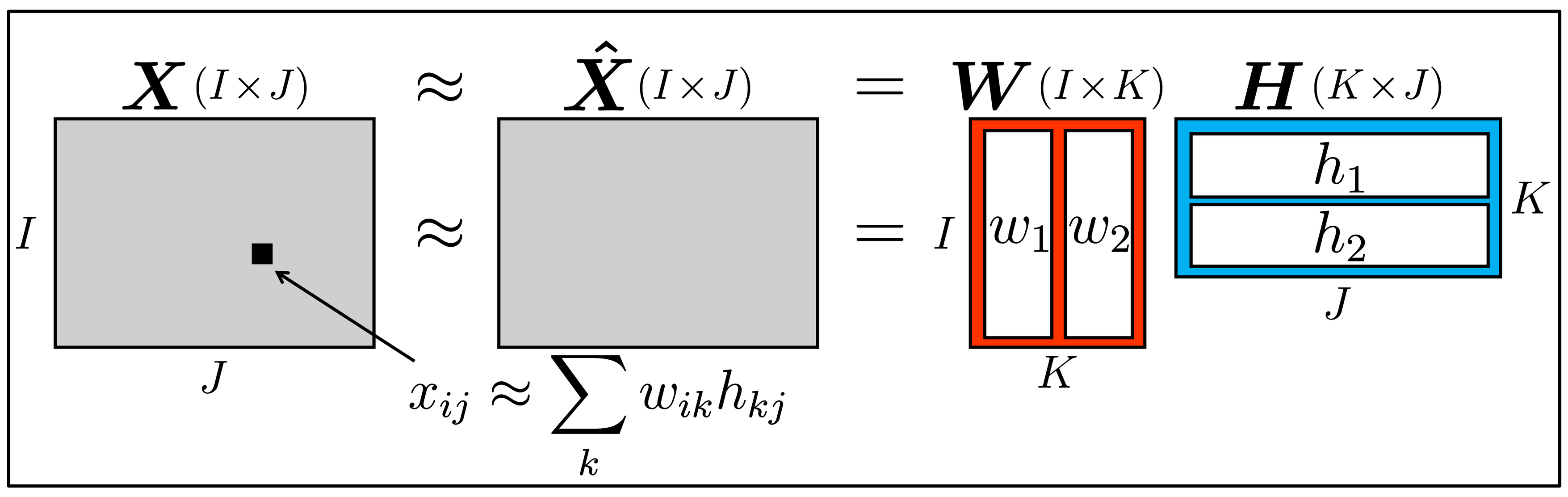
229

Nonnegative Matrix Factorization Using Dirichlet-Distribution-Based Regularization

Haru Ogawa, Daichi Kitamura, Shoma Ayano (NIT Kagawa, Japan)

1. Research Background

- **Nonnegative matrix factorization: NMF** [Lee+, 1999]
  - A method to approximate a nonnegative matrix by the product of two low-rank nonnegative matrices.
  - NMF has been widely applied to audio, image, and biomedical signal processing.



2. Regularized NMF

- **Sparse NMF** [Hoyer, 2004], [Le Roux+, 2015], [Marmin+, 2023], etc.
  - **L1-norm** regularized NMF that induces sparsity in the factor matrices.

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

- **Smooth NMF** [Virtanen+, 2007], etc.
  - Inducing smoothness via **adjacency-difference regularization**.

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

No existing regularization has unified sparsity and smoothness.
Scale indeterminacy between W and H requires caution.

- **MAP-based regularized NMF with a prior distribution**
  - MAP estimation assuming prior $p(\boldsymbol{W})$ on $\boldsymbol{W}$.

