1-Q-2

ディリクレ分布に基づく正則化付き非負値行列因子分解

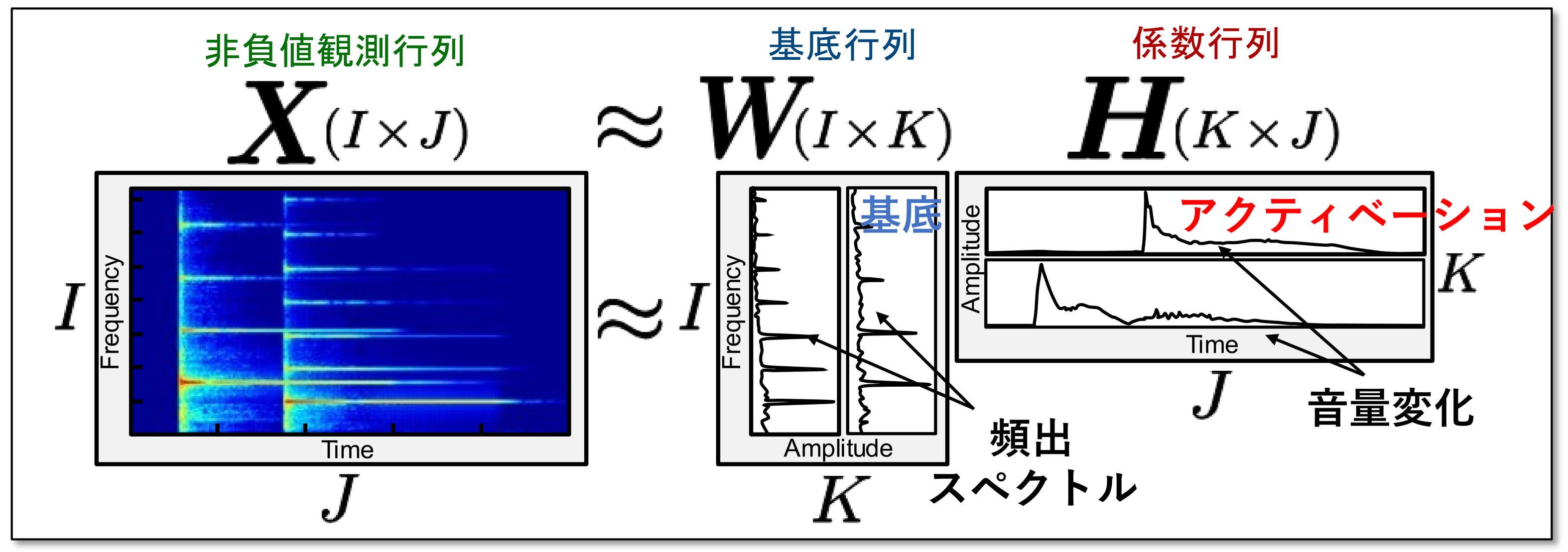
☆小川遼，北村大地，綾野翔馬（香川高専）

本発表の概要

- 非負値行列因子分解（nonnegative matrix factorization: NMF）は様々なタスクに活用可能
- 事前モデルを導入し最適化を誘導する正則化付きNMFが各種提案されてきた
  - L1ノルム正則化によるスパース性の誘導（スパースNMF）
  - 連続性正則化によるスムース性の誘導（スムースNMF）
- **スパース性とスムース性を統一的に表現できる新しい正則化付きNMF**を提案
  - ➡ **ディリクレ事前分布を仮定した最大事後確率推定によるNMF（ディリクレNMF）**
- **制約条件付き補助関数法**を適用し反復更新式を導出
- 乱数を用いた数値実験ではスパース・スムースな基底を完全に推定

1. 研究背景

- **非負値行列因子分解（nonnegative matrix factorization: NMF）** [Lee+, 1999]
  - 観測非負行列を別の2つの非負行列の行列積で低ランク近似する手法
  - 音響信号では時間周波数のモデリングに頻繁に用いられる



