

ノルム制約・スパース正則化付き KL 基準 NMF の 補助関数法に基づく最適化*

☆和気佑弥, 北村大地, 綾野翔馬 (香川高専)

1 はじめに

非負値行列因子分解 (nonnegative matrix factorization: NMF) [1,2] は, 非負観測行列 $\mathbf{X} \in \mathbb{R}_{\geq 0}^{I \times J}$ を別の 2 つの非負変数行列 $\mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_K] \in \mathbb{R}_{\geq 0}^{I \times K}$ 及び $\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_K]^T \in \mathbb{R}_{\geq 0}^{K \times J}$ に低ランク近似する手法であり, 数多くの応用が提案されてきた [3]. ここで, \mathbf{W} 及び \mathbf{H} をそれぞれ基底行列及び係数行列と呼び, I 及び J は \mathbf{X} の行及び列数, K は \mathbf{W} の列数 (基底数), \cdot^T は転置である. NMF は一般に次式の最小化問題として定式化される.

$$\underset{\mathbf{W}, \mathbf{H}}{\text{Minimize}} \mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H}) \text{ s.t. } w_{ik}, h_{kj} \geq 0 \quad \forall i, j, k \quad (1)$$

ここで, w_{ik} 及び h_{kj} はそれぞれ \mathbf{W} 及び \mathbf{H} の要素, i 及び j はそれぞれ \mathbf{X} の行及び列のインデックス, k は \mathbf{W} の列のインデックスである. $\mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H})$ は \mathbf{X} と $\mathbf{W}\mathbf{H}$ の乖離度を測る損失関数であり, 本稿では次式の一般化 Kullback-Leibler (KL) 擬距離を扱う.

$$\mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H}) = \sum_{i,j} \left(x_{ij} \log \frac{x_{ij}}{\sum_k w_{ik} h_{kj}} - x_{ij} + \sum_k w_{ik} h_{kj} \right) \quad (2)$$

最適化問題 (1) の反復最適化として, 補助関数法または上界最小化 (majorization-minimization: MM) アルゴリズム [4,5] と呼ばれる最適化法の乗算更新式 [2] が広く用いられる. この手法は目的関数の単調非増加性が保証され, ステップサイズのようなパラメータも存在しないという大きな利点がある.

\mathbf{W} や \mathbf{H} にスパース正則化を付与したスパース NMF (sparse NMF: SNMF) も数多く提案されている [6–12]. 例えば, \mathbf{H} に L_1 ノルム正則化を適用する SNMF は次式となり, 以降 simple SNMF と呼ぶ.

$$\underset{\mathbf{W}, \mathbf{H}}{\text{Minimize}} \mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H}) + \mu \|\mathbf{H}\|_1 \text{ s.t. } w_{ik}, h_{kj} \geq 0 \quad \forall i, j, k \quad (3)$$

ここで, $\|\cdot\|_1$ は L_1 ノルムであり, $\mu \geq 0$ は重み係数である. NMF の変数 \mathbf{W} 及び \mathbf{H} には一般に, 任意の $\alpha > 0$ に対して $\mathbf{W}\mathbf{H} = (\alpha\mathbf{W})(\mathbf{H}/\alpha)$ が成立するスケール任意性が存在するため, $\alpha \rightarrow \infty$ とすれば最適化問題 (3) のスパース正則化項 $\|\mathbf{H}\|_1$ を最小化で

きる. 従って, 最適化問題 (3) のような一方の行列に対するスパース正則化は, スケール任意性を回避する制約条件を変数に課すか, スケールに非依存な正則化項を設計しない限り, 本質的に無意味となってしまう問題がある [9,13].

初期の SNMF では, 前述のスケール任意性問題を回避するため, 変数に $\|\mathbf{w}_k\| = 1 \quad \forall k$ なるノルム制約を課す方法 [6] が提案された. これは, 次のように定義され, 本稿では norm-constrained SNMF と呼ぶ.

$$\underset{\mathbf{W}, \mathbf{H}}{\text{Minimize}} \mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H}) + \mu \|\mathbf{H}\|_1 \text{ s.t. } w_{ik}, h_{kj} \geq 0 \quad \forall i, j, k, \|\mathbf{w}_k\|_1 = 1 \quad \forall k \quad (4)$$

同時期には, スケール非依存なスパース制約の付与 [7] も提案されたが, いずれも MM アルゴリズムによる最適化が断念され斜影勾配法が用いられた. そのため, 単調非増加性は保証されず, 安定性がステップサイズに依存する課題があった. また, simple SNMF に対して MM アルゴリズムに基づく乗算更新式を導出し, スケール任意性を解消するために変数更新毎に \mathbf{w}_k を正規化し, $\mathbf{w}_k \mathbf{h}_k^T$ が不変となるように正規化係数の逆数を \mathbf{h}_k に乗じるアルゴリズム [8] も提案された. しかし, 反復毎の正規化処理は正則化項の値を変化させるため, この手法も単調非増加性は保証されていない. その後, \mathbf{W} の各列ベクトルを正規化した行列 $\widetilde{\mathbf{W}}$ で損失関数に用いることでスケール任意性問題を回避する手法 [9,10] が提案された. これは, 次の最適化問題として定義され, 本稿では normalized SNMF と呼ぶ.

