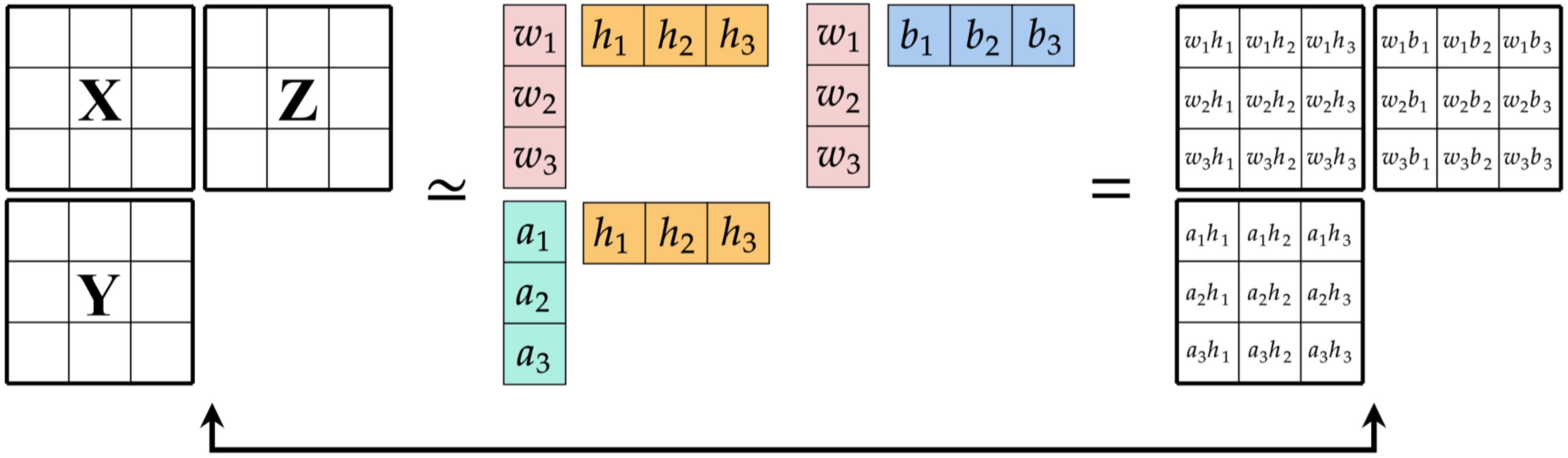
- ☐ Closed formula of the best rank-1 NMMF w.r.t. minimizing KL divergence



$$D(\mathbf{X}, w \otimes h) + \alpha D(\mathbf{Y}, a \otimes h) + \beta D(\mathbf{Z},\ w \otimes b)$$

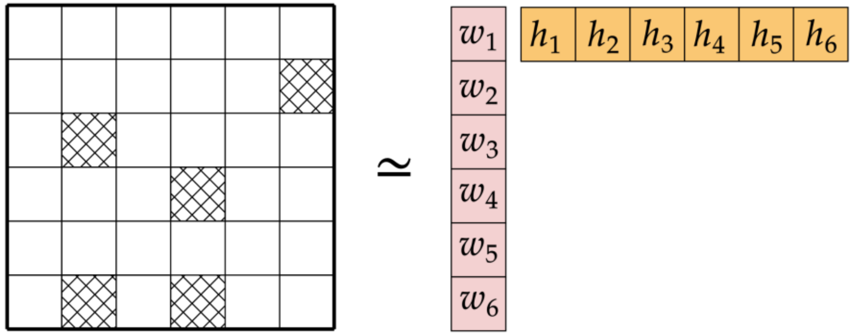
Contributions

- ☐ Closed formula of the best rank-1 NMMF w.r.t. minimizing KL divergence



$$D(\mathbf{X}, w \otimes h) + \alpha D(\mathbf{Y}, a \otimes h) + \beta D(\mathbf{Z}, w \otimes b)$$

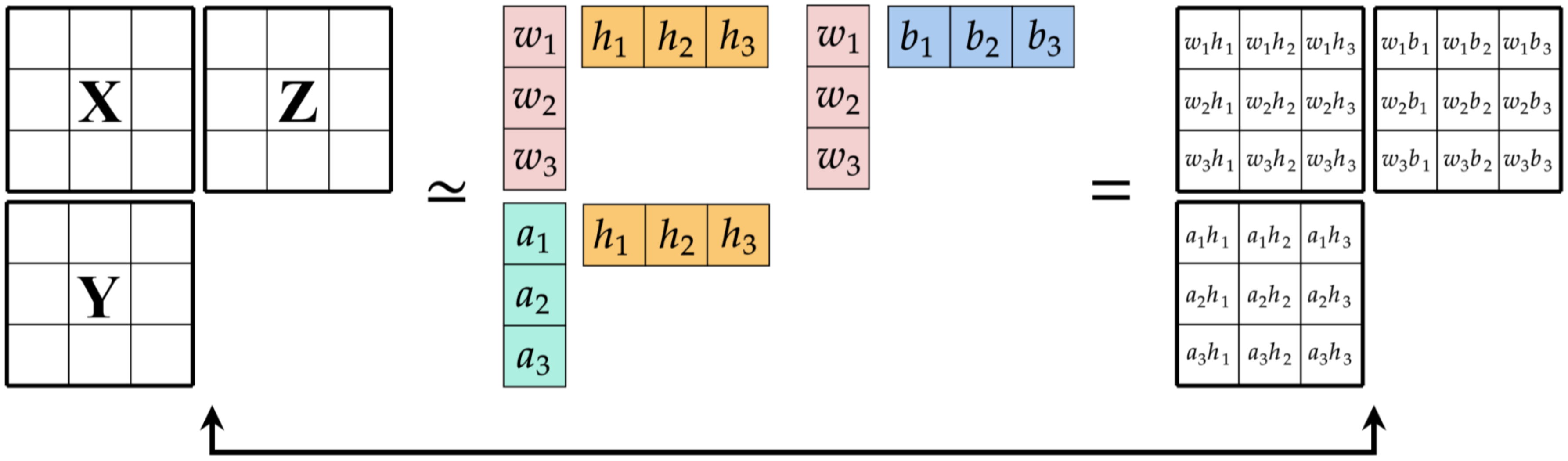
- ☐ A1GM: Faster method for rank-1 NMF with missing values



Non-gradient-based method.

No worries about initial values, stopping criterion and learning rate 😄

Closed formula of the best rank-1 NMMF



$$D(\mathbf{X}, w \otimes h) + \alpha D(\mathbf{Y}, a \otimes h) + \beta D(\mathbf{Z}, w \otimes b)$$

Theorem 1.

For given $\mathbf{X} \in \mathbb{R}_{>0}^{I \times J}$, $\mathbf{Y} \in \mathbb{R}_{>0}^{N \times J}$, and $\mathbf{Z} \in \mathbb{R}_{>0}^{I \times M}$ the best rank-1 NMMF is given as

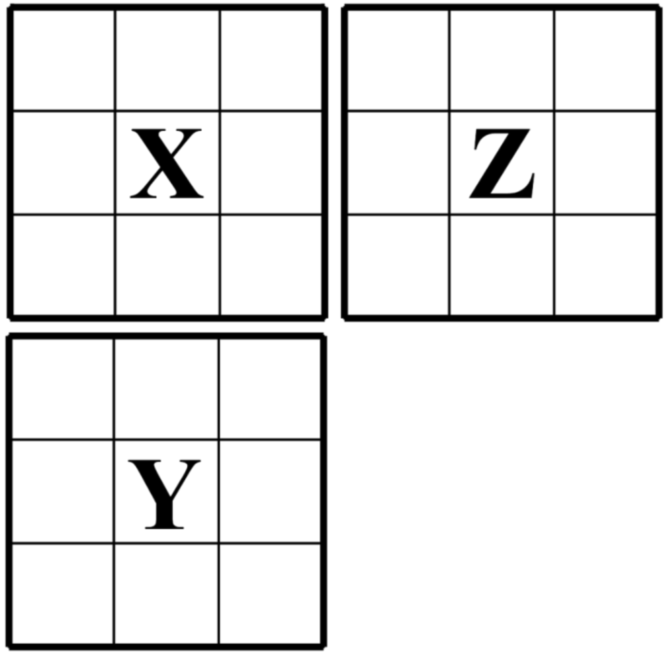
$$w_i = \frac{\sqrt{S(\mathbf{X})}}{S(\mathbf{X}) + \beta S(\mathbf{Z})} \left\{ \sum_{j=1}^{J} \mathbf{X}_{ij} + \beta \sum_{m=1}^{M} \mathbf{Z}_{im} \right\} \qquad a_n = \frac{\sum_{j=1}^{J} \mathbf{Y}_{nj}}{\sqrt{S(\mathbf{X})}}$$

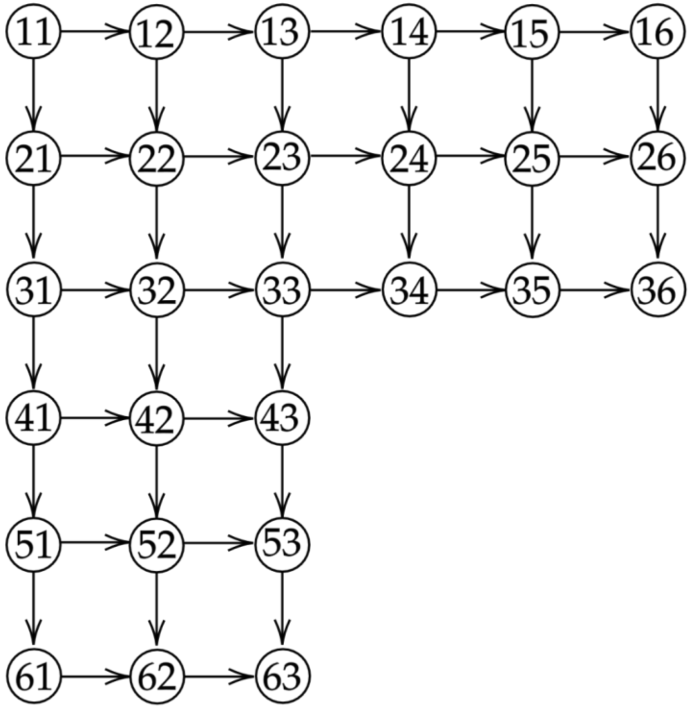
$$h_j = \frac{\sqrt{S(\mathbf{X})}}{S(\mathbf{X}) + \alpha S(\mathbf{Y})} \left\{ \sum_{i=1}^{I} \mathbf{X}_{ij} + \alpha \sum_{n=1}^{N} \mathbf{Y}_{nj} \right\} \qquad b_m = \frac{\sum_{i=1}^{I} \mathbf{Z}_{im}}{\sqrt{S(\mathbf{X})}}$$

$S(\mathbf{X})$ is sum of all elements of $\mathbf{X}$.

