Fast Rank-1 NMF for Missing Data with KL Divergence

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

[1]National Institute of Infomatics @ Tokyo, Japan [2]The Graduate University for Advanced Studies, SOKENDAI

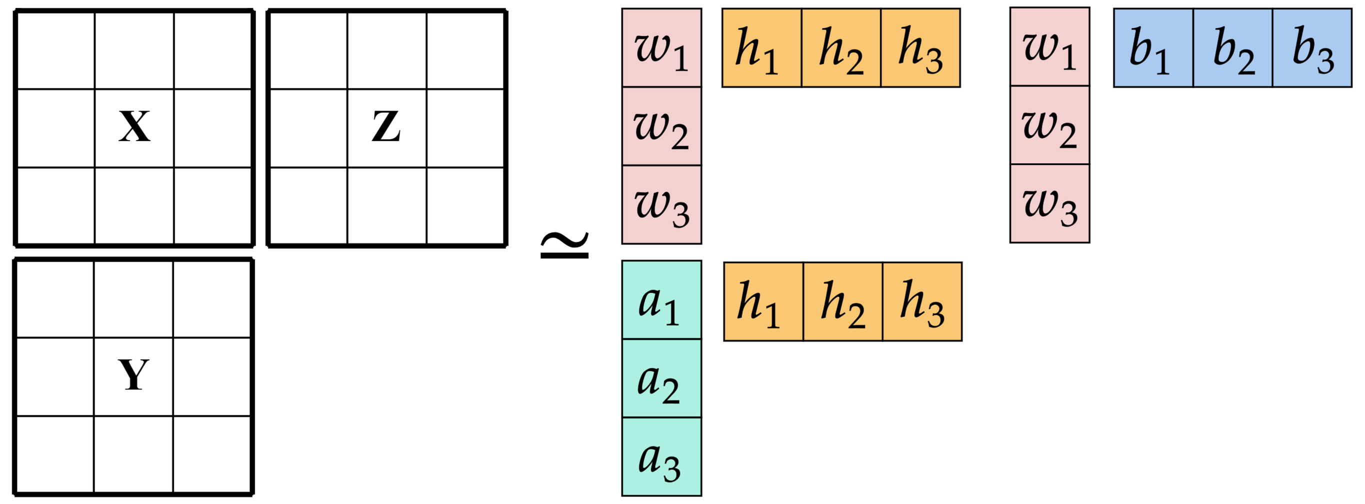
The 25th International Conference on Artificial Intelligence and Statistics (AISTATS 2022), March 28–30, 2022

Introduction

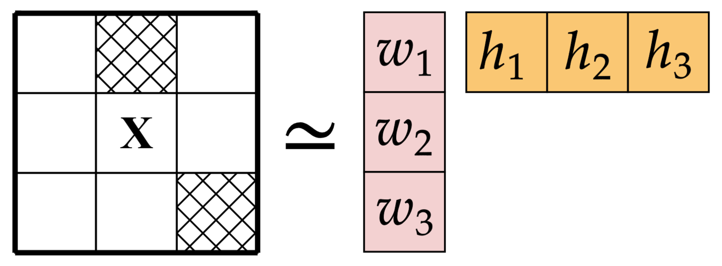
- We found **the formula of the best rank-1 factor sharing NMF** w.r.t. minimizing KL divergence.
- We developed a **non-gradient-based rank-1 NMF method with missing values** based on the formula.

Rank-1 Non-negative Multiple Matrix Factorization

For given $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$, rank-1 NMMF (non-negative multiple matrix factorization) finds $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{w}$ and $\boldsymbol{h}$ to minimize $D(\mathbf{X}, \boldsymbol{w} \otimes \boldsymbol{h}) + \alpha D(\mathbf{Y}, \boldsymbol{a} \otimes \boldsymbol{h}) + \beta D(\mathbf{Z}, \boldsymbol{w} \otimes \boldsymbol{b})$.