2. 正則化付きNMF

- **スパースNMF** [Hoyer, 2004], [Le Roux+, 2015], [Marmin+, 2023]等
  - NMFの変数行列をスパースに誘導する**L1ノルム正則化**付きNMF

$$\underset{\boldsymbol{W},\boldsymbol{H}}{\text{Minimize}}\ \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H})+\mu\|\boldsymbol{W}\|_1\ \text{s.t.}\ w_{i,k},h_{k,j}\geq 0\ \forall i,j,k,\ \|\boldsymbol{w}_k\|_1=1\ \forall k$$

- **スムースNMF** [Virtanen+, 2007]等
  - NMFの変数行列の列や行をスムースに誘導する**隣接差分正則化**付きNMF

$$\underset{\boldsymbol{W},\boldsymbol{H}}{\text{Minimize}}\ \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H})+\mu\sum_{i,k}|w_{i,k}-w_{i-1,k}|^2\ \text{s.t.}\ w_{i,k},h_{k,j}\geq 0\ \forall i,j,k,\ \|\boldsymbol{w}_k\|=1\ \forall k$$

スパース性とスムース性の両方を統一的に扱える正則化は無かった
NMFの$\boldsymbol{W}$と$\boldsymbol{H}$間のスケール不定性の存在は正則化において要注意

- **事前分布の最大事後確率推定（MAP推定）に基づく正則化付きNMF**
  - 基底行列$\boldsymbol{W}$が事前分布$p(\boldsymbol{W})$に従うと仮定するMAP推定NMF

$$\underset{\boldsymbol{W},\boldsymbol{H}}{\text{Minimize}}\ \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H})+\mathcal{R}(\boldsymbol{W})\quad \text{s.t.}\ w_{i,k},h_{k,j}\geq 0\ \forall i,j,k$$

  - $\mathcal{R}(\boldsymbol{W})=-\log p(\boldsymbol{W})$は事前分布由来の正則化項

新規性 **ディリクレ分布をNMFの事前分布に導入**
スパース・スムースを統一的に扱いつつスケール不定性を回避

3. 提案手法

- **ディリクレ分布**
  - **非負制約**と**ノルム制約**を満たす確率変数ベクトル（標準単体上のベクトル）の分布
  - 標準単体の定義
  - ディリクレ分布の確率密度関数

$$\mathcal{S}_{I-1}=\left\{\mathbf{x}\in\mathbb{R}^I \mid \underbrace{\mathrm{x}_i\geq 0\ \forall i}_{\text{非負制約}},\ \underbrace{\sum_i \mathrm{x}_i=1}_{\text{ノルム制約}}\right\}$$

$$\mathbf{x}\sim p(\mathbf{x};\boldsymbol{\alpha})=\frac{1}{B(\boldsymbol{\alpha})}\prod_i \mathrm{x}_i^{\alpha_i-1}$$

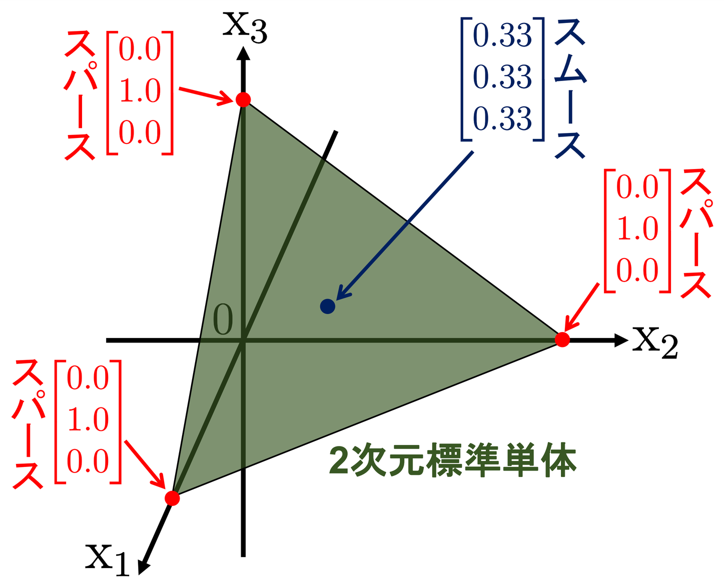
確率変数ベクトル　パラメータベクトル　多変量ベータ関数

$$B(\boldsymbol{\alpha})=\frac{\prod_i\Gamma(\alpha_i)}{\Gamma\left(\sum_i\alpha_i\right)}$$