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

  - $\mathcal{R}(\boldsymbol{W}) = -\log p(\boldsymbol{W})$: **Regularization from prior distribution**

Novelty
Introducing the Dirichlet distribution as a prior in NMF
Unifying sparsity and smoothness, avoiding scale indeterminacy

3. Proposed Method

- **Dirichlet Distribution**
  - A distribution over random vectors satisfying **non-negativity** and **norm** constraints (vectors on the standard simplex).
  - Standard Simplex
  - Dirichlet Distribution (PDF)

$$\mathcal{S}_{I-1} = \left\{\mathrm{x} \in \mathbb{R}^I \,\middle|\, \mathrm{x}_i \geq 0\ \forall i,\ \sum_i \mathrm{x}_i = 1\right\}$$

Non-negativity　Norm Constraint

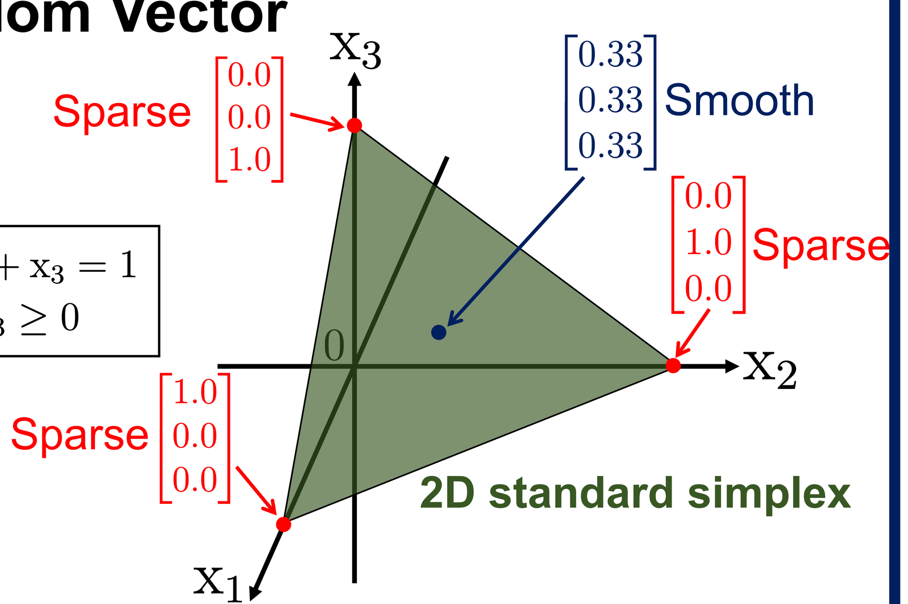
$$\mathrm{x} \sim p(\mathrm{x};\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})}\prod_i \mathrm{x}_i^{\alpha_i - 1} \qquad B(\boldsymbol{\alpha}) = \frac{\prod_i \Gamma(\alpha_i)}{\Gamma\left(\sum_i \alpha_i\right)}$$

Random Variable Vector　Parameter Vector　Multivariate Beta Function

- **Case of a 3-Dimensional Random Vector**
  - Random Vector on the 2D Standard Simplex