Rank-1 NMF with Missing Values



For given $\mathbf{X}$ with missing values, find $\boldsymbol{w}$ and $\boldsymbol{h}$ to minimize $D(\boldsymbol{\Phi} \circ \mathbf{X}, \boldsymbol{\Phi} \circ (\boldsymbol{w} \otimes \boldsymbol{h}))$.

Element-wise product

$$\Phi_{ij} = \begin{cases} 0 & \text{If } \mathbf{X}_{ij} \text{ is missing} \\ 1 & \text{otherwise} \end{cases}$$

Main Result

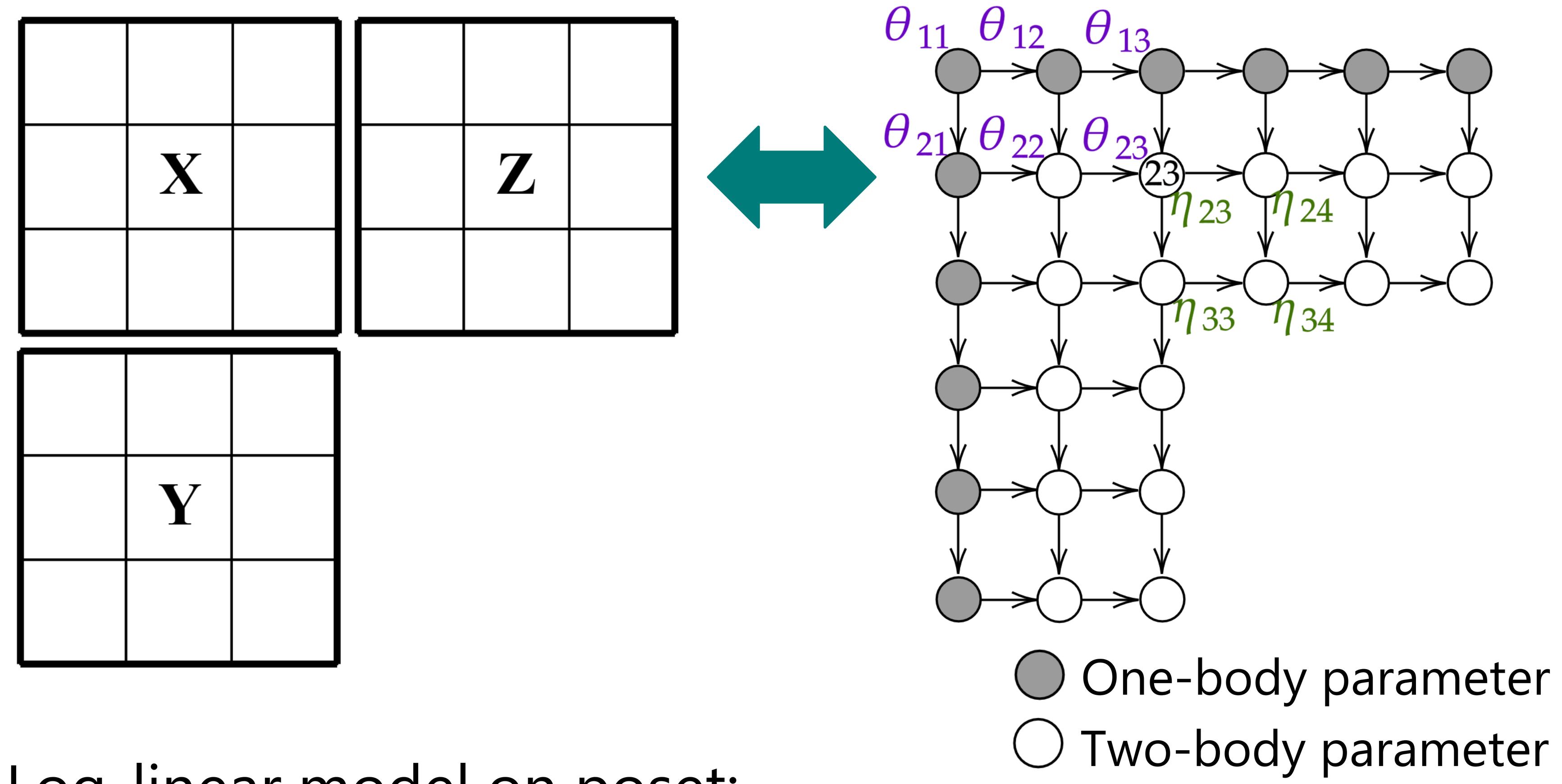
The best rank-1 NMMF of $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ is given as