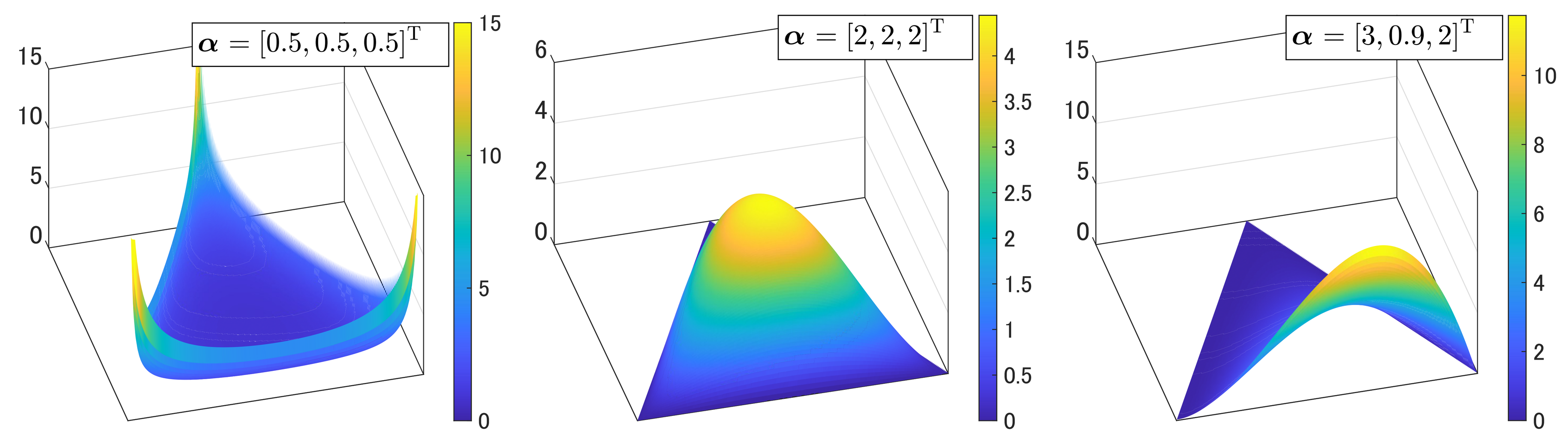
- **3次元確率変数ベクトルの場合**
  - 2次元標準単体上の確率変数ベクトル（3変数だがノルム制約より自由度は2）

3次元確率変数ベクトル $\mathbf{x}=\begin{bmatrix}\mathrm{x}_1\\ \mathrm{x}_2\\ \mathrm{x}_3\end{bmatrix}$　ノルム制約: $\mathrm{x}_1+\mathrm{x}_2+\mathrm{x}_3=1$　非負制約: $\mathrm{x}_1,\mathrm{x}_2,\mathrm{x}_3\geq 0$

例えば… $\mathbf{x}=\begin{bmatrix}0.2\\0.3\\0.5\end{bmatrix},\ \mathbf{x}=\begin{bmatrix}0.1\\0.1\\0.8\end{bmatrix},\ \mathbf{x}=\begin{bmatrix}0.0\\0.9\\0.1\end{bmatrix}$ 等