Modeling of NMMF

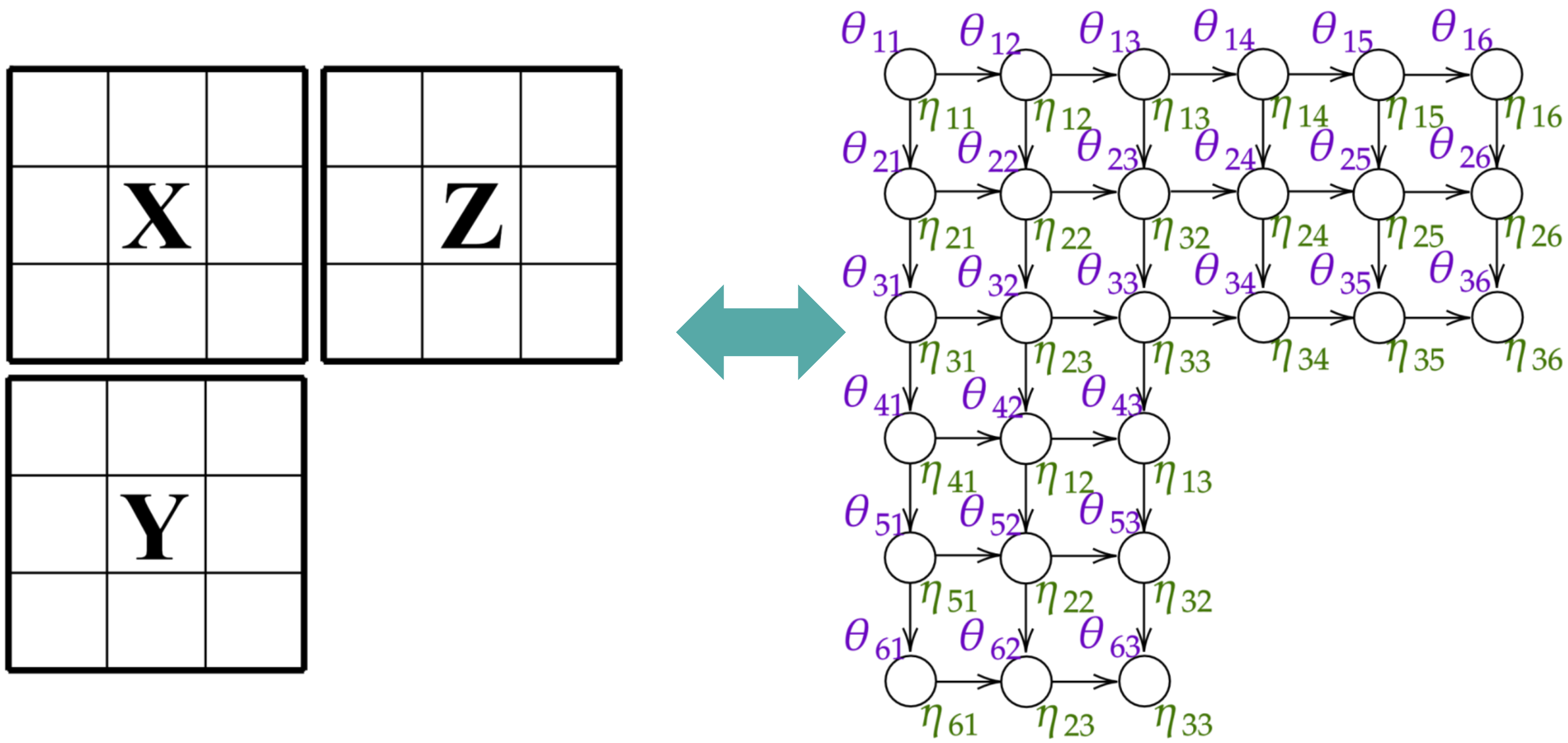
- Introduce partial order structure in the input of NMMF