$$\underset{\mathbf{W}, \mathbf{H}}{\text{Minimize}} \mathcal{D}(\mathbf{X}|\widetilde{\mathbf{W}}\mathbf{H}) + \mu \|\mathbf{H}\|_1 \text{ s.t. } w_{ik}, h_{kj} \geq 0 \quad \forall i, j, k \quad (5)$$

この手法の乗算更新式が単調非増加となることは実験的に確認されるが, その導出過程で目的関数の勾配を正值項と負値項に分割し乗算更新式を構成する heuristic update rule [14] が用いられており, 単調非増加性に対する理論的な裏付けがないことが指摘されている [12].

近年では, norm-constrained SNMF に対し, ラグランジュの未定乗数法に基づいて乗算更新式を導く手法 [11] が提案された. この手法は単調非増加性を

* Auxiliary-function-based optimization of KL NMF with norm constraint and sparsity regularization. By Yuya WAKE, Daichi KITAMURA, and Shoma AYANO (NIT, Kagawa).

保証しつつスケール任意性問題も回避できるため、前述の問題をすべて解決できるフレームワークである。また、損失関数や制約条件等についても一般化されており、その適用範囲は広い。しかしながら、各反復においてラグランジュ乗数が有理関数の根を求める問題となり解析的に導出できず、毎回の変数の反復更新で数値解を求める内部反復処理を必要とする問題が残る。さらに、norm-constrained SNMF と次の最適化問題が本質的に等価であることを示し、MM アルゴリズムに基づく乗算更新式を導出する手法 [12] も提案されている。

$$\begin{aligned} & \underset{\mathbf{W}, \mathbf{H}}{\text{Minimize}} \mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H}) + \mu \|\mathbf{W}\mathbf{H}\|_1 \\ & \text{s.t. } w_{ik}, h_{kj} \geq 0 \quad \forall i, j, k \end{aligned} \quad (6)$$

この手法は、正則化項が \mathbf{W} 及び \mathbf{H} の両変数にまたがることでスケール任意性問題を回避しつつ、 $\|\mathbf{W}\mathbf{H}\|_1 = \sum_k \|\mathbf{w}_k \mathbf{h}_k^T\|_1$ より、 \mathbf{X} の低ランク近似を構成する K 個のランク 1 行列 $\mathbf{w}_k \mathbf{h}_k^T$ のそれぞれに対してスパース性を誘導することで、行列積 $\mathbf{W}\mathbf{H}$ がスパースとなる。しかし、SNMF を適用する目的が \mathbf{W} 及び \mathbf{H} の一方の変数にスパース性を与えることである場合、最適化問題 (6) はその目的に適合しない。

本稿では、上記の課題を克服する新たな SNMF の最適化法を提案する。提案法は文献 [11] と同様に、norm-constrained SNMF に対し MM アルゴリズムに基づく単調非増加性が保証された乗算更新式を導出する。このとき、一般化 KL 擬距離と L_1 ノルム制約の組み合わせならば制約条件のラグランジュ乗数が解析的に解けることを示し、文献 [11] の手法で必要だった数値解計算を回避している。なお、本稿は \mathbf{H} に対するスパース正則化を扱うが、 \mathbf{W} に対するスパース正則化を考へても一般性を失うことなく議論できる。

2 提案手法

2.1 ノルム制約付き KL 基準 SNMF の定式化

本稿では、norm-constrained SNMF に MM アルゴリズムを適用し、単調非増加性を保証した乗算更新式を導出する。改めて norm-constrained SNMF は次の最適化問題となる。

$$\begin{aligned} & \underset{\mathbf{W}, \mathbf{H}}{\text{Minimize}} \mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H}) + \mu \sum_{k,j} h_{kj} \\ & \text{s.t. } w_{ik}, h_{kj} \geq 0 \quad \forall i, j, k, \quad \sum_i w_{ik} = 1 \quad \forall k \end{aligned} \quad (7)$$

各基底ベクトル \mathbf{w}_k に L_1 ノルム制約 $\sum_i w_{ik} = 1$ を課すことで、 \mathbf{W} 及び \mathbf{H} 間のスケール任意性問題を回避する。以後、最適化の目的関数を \mathcal{J} と表す。