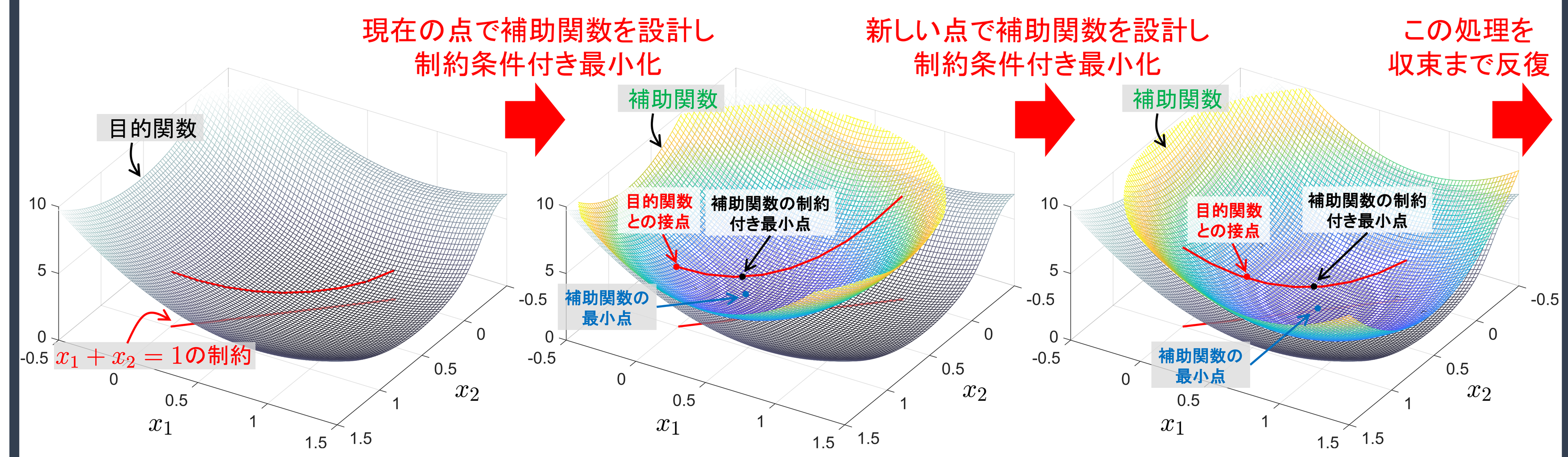
  - パラメータベクトル$\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\alpha_3]^{\mathrm{T}}$は**確率密度の各頂点への集中度**に対応
  - ➡ スパースやスムースなベクトルの生成確率密度を$\boldsymbol{\alpha}$で制御可能



- **ディリクレNMFの目的関数と最適化の方針**
  - ディリクレ分布から基底行列の各列（基底ベクトル）が独立に生成されると仮定

$$\underset{\boldsymbol{W},\boldsymbol{H}}{\text{Minimize}}\ \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H})+\sum_{i,k}(\alpha_{ik}-1)\log w_{ik}^{-1}\ \text{s.t.}\ w_{ik},h_{kj}\geq 0\ \forall i,j,k,\quad \sum_i w_{ik}=1\ \forall k$$

  - NMFでは常套手段の**補助関数法**を用いるが，ディリクレ分布由来の**ノルム制約**を陽に扱って解く必要がある ➡ **補助関数法の上限関数最小化に制約条件を付与**
- **制約条件付き補助関数法**
  - 「本来の目的関数に対する最適化問題の制約条件」を「補助関数の最小化問題」にも引き継いで考慮しながら最小化



- **補助関数の非負制約・ノルム制約条件付き最適化問題**

$$\underset{\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta}}{\text{Minimize}}\ \mathcal{J}^+\ \text{s.t.}\ w_{ik},h_{kj}\geq\ 0\ \forall i,j,k,\quad \sum_i w_{ik}=1\ \forall k$$