Modeling of NMMF

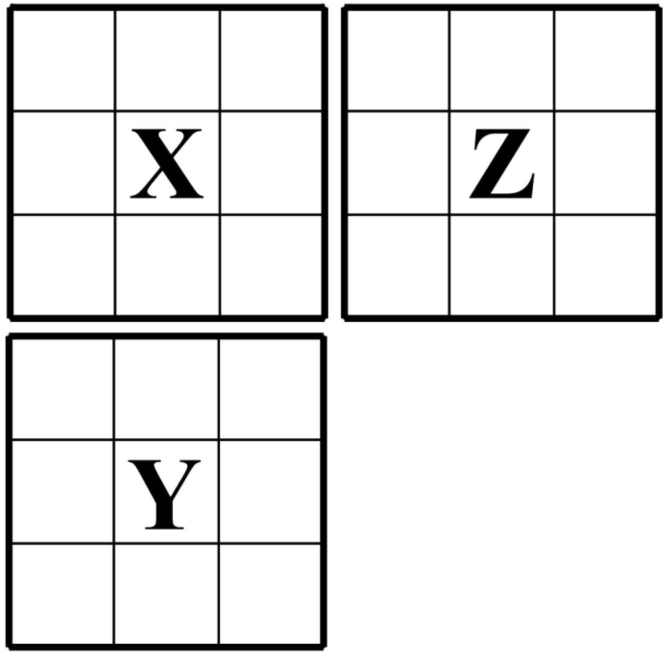
- Each node has θ- and η-parameters

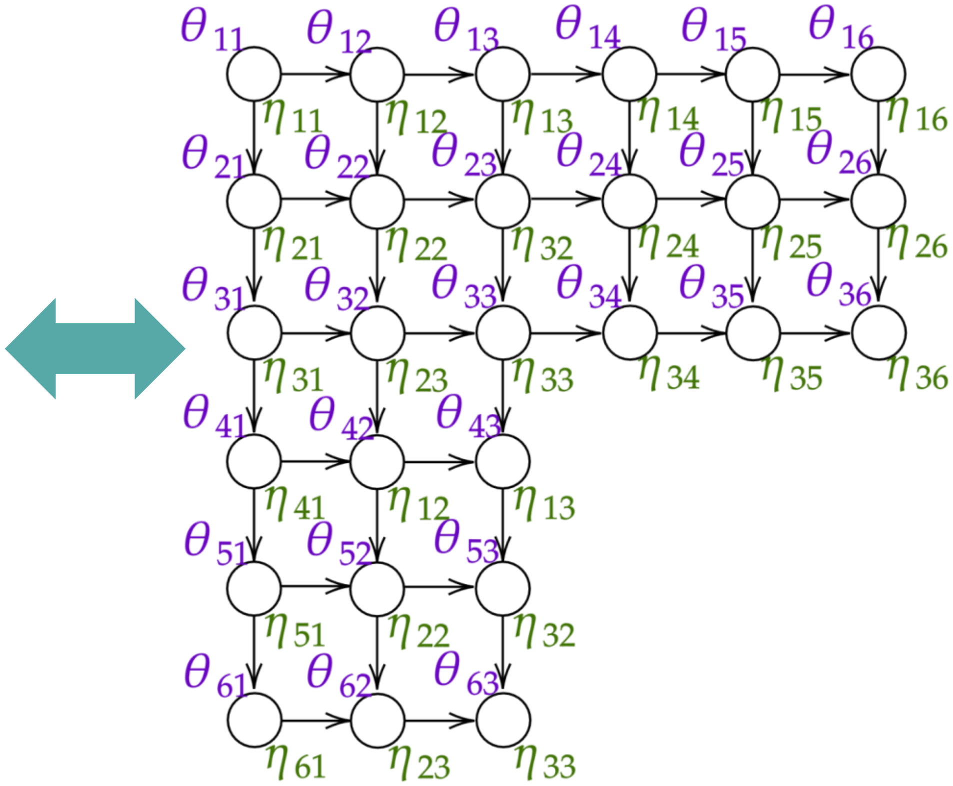


Modeling of NMMF

$$S(\mathbf{X}) + S(\mathbf{Y}) + S(\mathbf{Z}) = 1$$

- Each node has θ- and η-parameters

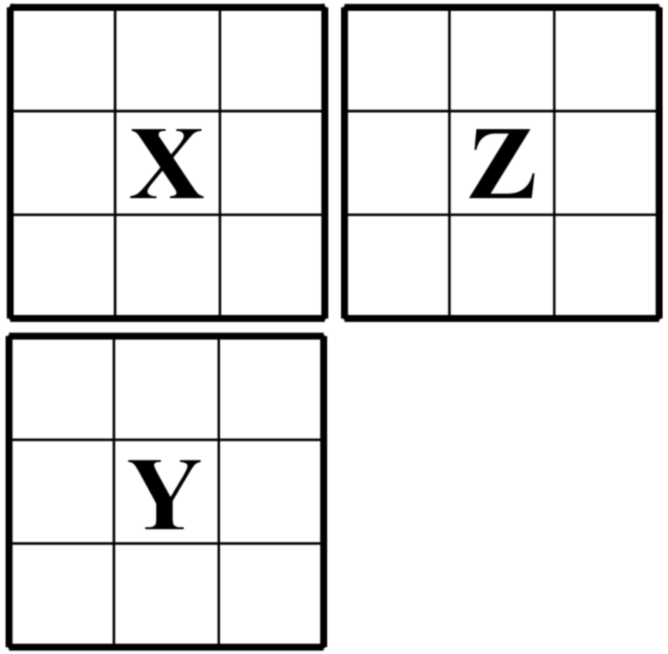


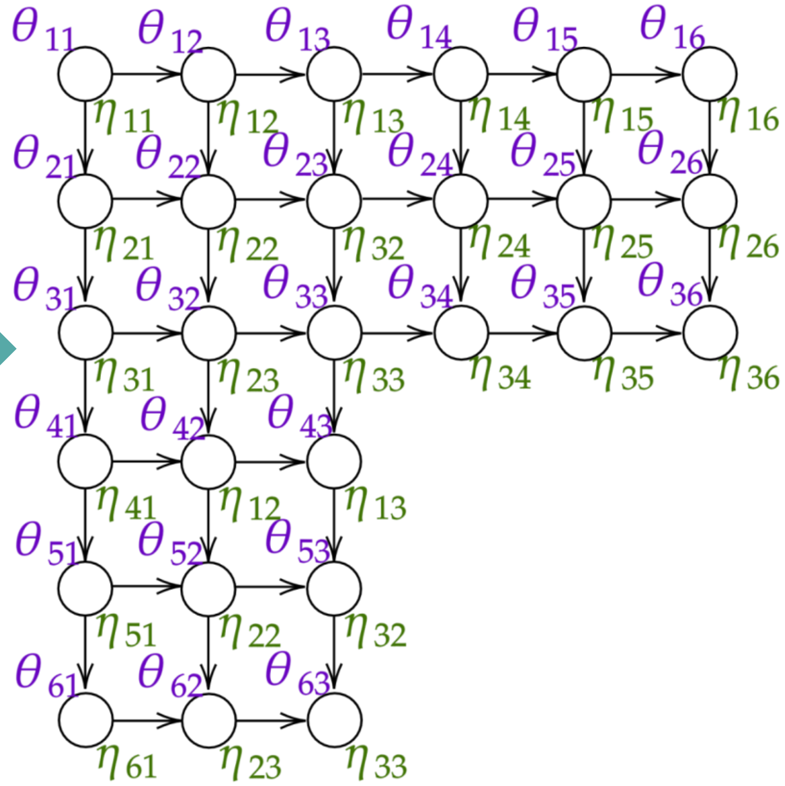


Modeling of NMMF

- ☐ Each node has θ- and η-parameters

$$S(\mathbf{X}) + S(\mathbf{Y}) + S(\mathbf{Z}) = 1$$



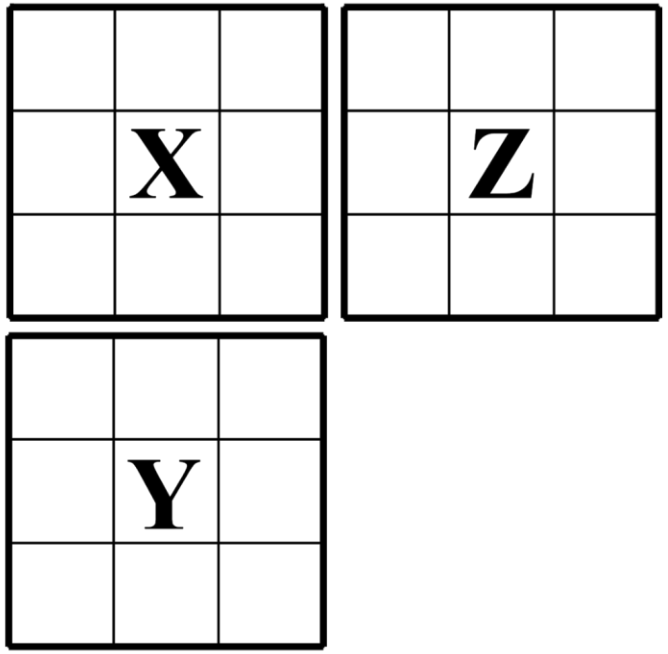


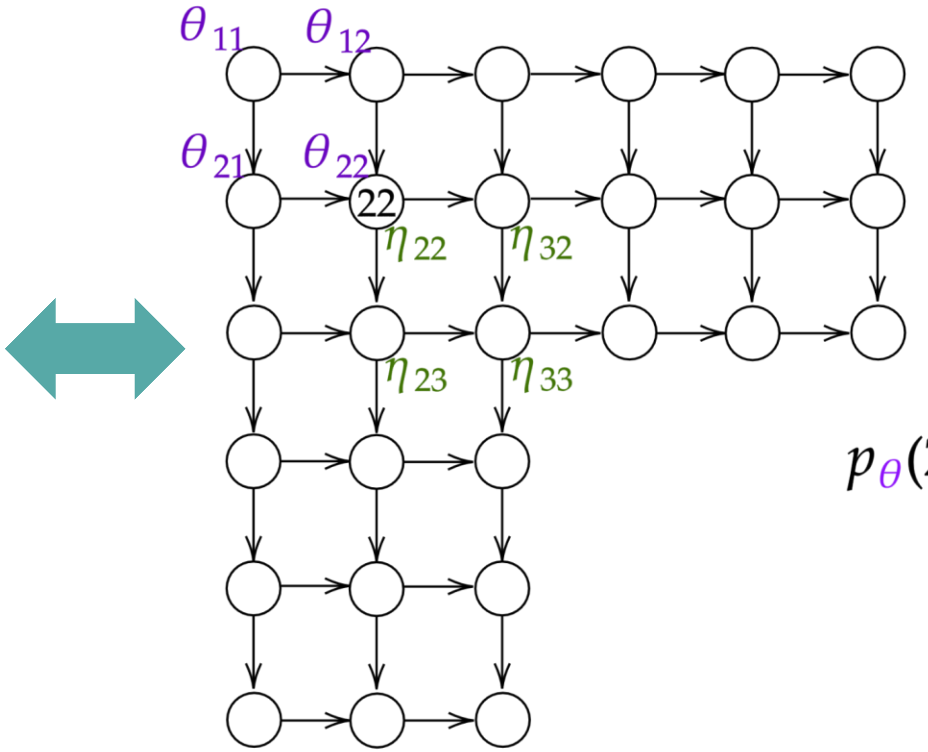
$$p_\theta(k,l) = \exp\left(\sum_{(s,t)\le(k,l)} \theta_{st} \right), \quad \eta_{kl} = \sum_{(k,l)\le(s,t)} p(s,t).$$

Modeling of NMMF

- ☐ Example for $p(2,2)$.

$$S(\mathbf{X}) + S(\mathbf{Y}) + S(\mathbf{Z}) = 1$$





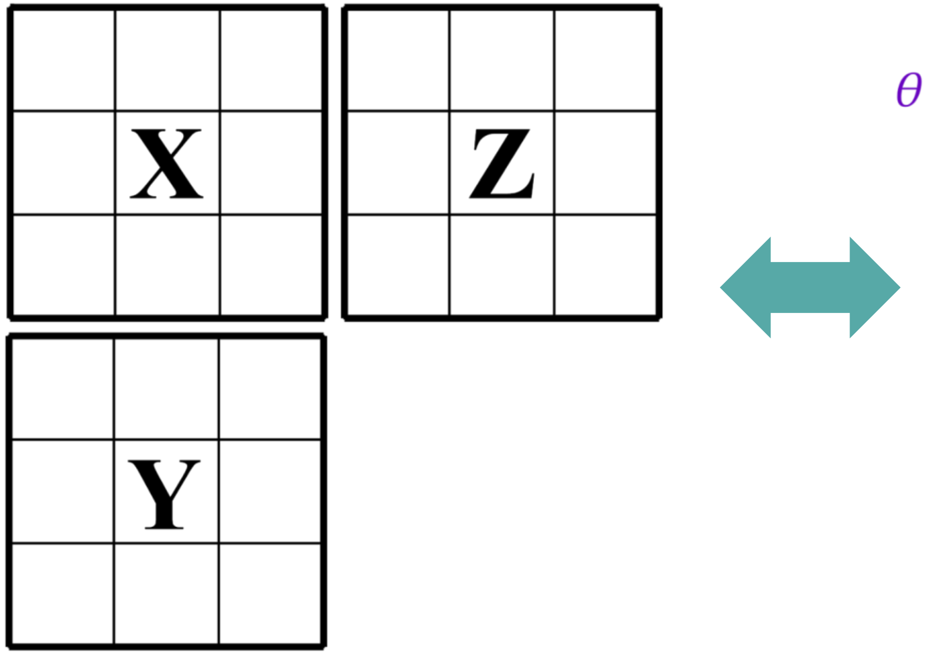
$$p_\theta(2,2) = \exp(\theta_{11} + \theta_{21} + \theta_{12} + \theta_{22})$$

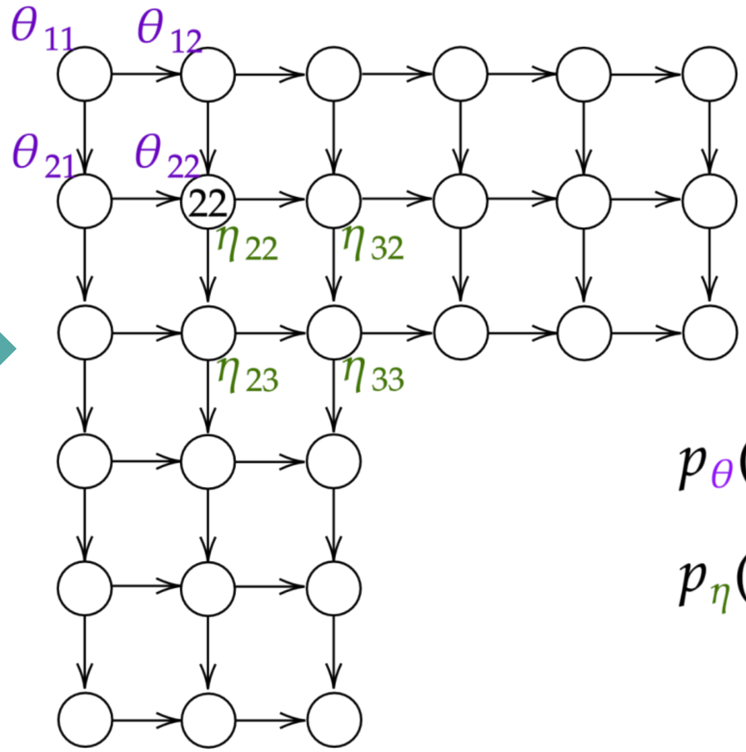
$$p_\theta(k,l) = \exp\left(\sum_{(s,t)\le(k,l)} \theta_{st}\right), \quad \eta_{kl} = \sum_{(k,l)\le(s,t)} p(s,t).$$

Modeling of NMMF

$$S(\mathbf{X}) + S(\mathbf{Y}) + S(\mathbf{Z}) = 1$$

- Example for $p(2,2)$.





$$p_\theta(2,2) = \exp(\theta_{11} + \theta_{21} + \theta_{12} + \theta_{22})$$

$$p_\eta(2,2) = \eta_{22} - \eta_{23} - \eta_{32} + \eta_{33}$$

Möbius inversion formula

$$p_\eta(i,j) = \sum_{(k,l)} \mu_{ij}^{kl} \eta_{kl}$$

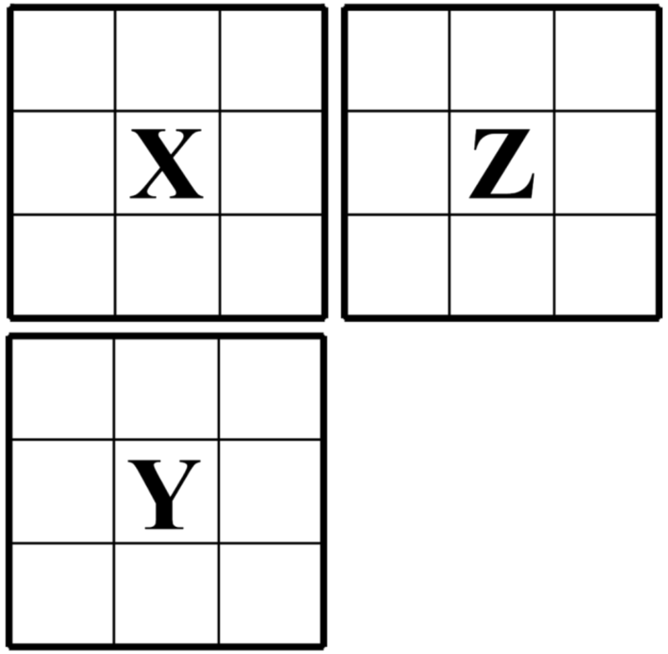
$$\mu_{ij}^{kl} = \begin{cases} -\sum_{(i,j)\le(s,t)<(k,l)} \mu_{ij}^{st} & (i,j) < (k,l) \\ 1 & (i,j) = (k,l) \\ 0 & \textit{otherwiese} \end{cases}$$