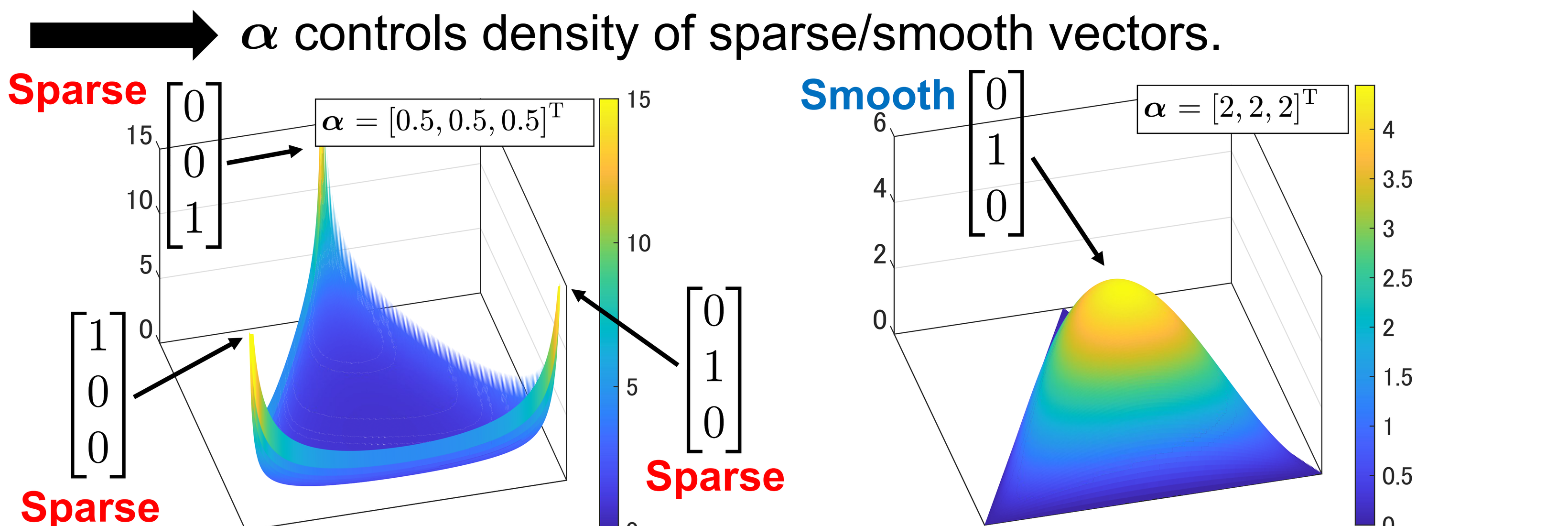
3D Random Vector $\mathrm{x} = \begin{bmatrix}\mathrm{x}_1\\ \mathrm{x}_2\\ \mathrm{x}_3\end{bmatrix}$　Norm Constraint: $\mathrm{x}_1 + \mathrm{x}_2 + \mathrm{x}_3 = 1$　Non-negativity: $\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3 \geq 0$

Example⋯ $\mathrm{x} = \begin{bmatrix}0.2\\0.3\\0.5\end{bmatrix}$, $\mathrm{x} = \begin{bmatrix}0.1\\0.1\\0.8\end{bmatrix}$, $\mathrm{x} = \begin{bmatrix}0.0\\0.9\\0.1\end{bmatrix}$ etc.



- $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3]^{\mathrm{T}}$ controls **concentration toward vertices**.

➡ $\boldsymbol{\alpha}$ controls density of sparse/smooth vectors.



- **Objective function and optimization policy of Dirichlet NMF**
  - Assume that each column of the basis matrix (basis vector) is independently generated from the Dirichlet distribution.

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

  - NMF uses the auxiliary function method, but must handle the **norm constraint**. ➡ **Add constraints to the upper bound minimization.**

- **Auxiliary function optimization with non-negativity and norm constraints**

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

  - Auxiliary minimization with **non-negativity** & **norm constraints**
  - **For generalized KL divergence** + **Dirichlet prior regularization**