  - 補助関数の最小化問題に対して**非負制約**と**ノルム制約**の両方を考慮
  - **一般化KL擬距離に対する補助関数**と**ディリクレ分布由来の正則化項（そのまま）**

$$\mathcal{J}^+(\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta})=\sum_{i,j}\left[x_{ij}\log x_{ij}-x_{ij}\sum_k\delta_{i,j,k}\log\frac{w_{i,k}h_{k,j}}{\delta_{i,j,k}}-x_{ij}+\sum_k w_{i,k}h_{k,j}\right]+\sum_{i,k}(\alpha_{i,k}-1)\log w_{i,k}$$

- **ディリクレNMFの反復更新式**

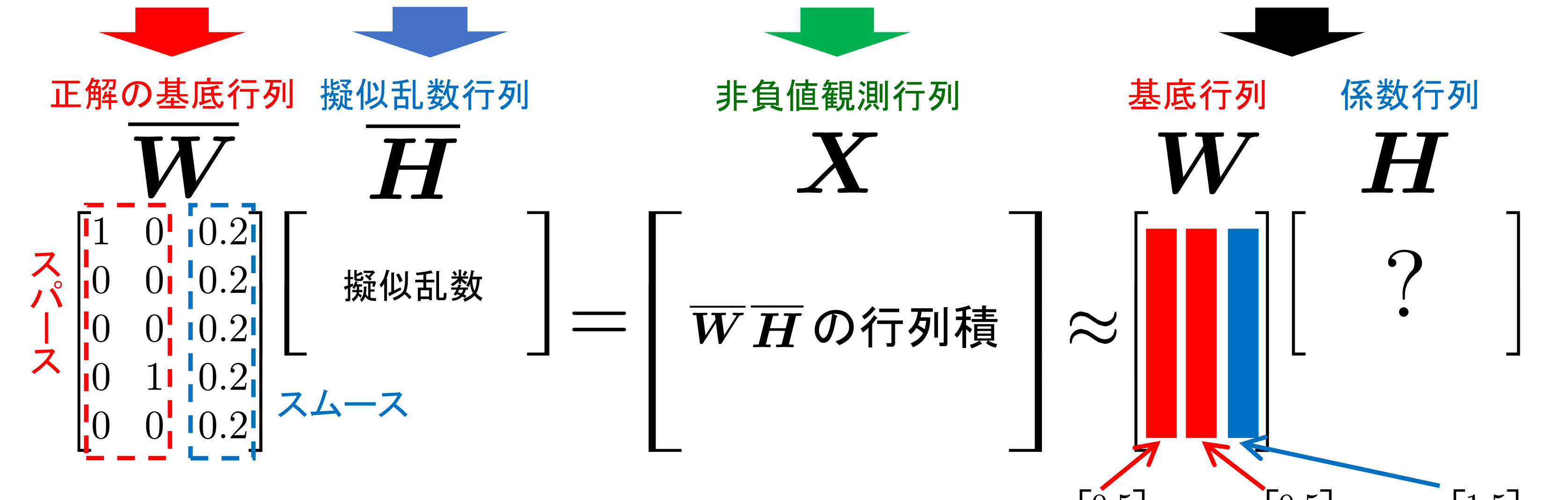
$$w_{ik}=\begin{cases}\hat{w}_{ik} & (\text{if}\ \hat{w}_{ik}\geq 0)\\ 0 & (\text{otherwise})\end{cases}\qquad \hat{w}_{ik}=\frac{w_{ik}\sum_j\frac{x_{ij}}{\sum_{k'}w_{ik'}h_{k'j}}h_{kj}+\alpha_{ik}-1}{\sum_{i\in\mathcal{I}_k}\left(w_{ik}\sum_j\frac{x_{ij}}{\sum_{k'}w_{ik'}h_{k'j}}h_{kj}+\alpha_{ik}-1\right)}$$

補助関数の最適化問題に対するラグランジアン$\mathcal{L}$を考えたときに，$\partial\mathcal{L}/\partial w_{ik}=0$を満たす$w_{ik}$が非負となるインデクス$i$の集合（$k$毎にこの集合を定義している）