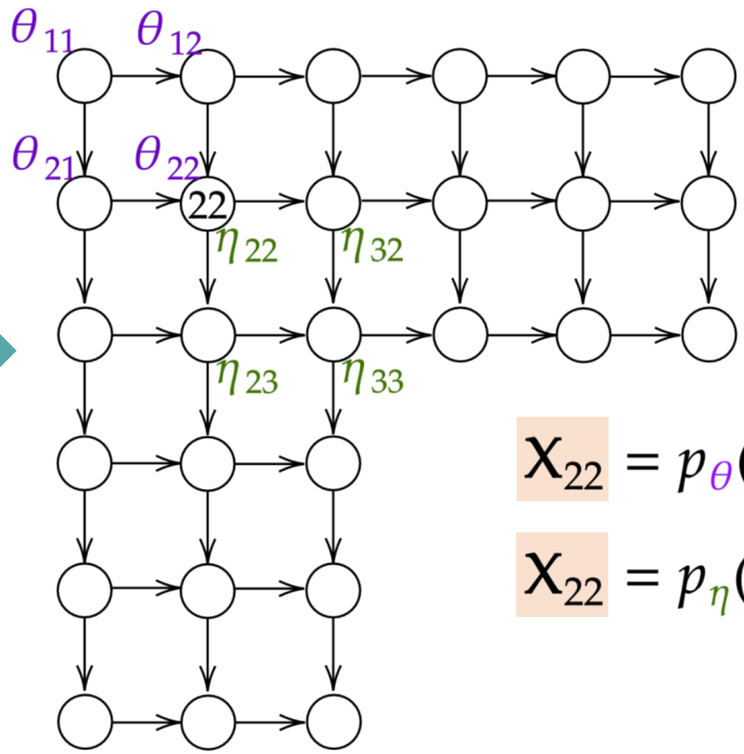
$$p_\theta(k,l) = \exp\left(\sum_{(s,t)\le(k,l)} \theta_{st} \right), \quad \eta_{kl} = \sum_{(k,l)\le(s,t)} p(s,t).$$

Modeling of NMMF

$$S(\mathbf{X}) + S(\mathbf{Y}) + S(\mathbf{Z}) = 1$$

- Example for $p(2,2)$.





$$\mathbf{X}_{22} = p_\theta(2,2) = \exp(\theta_{11} + \theta_{21} + \theta_{12} + \theta_{22})$$

$$\mathbf{X}_{22} = p_\eta(2,2) = \eta_{22} - \eta_{23} - \eta_{32} + \eta_{33}$$

Möbius inversion formula

$$p_\eta(i,j) = \sum_{(k,l)} \mu_{ij}^{kl} \eta_{kl}$$

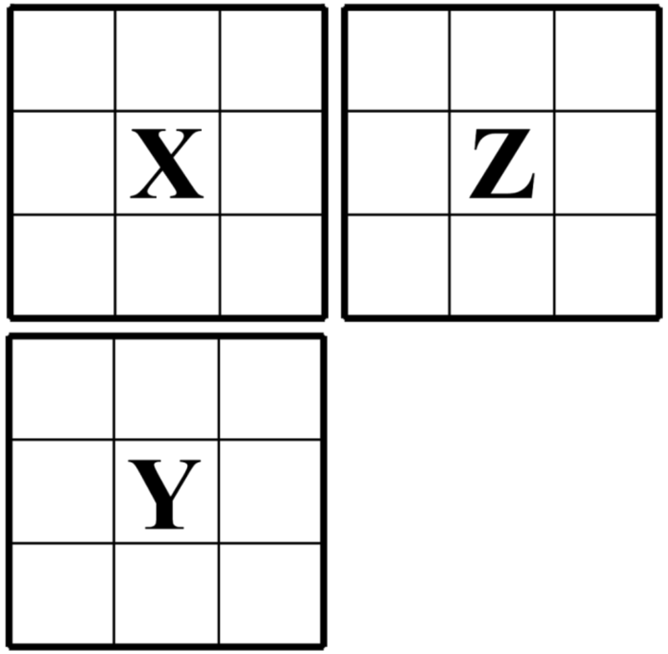
$$\mu_{ij}^{kl} = \begin{cases} -\sum_{(i,j)\le(s,t)<(k,l)} \mu_{ij}^{st} & (i,j) < (k,l) \\ 1 & (i,j) = (k,l) \\ 0 & otherwiese \end{cases}$$

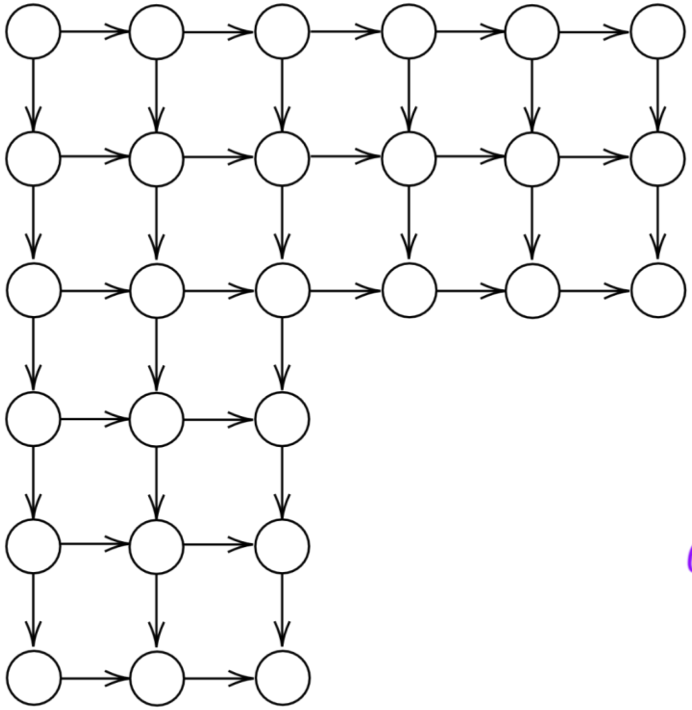
$$p_\theta(k,l) = \exp\left(\sum_{(s,t)\le(k,l)} \theta_{st} \right), \quad \eta_{kl} = \sum_{(k,l)\le(s,t)} p(s,t).$$

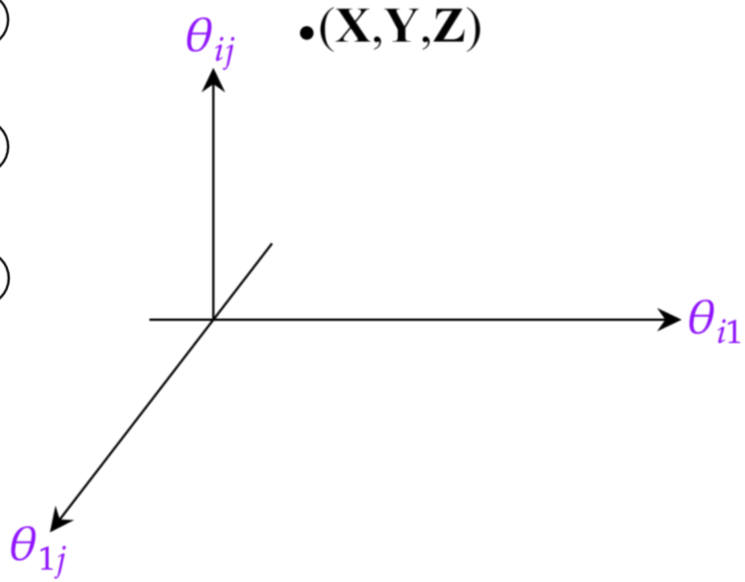
Information geometry of rank-1 NMMF

- ☐ Introducing θ- and η-coordinate

$$S(\mathbf{X}) + S(\mathbf{Y}) + S(\mathbf{Z}) = 1$$



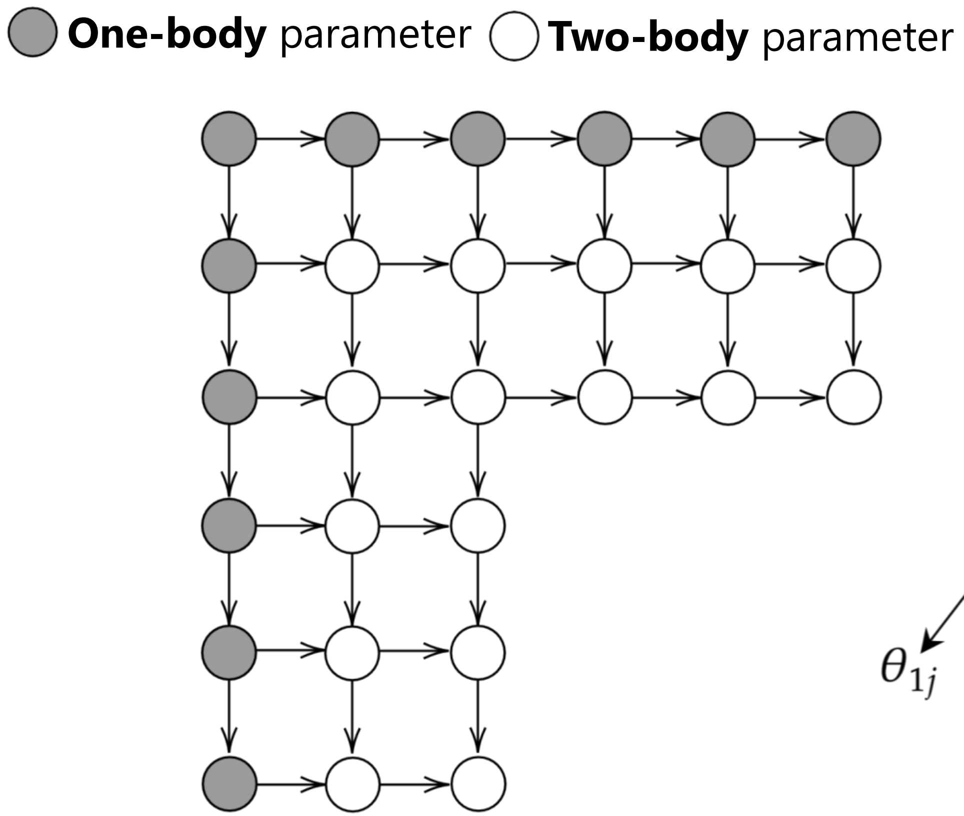


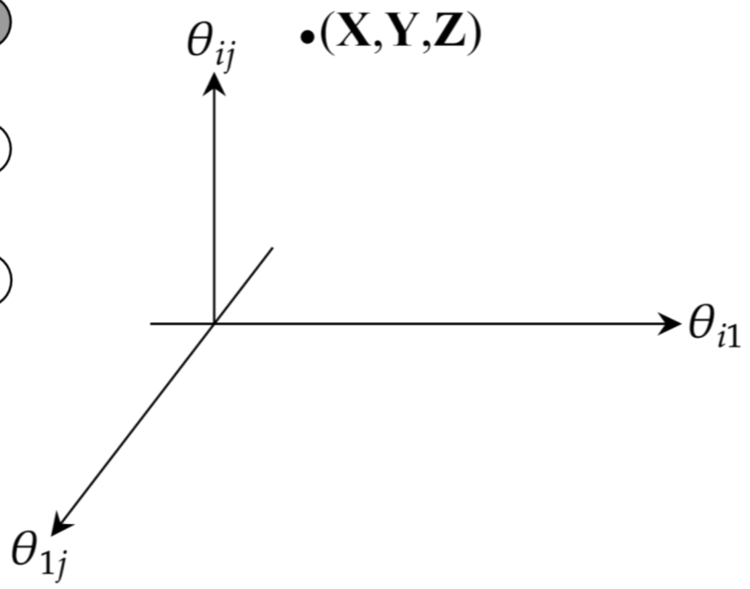


$$p_\theta(k,l) = \exp\left(\sum_{(s,t)\leq(k,l)} \theta_{st} \right), \quad \eta_{kl} = \sum_{(k,l)\leq(s,t)} p(s,t).$$

Information geometry of rank-1 NMMF

$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is *simultaneously rank-1 decomposable.* $\Leftrightarrow$ It can be written as $(\boldsymbol{w} \otimes \boldsymbol{h}, \boldsymbol{a} \otimes \boldsymbol{h}, \boldsymbol{w} \otimes \boldsymbol{b})$.