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

- **Iterative update rules of Dirichlet NMF**

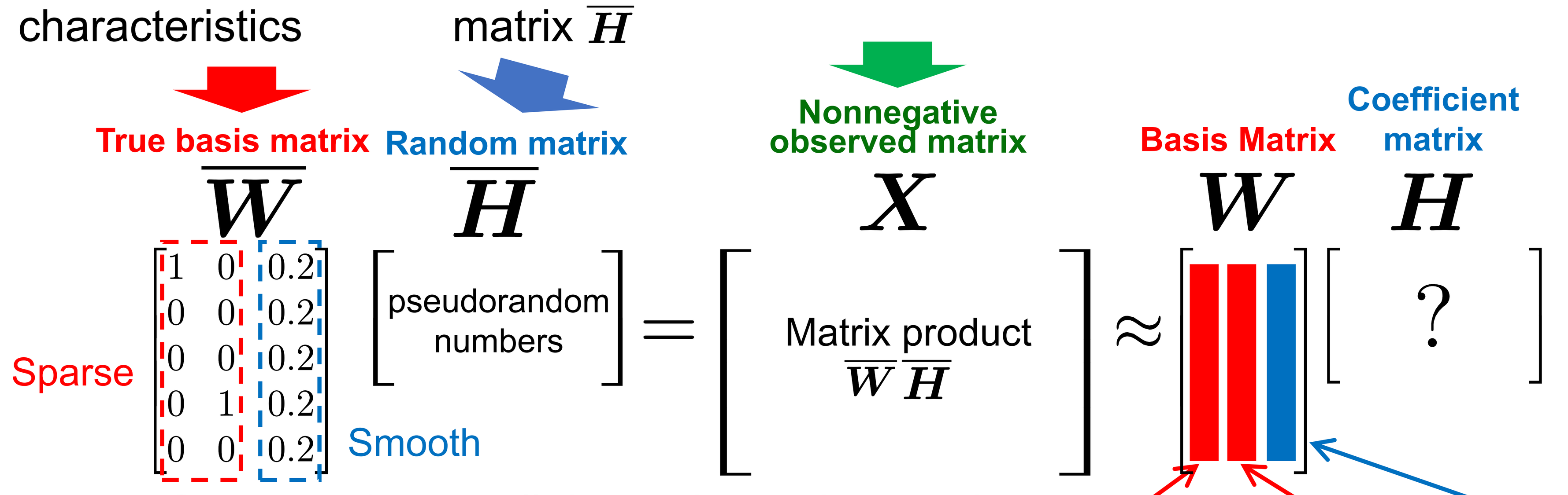
$$w_{ik} = \begin{cases}\hat{w}_{ik} & (\text{if } \hat{w}_{ik} \geq 0)\\ 0 & (\text{otherwise})\end{cases} \qquad \hat{w}_{ik} = \frac{w_{ik}\sum_j \frac{x_{ij}}{\sum_{k'} w_{ik'}h_{k'j}}h_{kj} + \alpha_{ik} - 1}{\sum_{i\in\mathcal{I}_k}\left(w_{ik}\sum_j \frac{x_{ij}}{\sum_{k'} w_{ik'}h_{k'j}}h_{kj}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

For the Lagrangian $\mathcal{L}$, the stationary point $\partial\mathcal{L}/\partial w_{ik} = 0$ gives non-negative w_{ik} within the index set i (defined for each k).

4. Numerical Experiment

- **Conditions**

①Create the true basis matrix $\overline{\boldsymbol{W}}$ with sparse and smooth characteristics
②Generate random nonnegative matrix $\overline{\boldsymbol{H}}$
③Construct the observed matrix $\boldsymbol{X} = \overline{\boldsymbol{W}}\,\overline{\boldsymbol{H}}$
④Apply conventional NMF and **Dirichlet NMF**



  - Compare how well the estimated $\boldsymbol{W}$ and $\boldsymbol{H}$ match the true basis matrix
  - Dirichlet parameters are set to induce sparse and smooth bases as shown in the figure

$\boldsymbol{\alpha}_1 = \begin{bmatrix}0.5\\0.5\\0.5\\0.5\\0.5\end{bmatrix}$　$\boldsymbol{\alpha}_2 = \begin{bmatrix}0.5\\0.5\\0.5\\0.5\\0.5\end{bmatrix}$　$\boldsymbol{\alpha}_3 = \begin{bmatrix}1.5\\1.5\\1.5\\1.5\\1.5\end{bmatrix}$

Parameters inducing sparsity/smoothness

- **Results**

True basis matrices (oracle)

Basis mat. (oracle)

0.00e+00	1.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01
1.00e+00	0.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01

Basis matrices (simple NMF)

Basis mat. (simple NMF)

2.53e−02	7.87e−01	3.73e−01
2.10e−02	1.42e−14	1.80e−01
9.35e−01	2.20e−01	1.56e−02
2.10e−02	1.16e−15	1.80e−01
2.10e−02	7.39e−15	1.80e−01

Basis matrices (Dirichlet NMF)