4. 乱数を用いた数値実験

- **実験目的**
  - ディリクレNMFの正則化の効果及び導出の正当性の確認実験
- **実験条件**

① 特徴的な基底を持つ正解の基底行列$\boldsymbol{W}$を作成
② 非負の乱数から擬似乱数行列$\boldsymbol{H}$を作成
③ $\boldsymbol{W}$と$\boldsymbol{H}$の行列積から非負観測行列$\boldsymbol{X}$を作成
④ ディリクレNMFと正則化なしNMFを適用



  - ディリクレNMFと正則化なしNMFにより$\boldsymbol{W}$と$\boldsymbol{H}$を推定
  - 得られた$\boldsymbol{W}$と$\boldsymbol{H}$が正解の基底行列をどれほど精度よく推定出来ているかを比較
  - ディリクレ分布のパラメータは**スパース**・**スムース**度合に応じて右図のように決定

$$\alpha_1=\begin{bmatrix}0.5\\0.5\\0.5\\0.5\\0.5\end{bmatrix}\quad \alpha_2=\begin{bmatrix}0.5\\0.5\\0.5\\0.5\\0.5\end{bmatrix}\quad \alpha_3=\begin{bmatrix}1.5\\1.5\\1.5\\1.5\\1.5\end{bmatrix}$$

スパースに誘導するパラメータ　スパースに誘導するパラメータ　スムースに誘導するパラメータ

- **実験結果**

正解の基底行列及び係数行列

Basis mat. (oracle)

	1	2	3
1	0	1	0.2
2	0	0	0.2
3	1	0	0.2
4	0	0	0.2
5	0	0	0.2

Coef. mat. (oracle)

	1	2	3	4	5	6	7	8	9	10
1	0.417	0.3023	0.1863	0.5388	0.2045	0.6705	0.1404	0.9683	0.8764	0.03905
2	0.7203	0.1468	0.3456	0.4192	0.8781	0.4173	0.1981	0.3134	0.8946	0.1698
3	0.0001144	0.09234	0.3968	0.6852	0.02739	0.5587	0.8007	0.6923	0.08504	0.8781

正則化なしNMFで得られた基底行列及び係数行列

Basis mat. (simple NMF)

	1	2	3
1	0.02527	0.7875	0.3726
2	0.02097	1.424e−14	0.1805
3	0.9351	0.2203	0.01564
4	0.02097	1.16e−15	0.1805
5	0.02097	7.385e−15	0.1805

Coef. mat. (simple NMF)

	1	2	3	4	5	6	7	8	9	10
1	0.1908	0.3024	0.1971	0.6257	2.22e−16	0.7339	0.3024	1.117	0.6972	0.2141
2	0.935	0.1694	0.3375	0.3636	1.073	0.3982	0.01888	0.2418	1.134	0.0002197
3	2.22e−16	0.06597	0.4161	0.6848	0.03104	0.5316	0.8722	0.6323	1.885e−07	0.9334

ディリクレNMFで得られた基底行列及び係数行列

Basis mat. (Dirichlet NMF)

	1	2	3
1	0	1	0.2
2	0	0	0.2
3	1	0	0.2
4	0	0	0.2
5	0	0	0.2

Coef. mat. (Dirichlet NMF)

	1	2	3	4	5	6	7	8	9	10
1	0.417	0.3023	0.1863	0.5388	0.2045	0.6705	0.1404	0.9683	0.8764	0.03904
2	0.7203	0.1468	0.3456	0.4192	0.8781	0.4173	0.1981	0.3134	0.8946	0.1698
3	0.0001144	0.09234	0.3968	0.6852	0.02739	0.5587	0.8008	0.6923	0.08504	0.8782

5. まとめ

- ディリクレ分布を事前分布に仮定した正則化付きNMFの導出を行った
- 提案手法の効果と導出の正当性を実験的に確認した

原稿の誤りに関する訂正

訂正①

集合である．ラグランジアンは次式となる．

$$\mathcal{L} = \mathcal{J}^{+} + \sum_k \lambda_k \left(\sum_i w_{ik} - 1 \right) \tag{12}$$