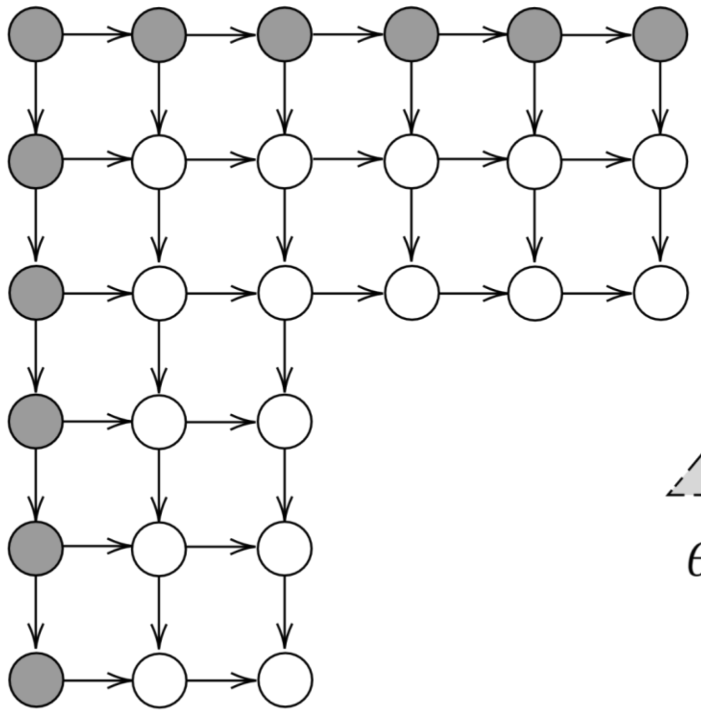
Information geometry of rank-1 NMMF

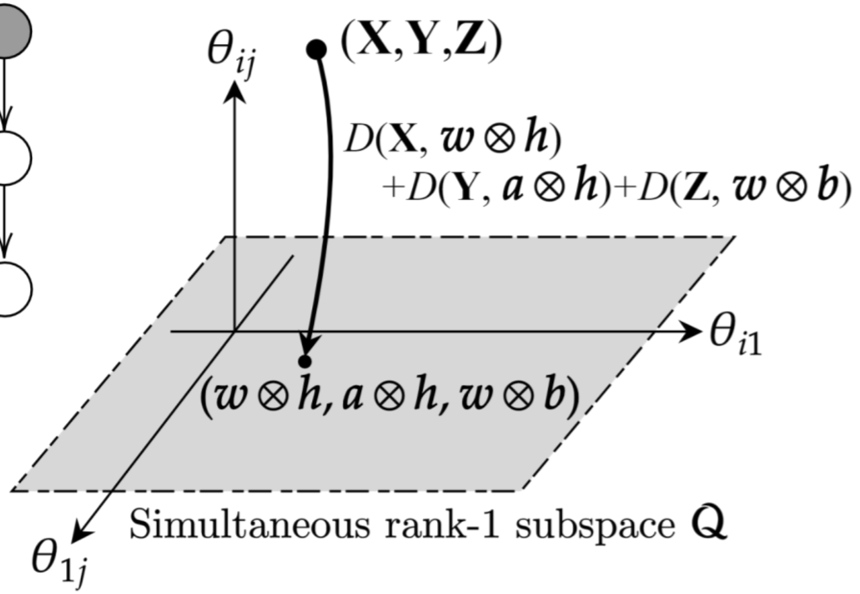
$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is *simultaneously rank-1 decomposable.* $\Leftrightarrow$ It can be written as $(\boldsymbol{w} \otimes \boldsymbol{h}, \boldsymbol{a} \otimes \boldsymbol{h}, \boldsymbol{w} \otimes \boldsymbol{b})$.

One-body parameter **Two-body** parameter

Simultaneous Rank-1 $\boldsymbol{\theta}$-condition

Its all two-body θ-parameters are 0.





Information geometry of rank-1 NMMF

$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is *simultaneously rank-1 decomposable*. $\Leftrightarrow$ It can be written as $(\boldsymbol{w} \otimes \boldsymbol{h}, \boldsymbol{a} \otimes \boldsymbol{h}, \boldsymbol{w} \otimes \boldsymbol{b})$.

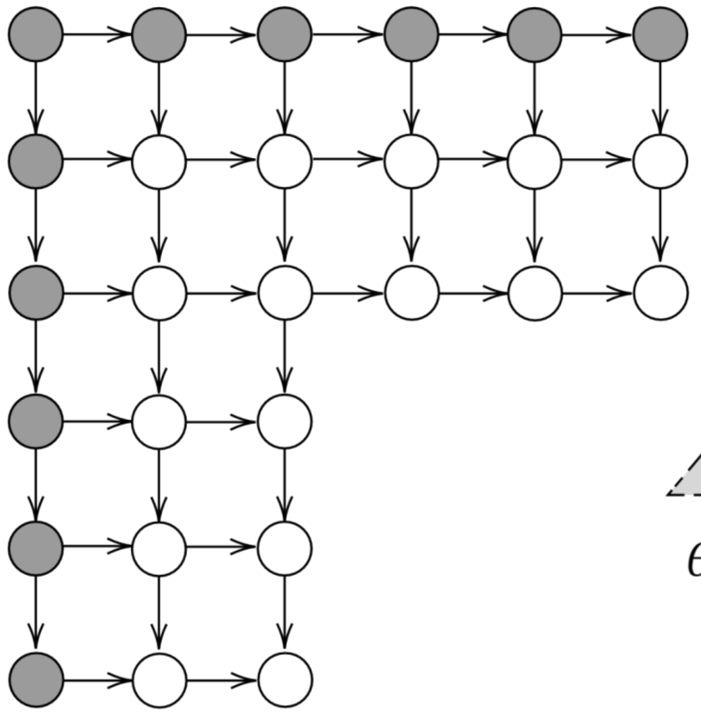
One-body parameter **Two-body** parameter

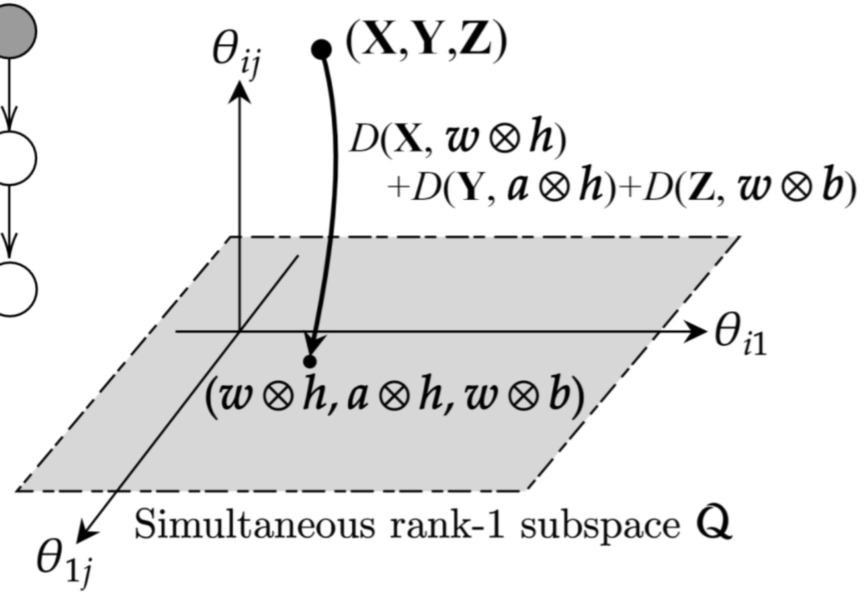
Simultaneous Rank-1 $\boldsymbol{\theta}$-condition

Its all two-body θ-parameters are 0.

Simultaneous Rank-1 $\boldsymbol{\eta}$-condition

$$\eta_{ij} = \eta_{i1}\eta_{1j}$$





Information geometry of rank-1 NMMF

$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is *simultaneously rank-1 decomposable*. ⇔ It can be written as $(\boldsymbol{w} \otimes \boldsymbol{h}, \boldsymbol{a} \otimes \boldsymbol{h}, \boldsymbol{w} \otimes \boldsymbol{b})$.

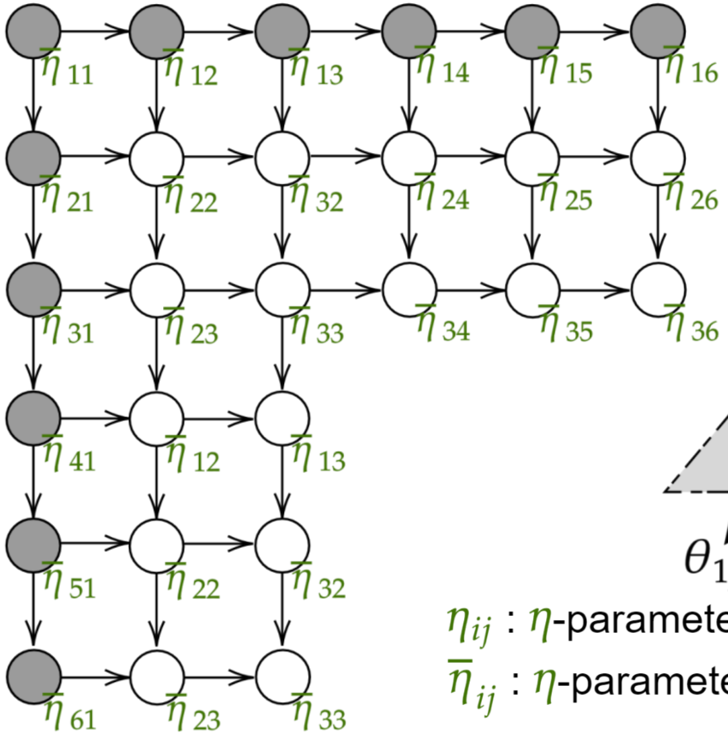
One-body parameter **Two-body** parameter

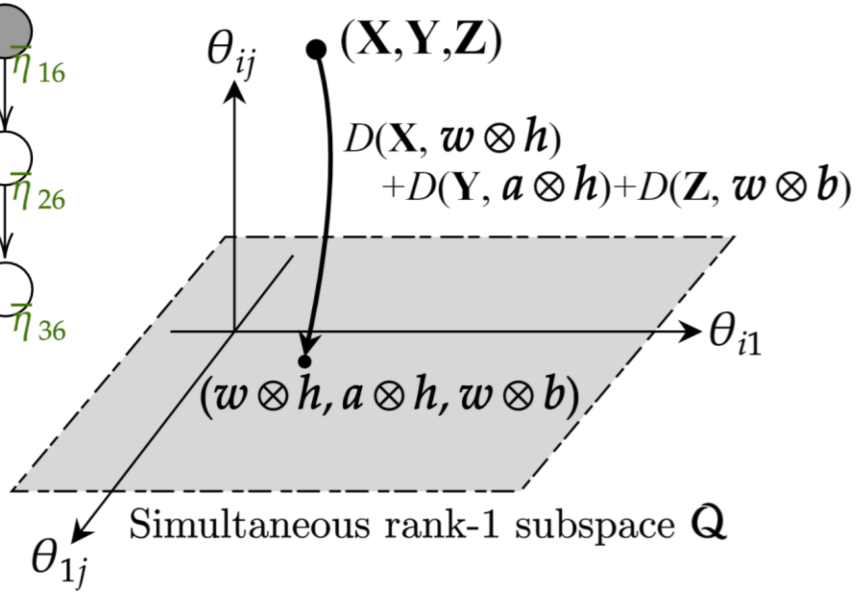
Simultaneous Rank-1 $\boldsymbol{\theta}$-condition

Its all two-body θ-parameters are 0.