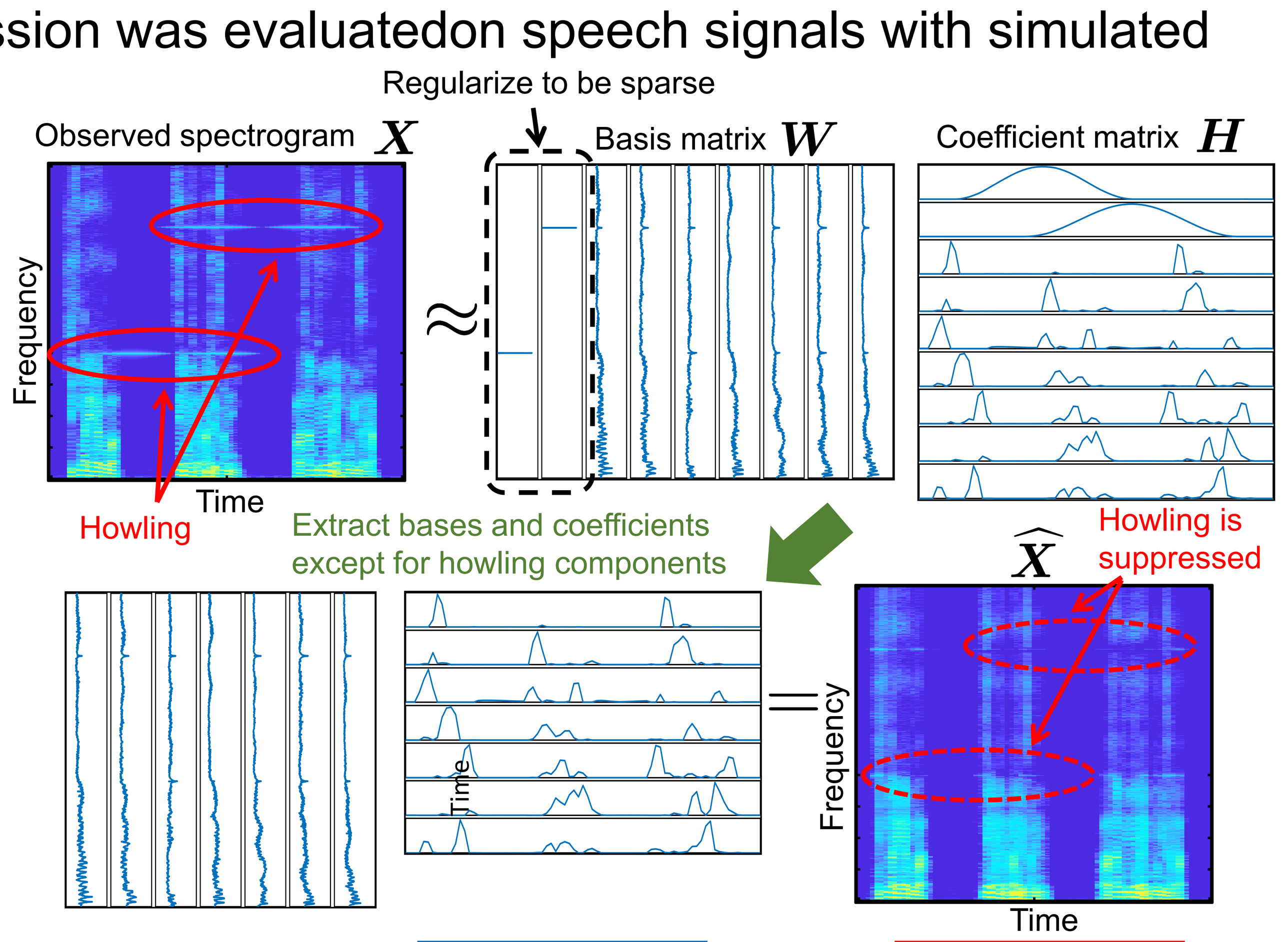
Basis mat. (Dirichlet NMF)

0.00e+00	1.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01
1.00e+00	0.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01

5. Application to Howling Suppression

- **Conditions**
  - Howling suppression was evaluatedon speech signals with simulated feedback noise.

Item	Description
Speech signals	3 utterances (JVS corpus)
Sampling rate	16 kHz
Howling signals	Two sinusoids added (1–8 kHz)
Compared methods	L1-sparse NMF Dirichlet NMF
Trials	50 random initializations
Metric	SDR



- **Results**
  - Violin plots of the SDR values, averaged over three speech signals and 50 random initializations
  - Dirichlet NMF achieved a higher average SDR (8.83 dB) than L1-sparse NMF (0.80 dB), confirming superior howling suppression