$$w_i = \frac{\sqrt{S(\mathbf{X})}}{S(\mathbf{X}) + \beta S(\mathbf{Z})} \left\{ \sum_j \mathbf{X}_{ij} + \sum_m \beta \mathbf{Z}_{im} \right\}, \quad a_n = \frac{\sum_j \mathbf{Y}_{nj}}{\sqrt{S(\mathbf{X})}},$$

$$h_j = \frac{\sqrt{S(\mathbf{X})}}{S(\mathbf{X}) + \alpha S(\mathbf{Y})} \left\{ \sum_i \mathbf{X}_{ij} + \sum_n \alpha \mathbf{Y}_{nj} \right\}, \quad b_m = \frac{\sum_i \mathbf{Z}_{im}}{\sqrt{S(\mathbf{X})}}.$$

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

Theory



Log-linear model on poset:

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

Index set of the poset is the sample space.

Information Geometry of Rank-1 NMMF

$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ is *simultaneously rank-1 decomposable.*

⇔ It can be written as $(\boldsymbol{w} \otimes \boldsymbol{h}, \boldsymbol{a} \otimes \boldsymbol{h}, \boldsymbol{w} \otimes \boldsymbol{b})$.

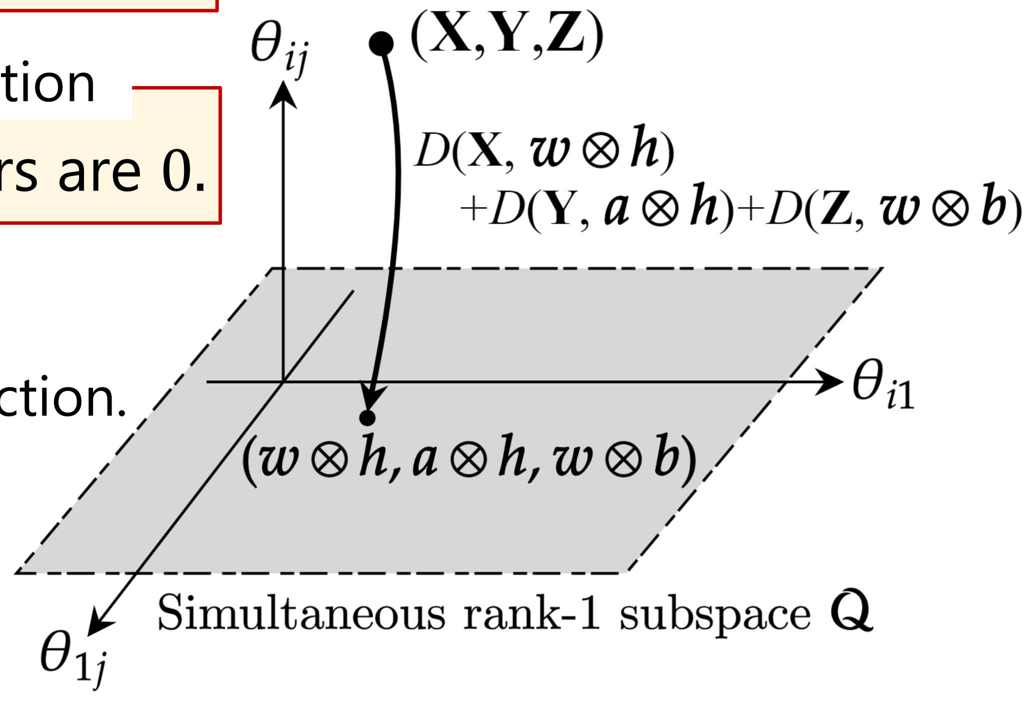
Simultaneous Rank-1 $\boldsymbol{\eta}$-condition