Simultaneous Rank-1 $\boldsymbol{\eta}$-condition

$$\eta_{ij} = \eta_{i1}\eta_{1j}$$





η_{ij} : η-parameter before the projection.

$\bar{\eta}_{ij}$: η-parameter after the projection.

Information geometry of rank-1 NMMF

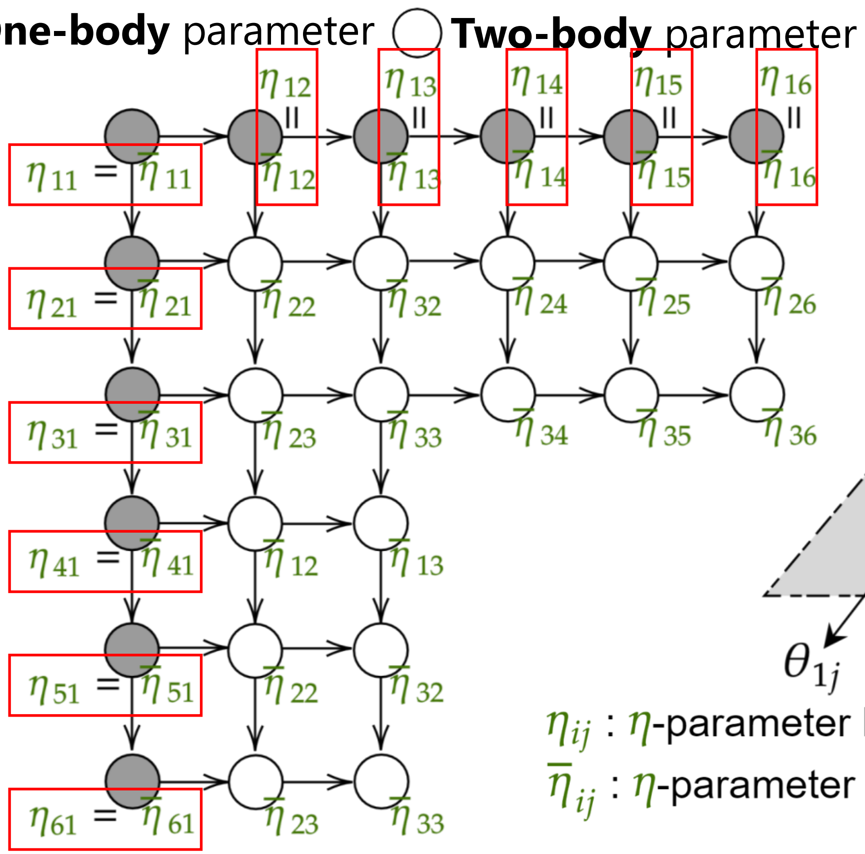
$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is *simultaneously rank-1 decomposable*. $\Leftrightarrow$ It can be written as $(\boldsymbol{w} \otimes \boldsymbol{h}, \boldsymbol{a} \otimes \boldsymbol{h}, \boldsymbol{w} \otimes \boldsymbol{b})$.

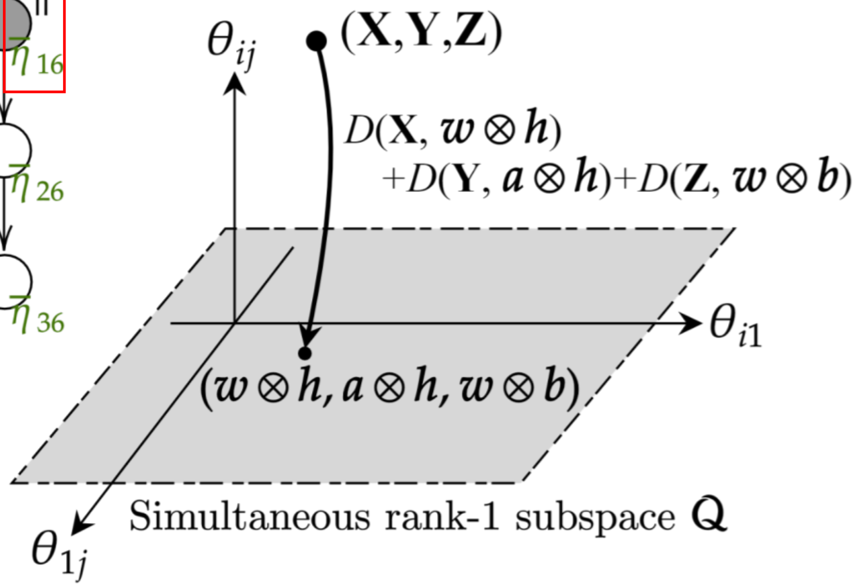
Simultaneous Rank-1 $\boldsymbol{\theta}$-condition

Its all two-body θ-parameters are 0.

Simultaneous Rank-1 $\boldsymbol{\eta}$-condition

$$\eta_{ij} = \eta_{i1}\eta_{1j}$$





η_{ij} : η-parameter before the projection.

$\bar{\eta}_{ij}$: η-parameter after the projection.

One-body $\boldsymbol{\eta}$-parameters do not change before or after the projection.

Information geometry of rank-1 NMMF

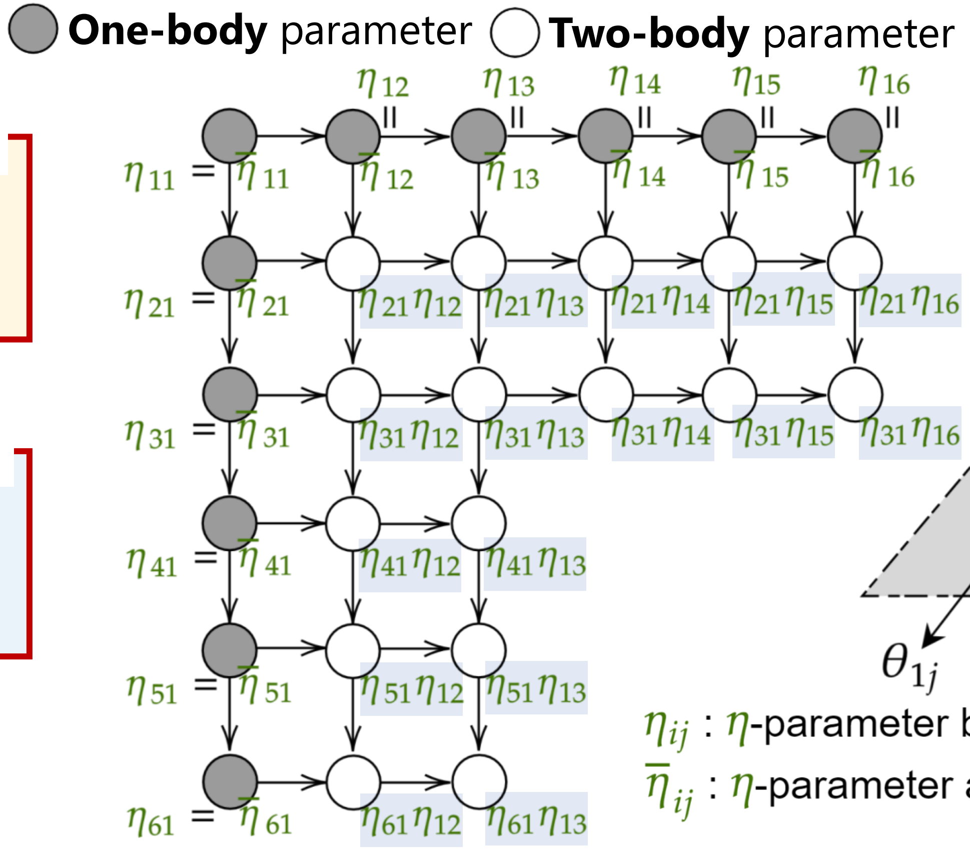
$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is *simultaneously rank-1 decomposable.* $\Leftrightarrow$ It can be written as $(\boldsymbol{w} \otimes \boldsymbol{h}, \boldsymbol{a} \otimes \boldsymbol{h}, \boldsymbol{w} \otimes \boldsymbol{b})$.

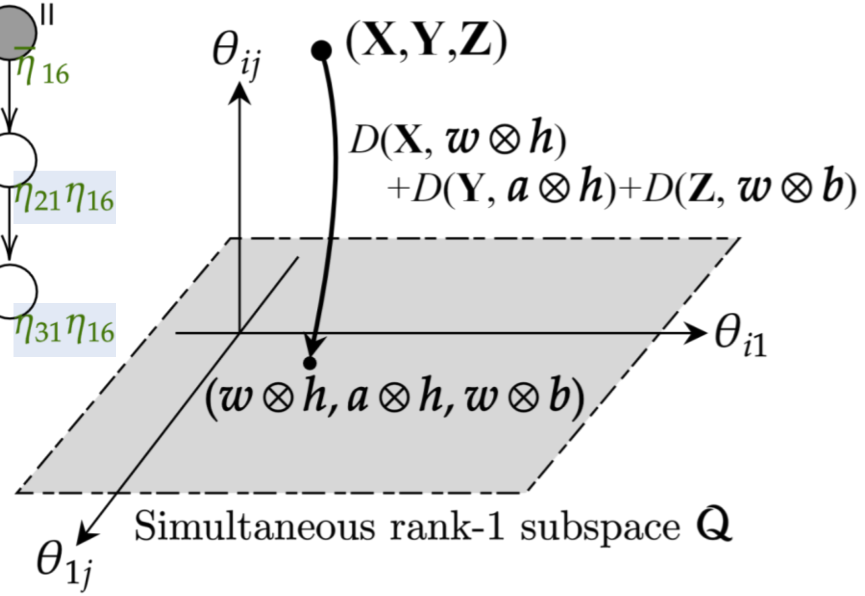
Simultaneous Rank-1 $\boldsymbol{\theta}$-condition

Its all two-body θ-parameters are 0.

Simultaneous Rank-1 $\boldsymbol{\eta}$-condition

$$\eta_{ij} = \eta_{i1}\eta_{1j}$$





η_{ij} : η-parameter before the projection.

$\bar{\eta}_{ij}$: η-parameter after the projection.