削除

ここで，λ_k ~~≥ 0~~ はラグランジュ乗数を表す．非負制約

訂正②

訂正前

以後，$\hat{w}_{ik}$ を求める．$\partial \mathcal{L} / \partial w_{ik} = 0$ より

$$\sum_j \left(-x_{ij} \frac{\delta_{ijk}}{\hat{w}_{ik}} + h_{kj} \right) - (\alpha_{ik} - 1) \frac{1}{\hat{w}_{ik}} + \lambda_k = 0$$

を得る．これを整理して，$\hat{w}_{ik}$ は次式となる．

$$\hat{w}_{ik} = \frac{\sum_j x_{ij} \delta_{ijk} + \alpha_{ik} - 1}{\sum_j h_{kj} + \lambda_k} \tag{17}$$

また，式 (15) より $\sum_i w_{ik} - 1 = 0$ を得るが，これはノルム制約 $\sum_i w_{ik} = 1\ \forall k$ と等価である．このノルム制約に式 (16) を代入して，次式を得る．

$$\sum_i \max(\hat{w}_{ik}, 0) = 1 \tag{18}$$

ここで，$\max(\cdot, \cdot)$ は入力値の大きい方を返す．式 (17) を式 (18) に代入すると，$\hat{w}_{ik}$ の分母は非負値のため

$$\sum_j h_{kj} + \lambda_k = \sum_i \max \left(\sum_j x_{ij} \delta_{ijk} + \alpha_{ik} - 1, 0 \right) \tag{19}$$

が得られる．式 (19) を式 (17) の分母に代入してラグランジュ乗数 λ_k を消去すると，次式となる．

$$\hat{w}_{ik} = \frac{\sum_j x_{ij} \delta_{ijk} + \alpha_{ik} - 1}{\sum_i \max \left(\sum_j x_{ij} \delta_{ijk} + \alpha_{ik} - 1, 0 \right)} \tag{20}$$

訂正後

以後，$\hat{w}_{ik}$ を求める．$\partial \mathcal{L} / \partial w_{ik} = 0$ より

$$\sum_j \left(-x_{ij} \frac{\delta_{ijk}}{\hat{w}_{ik}} + h_{kj} \right) - (\alpha_{ik} - 1) \frac{1}{\hat{w}_{ik}} + \lambda_k = 0$$

を得る．これを整理して，$\hat{w}_{ik}$ は次式となる．

$$\hat{w}_{ik} = \frac{\sum_j x_{ij} \delta_{ijk} + \alpha_{ik} - 1}{\sum_j h_{kj} + \lambda_k} \tag{17}$$

式 (16) を考慮すると，式 (17) で求まる $\hat{w}_{ik}$ が非負になる要素のみ着目すればよいため，$\hat{w}_{ik} \geq 0$ を満たすインデクス i の集合を k 毎に $\mathcal{I}_k$ とおく．また，式 (15) より $\sum_i w_{ik} - 1 = 0$ を得るが，これはノルム制約と等価であり，次式に書き換えられる．

$$\sum_i w_{ik} = \sum_{i \in \mathcal{I}_k} \hat{w}_{ik} = 1\ \forall k \tag{18}$$

式 (17) を式 (18) に代入して整理すると

$$\sum_j h_{kj} + \lambda_k = \sum_{i \in \mathcal{I}_k} \left(\sum_j x_{ij} \delta_{ijk} + \alpha_{ik} - 1 \right) \tag{19}$$

が得られる．式 (19) を式 (17) の分母に代入してラグランジュ乗数 λ_k を消去すると，次式となる．

$$\hat{w}_{ik} = \frac{\sum_j x_{ij} \delta_{ijk} + \alpha_{ik} - 1}{\sum_{i \in \mathcal{I}_k} \left(\sum_j x_{ij} \delta_{ijk} + \alpha_{ik} - 1 \right)}\ \forall i \in \mathcal{N}_k, \forall k \tag{20}$$