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

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

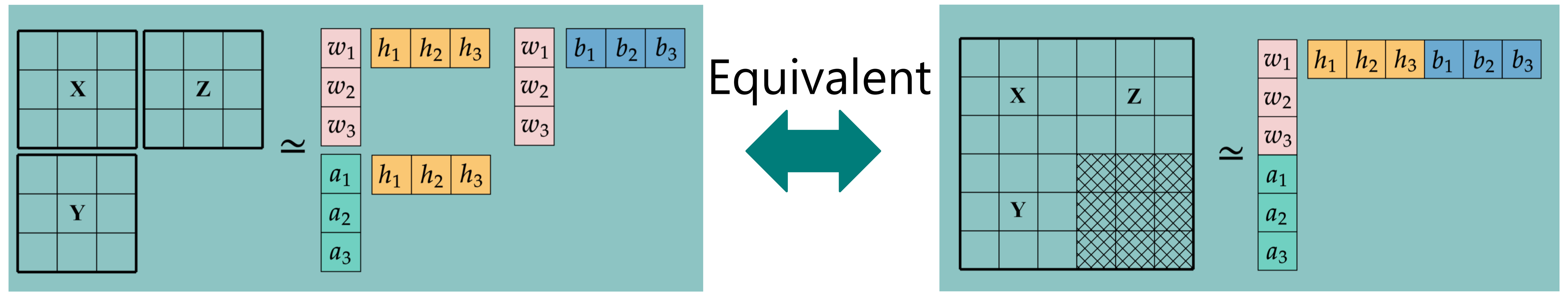
Its all two-body θ-parameters are 0.

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



Faster Rank-1 NMF with Missing Values

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



A1GM: Proposed Method for Rank-1 NMF with Missing Values

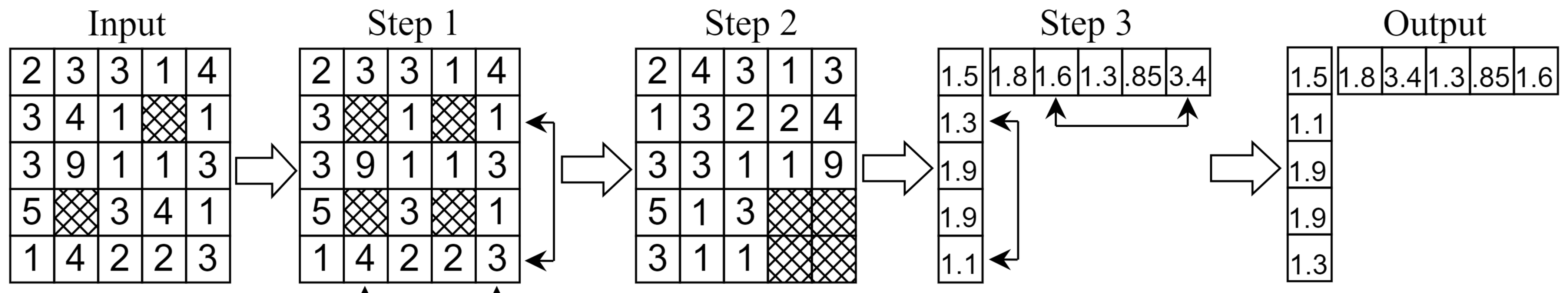
A1GM is a non-gradient-based method.

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

Step1: Increase the number of missing values.

Step2: Gather missing values in the bottom right.

Step3: Use the formula of rank-1 NMMF and repermutate.



Experiments on Real Data

Our method 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.

DataSet	size	# missing values	increase rate	relative error	relative runtime
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
CreditCardApproval	(590, 7)	25	1.92	1.0018	0.12212
HumanResourceAnaly	(14999, 7)	519	1.96146	1.0168	0.11858
PerthHousePrice	(33656, 14)	16585	2.61345	1.0004	0.15382
SleepData	(62, 8)	12	2.75	1.0211	0.18208
Bostonhousing	(506, 14)	120	5.6	1.003	0.1097

Increase rate is the ratio of # missing values after addition of missing values at step1.

Much faster!!