All $\boldsymbol{\eta}$-parameters after the projection are identified.

Information geometry of rank-1 NMMF

$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is *simultaneously rank-1 decomposable.* $\Leftrightarrow$ It can be written as $(\boldsymbol{w} \otimes \boldsymbol{h}, \boldsymbol{a} \otimes \boldsymbol{h}, \boldsymbol{w} \otimes \boldsymbol{b})$.

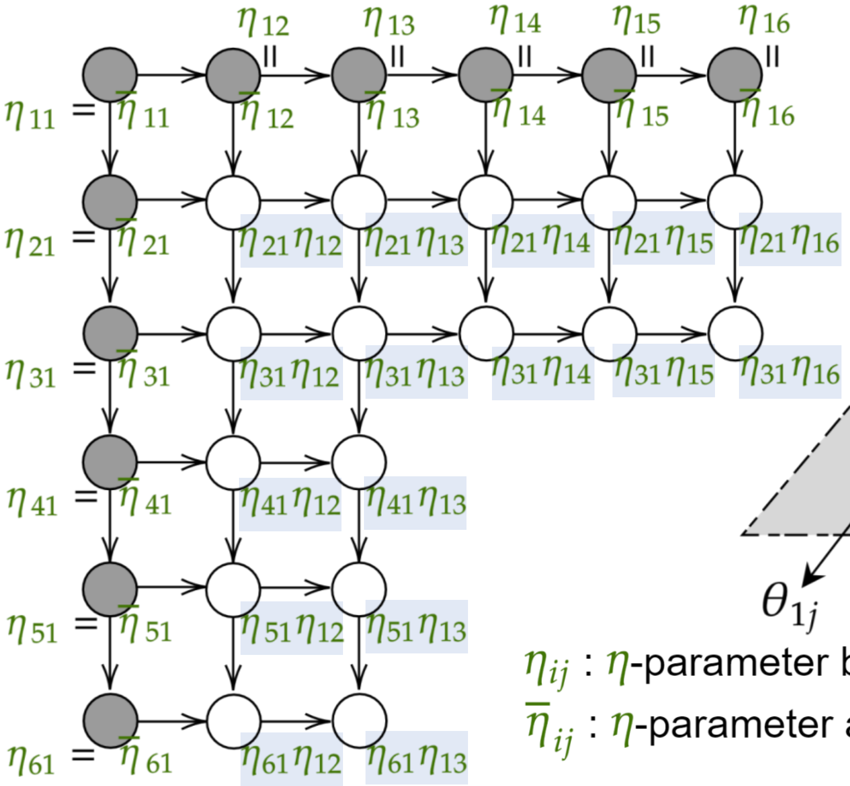
Simultaneous Rank-1 $\boldsymbol{\theta}$-condition

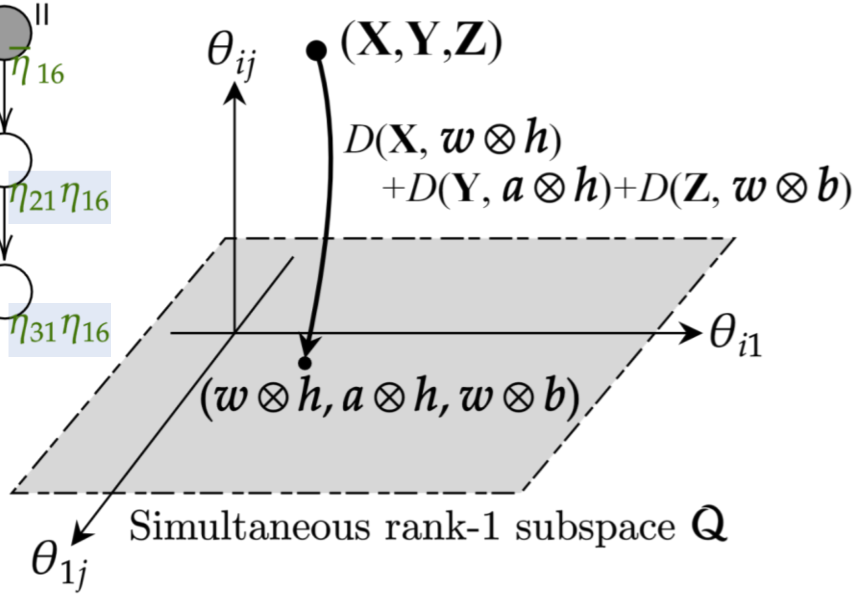
Its all two-body θ-parameters are 0.

Simultaneous Rank-1 $\boldsymbol{\eta}$-condition

$$\eta_{ij} = \eta_{i1}\eta_{1j}$$

● **One-body** parameter ○ **Two-body** parameter





η_{ij} : η-parameter before the projection.

$\bar\eta_{ij}$: η-parameter after the projection.

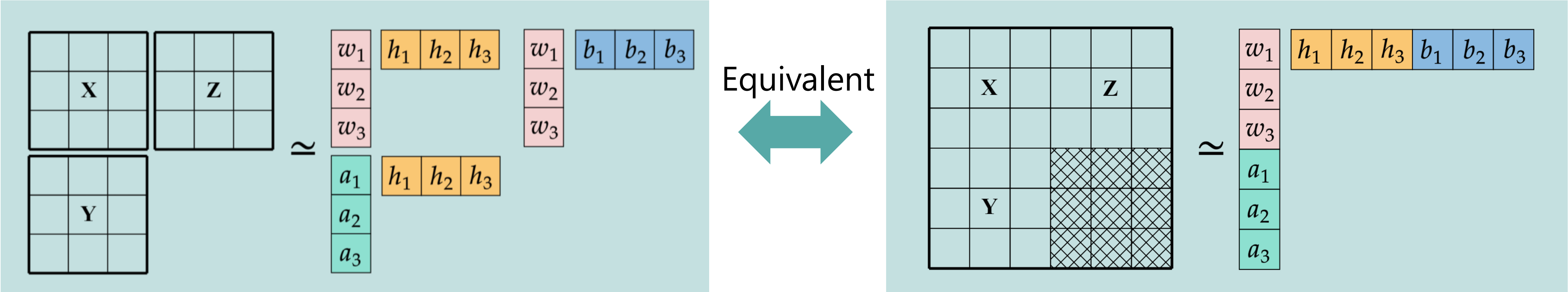
Möbius inversion formula

$$p_\eta(i,j) = \sum_{(k,l)} \mu_{ij}^{kl} \eta_{kl}$$

All $\boldsymbol{\eta}$-parameters after the projection are identified.
Using inversion formula, we found the projection destination.

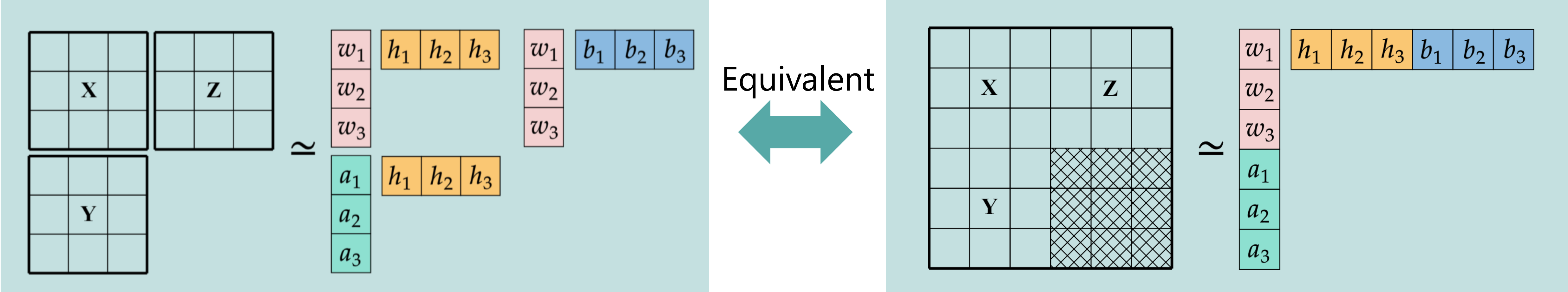
Rank-1 NMF with missing values

- NMMF can be viewed as a special case of NMF with missing values.

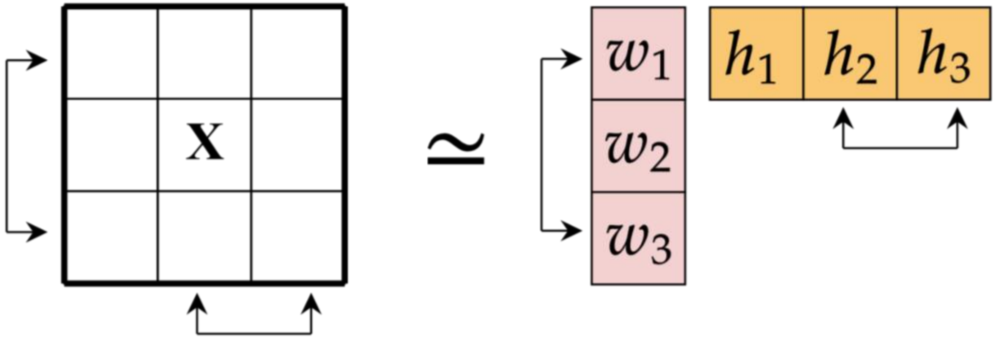


Rank-1 NMF with missing values

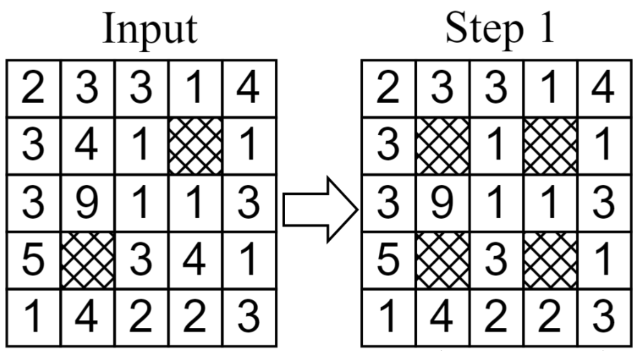
☐ NMMF can be viewed as a special case of NMF with missing values.



☐ NMF is homogeneous for row and column permutations

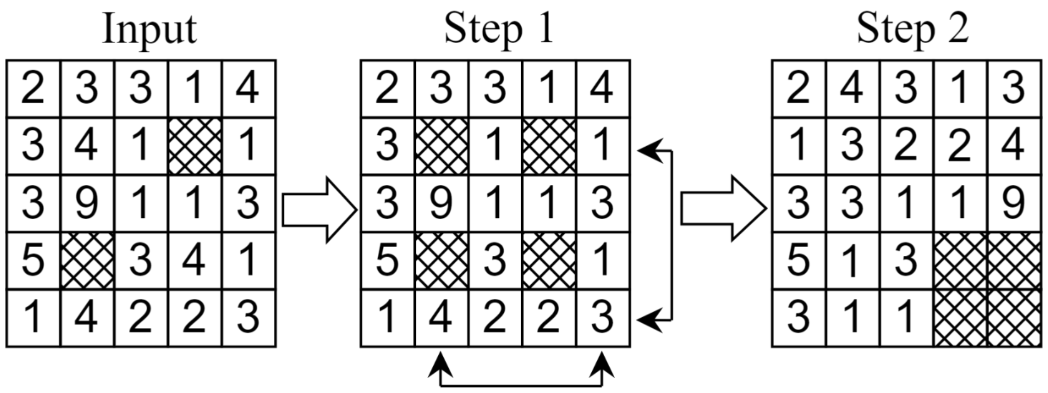


A1GM: Method



Step 1 : Increase the number of missing values.

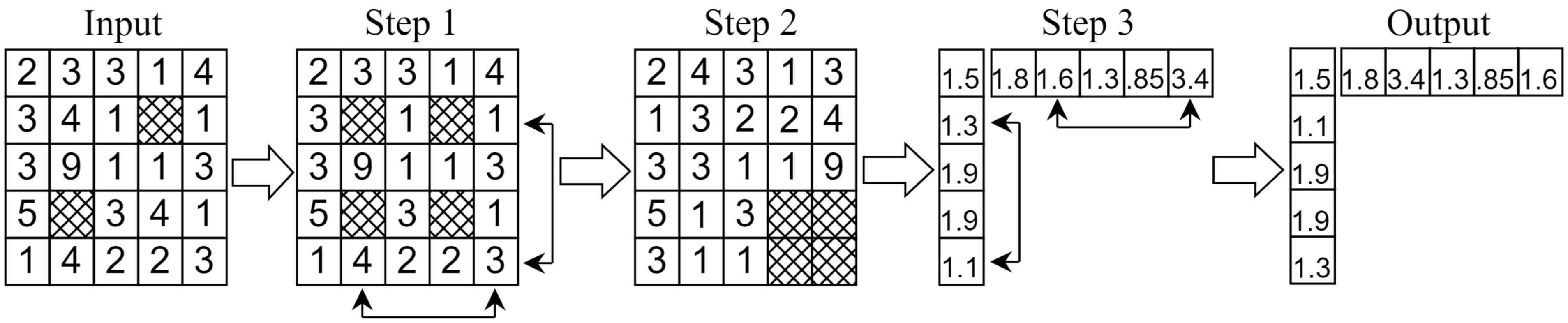
A1GM: Method



Step 1 : Increase the number of missing values.

Step 2 : Gather missing values in the bottom right.

A1GM: Method



Step 1 : Increase the number of missing values.

Step 2 : Gather missing values in the bottom right.

Step 3 : Use the formula of rank-1 NMMF and repermutate.

Experiments on real data

☐ A1GM is compared with gradient-based KL-WNMF

- Relative runtime < 1 means A1GM is faster than KL-WNMF.
- Relative error > 1 means worse reconstruction error of A1GM than KL-WNMF.
- Increase rate is the ratio of # missing values after addition of missing values at step1.

DataSet	size	# missing values	increase rate	relative error	relative runtime
Autompg	(398, 8)	6	1	1	0.12957
DailySunSpot	(73718, 9)	3247	1	1	0.12845
CaliforniaHousing	(20640, 9)	207	1	1	0.11821
MTSLibrary	(1533078, 4)	1247722	1	1	0.18327
BigMartSaleForecas	(8522, 5)	1463	1	1	0.12699
BoardGameGeekData	(101375, 17)	21	1	1	0.14625
CreditCardApproval	(590, 7)	25	1.92	1.0018	0.12212
HumanResourceAnaly	(14999, 7)	519	1.96146	1.0168	0.11858
heartdisease	(303, 14)	6	2	1	0.12259
lungcancer	(32, 57)	5	2	1.0001	0.13803
PerthHousePrice	(33656, 14)	16585	2.61345	1.0004	0.15382
SleepData	(62, 8)	12	2.75	1.0211	0.18208
arrhythmia	(452, 280)	408	4.70588	1.0148	0.11387
Bostonhousing	(506, 14)	120	5.6	1.003	0.1097
LifeExpectancyData	(2938, 19)	2563	7.04097	5.7983	0.095773
HCCSurvivalDataSet	(165, 50)	826	8.3632	3.2898	0.07113
wiki4HE	(913, 53)	1995	18.10175	1.2363	0.066256

Experiments on real data

- A1GM is compared with gradient-based KL-WNMF
  - Relative runtime < 1 means A1GM is faster than KL-WNMF.
  - Relative error > 1 means worse reconstruction error of A1GM than KL-WNMF.
  - Increase rate is the ratio of # missing values after addition of missing values at step1.

DataSet	size	# missing values	increase rate	relative error	relative runtime
Autompg	(398, 8)	6	1	1	0.12957
DailySunSpot	(73718, 9)	3247	1	1	0.12845
CaliforniaHousing	(20640, 9)	207	1	1	0.11821
MTSLibrary	(1533078, 4)	1247722	1	1	0.18327
BigMartSaleForecas	(8522, 5)	1463	1	1	0.12699
BoardGameGeekData	(101375, 17)	21	1	1	0.14625
CreditCardApproval	(590, 7)	25	1.92	1.0018	0.12212
HumanResourceAnaly	(14999, 7)	519	1.96146	1.0168	0.11858
heartdisease	(303, 14)	6	2	1	0.12259
lungcancer	(32, 57)	5	2	1.0001	0.13803
PerthHousePrice	(33656, 14)	16585	2.61345	1.0004	0.15382
SleepData	(62, 8)	12	2.75	1.0211	0.18208
arrhythmia	(452, 280)	408	4.70588	1.0148	0.11387
Bostonhousing	(506, 14)	120	5.6	1.003	0.1097
LifeExpectancyData	(2938, 19)	2563	7.04097	5.7983	0.095773
HCCSurvivalDataSet	(165, 50)	826	8.3632	3.2898	0.07113
wiki4HE	(913, 53)	1995	18.10175	1.2363	0.066256

Experiments on real data

☐ A1GM is compared with gradient-based KL-WNMF

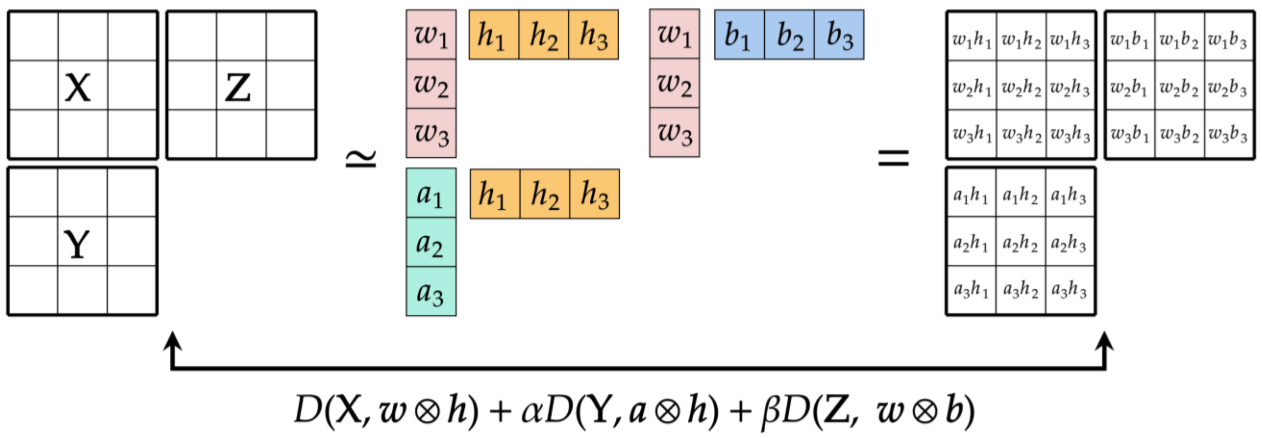
- Relative runtime < 1 means A1GM is faster than KL-WNMF.
- Relative error > 1 means worse reconstruction error of A1GM than KL-WNMF.
- Increase rate is the ratio of # missing values after addition of missing values at step1.

Much faster!

DataSet	size	# missing values	increase rate	relative error	relative runtime
Autompg	(398, 8)	6	1	1	0.12957
DailySunSpot	(73718, 9)	3247	1	1	0.12845
CaliforniaHousing	(20640, 9)	207	1	1	0.11821
MTSLibrary	(1533078, 4)	1247722	1	1	0.18327
BigMartSaleForecas	(8522, 5)	1463	1	1	0.12699
BoardGameGeekData	(101375, 17)	21	1	1	0.14625
CreditCardApproval	(590, 7)	25	1.92	1.0018	0.12212
HumanResourceAnaly	(14999, 7)	519	1.96146	1.0168	0.11858
heartdisease	(303, 14)	6	2	1	0.12259
lungcancer	(32, 57)	5	2	1.0001	0.13803
PerthHousePrice	(33656, 14)	16585	2.61345	1.0004	0.15382
SleepData	(62, 8)	12	2.75	1.0211	0.18208
arrhythmia	(452, 280)	408	4.70588	1.0148	0.11387
Bostonhousing	(506, 14)	120	5.6	1.003	0.1097
LifeExpectancyData	(2938, 19)	2563	7.04097	5.7983	0.095773
HCCSurvivalDataSet	(165, 50)	826	8.3632	3.2898	0.07113
wiki4HE	(913, 53)	1995	18.10175	1.2363	0.066256

Summary

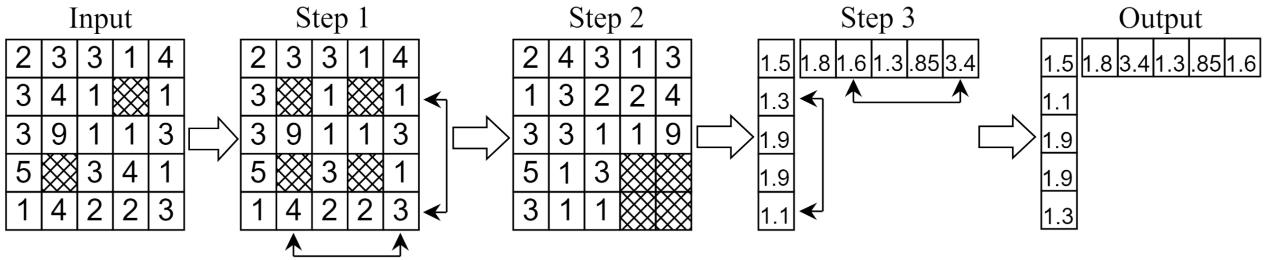
☐ Closed Formula of the Best Rank-1 NMMF



$$w_i = \frac{\sqrt{S(X)}}{S(X) + \beta S(Z)} \left\{ \sum_{j=1}^{J} X_{ij} + \beta \sum_{m=1}^{M} Z_{im} \right\} \qquad a_n = \frac{\sum_{j=1}^{J} Y_{nj}}{\sqrt{S(X)}}$$

$$h_j = \frac{\sqrt{S(X)}}{S(X) + \alpha S(Y)} \left\{ \sum_{i=1}^{I} X_{ij} + \alpha \sum_{n=1}^{N} Y_{nj} \right\} \qquad b_m = \frac{\sum_{i=1}^{I} Z_{im}}{\sqrt{S(X)}}$$

☐ A1GM: Faster Rank-1 NMF with missing values



Non-gradient-based method.

No worries about initial values, stopping criterion and learning rate.