2.2 MM アルゴリズムに基づく反復更新式

MM アルゴリズムは、直接最適化が難しい目的関数に対して、(a) 現在の変数の地点で一定の要件を満たす補助関数 (上界関数) を設計、(b) 補助関数の停留点を閉形式で求めて次の地点の変数として更新、の 2 ステップを反復することで、本来の目的関数を間接的に最小化する。本稿においても、最適化問題 (7) に対して同様のアプローチを適用する。ただし、従来の NMF における MM アルゴリズム [2] では、変数の非負制約を陽に扱うことなく、「乗算型の更新式を導出すること」及び「変数の初期値を正值とすること」の 2 点で暗に非負制約を満たすような変数更新を実現している。最適化問題 (7) では、非負制約に加えて \mathbf{w}_k のノルム制約も課す必要があるため、前述の (a) ステップで設計した補助関数に対して、(b) ステップを等式・不等式制約条件付き最小化問題と捉え、ラグランジュの未定乗数法とカルーシュ・クーン・タッカー (Karush–Kuhn–Tucker: KKT) 条件を用いて解く。

まず、補助関数を設計する。一般化 KL 擬距離 (2) にイエンゼンの不等式を適用することで、次の不等式が得られる。

$$\begin{aligned} & \mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H}) \\ & \stackrel{c}{=} \sum_{i,j} \left(-x_{ij} \log \sum_k w_{ik} h_{kj} + \sum_k w_{ik} h_{kj} \right) \\ & \leq \sum_{i,j} \left(-x_{ij} \sum_k \delta_{ijk} \log \frac{w_{ik} h_{kj}}{\delta_{ijk}} + \sum_k w_{ik} h_{kj} \right) \end{aligned} \quad (8)$$

ここで、 $\stackrel{c}{=}$ は定数項の違いを除いて等しいことを表す。また、 $\delta_{ijk} > 0$ はイエンゼンの不等式の補助変数であり、 $\sum_k \delta_{ijk} = 1$ を満たす。従って、目的関数全体の補助関数 $\mathcal{J}^+ \geq \mathcal{J}$ は次のように設計できる。

$$\begin{aligned} \mathcal{J}^+ & \stackrel{c}{=} \sum_{i,j} \left(-x_{ij} \sum_k \delta_{ijk} \log \frac{w_{ik} h_{kj}}{\delta_{ijk}} + \sum_k w_{ik} h_{kj} \right) \\ & \quad + \mu \sum_{k,j} h_{kj} \end{aligned} \quad (9)$$

ここで、 $\mathcal{J}^+ = \mathcal{J}$ を満たす補助変数は次式となる。

$$\delta_{ijk} = \frac{w_{ik} h_{kj}}{\sum_{k'} w_{ik'} h_{k'j}} \quad (10)$$

次に、補助関数 \mathcal{J}^+ に対して、次の等式・不等式制約条件付き最小化問題を考える。

$$\begin{aligned} & \underset{\mathbf{W}, \mathbf{H}, \Delta}{\text{Minimize}} \mathcal{J}^+ \\ & \text{s.t. } w_{ik}, h_{kj} \geq 0 \quad \forall i, j, k, \quad \sum_i w_{ik} = 1 \quad \forall k \end{aligned} \quad (11)$$

ここで、 Δ は全インデクスの δ_{ijk} の集合を表す。この最適化問題のラグランジアンは次式で与えられる。

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

ここで、 $\lambda_k \geq 0$ はノルム制約に対応するラグランジュ乗数である。非負制約条件付き最適化問題の KKT 条件より、次式を得る。

$$w_{ik} \frac{\partial \mathcal{L}}{\partial w_{ik}} = 0 \quad \forall i, k \quad (13)$$

$$\frac{\partial \mathcal{L}}{\partial w_{ik}} \geq 0 \quad \forall i, k \quad (14)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_k} = 0 \quad \forall k \quad (15)$$

式(13)より、 \mathcal{L} の最小解は $w_{ik} = 0$ または $\partial \mathcal{L} / \partial w_{ik} = 0$ のいずれかを満たす。

以後、 w_{ik} を求める。式(13)より、次式を得る。

$$\sum_j \left(-x_{ij} \frac{\delta_{ijk}}{w_{ik}} + h_{kj} \right) - \lambda_k = 0 \quad (16)$$

これを整理して、 w_{ik} は次式となる。

$$w_{ik} = \frac{\sum_j x_{ij} \delta_{ijk}}{\sum_j h_{kj} - \lambda_k} \quad (17)$$

また、式(15)より $\sum_i w_{ik} - 1 = 0$ を得るが、これはノルム制約 $\sum_i w_{ik} = 1 \quad \forall k$ と等価である。このノルム制約に式(17)を代入して整理することで、ラグランジュ乗数を解析的に解くことができる。

$$\lambda_k = \sum_j h_{kj} - \sum_{i,j} x_{ij} \delta_{ijk} \quad (18)$$

式(18)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。

$$w_{ik} = \frac{\sum_j x_{ij} \delta_{ijk}}{\sum_{i,j} x_{ij} \delta_{ijk}} \quad (19)$$

以上より、補助関数 \mathcal{J}^+ の制約条件付き最小解は式(19)である。なお、補助変数の等式条件(10)を式(19)に代入すると、 w_{ik} は次式となる。

$$w_{ik} = w_{ik} \frac{\sum_j \frac{x_{ij} h_{kj}}{\sum_{k'} w_{ik'} h_{k'j}}}{\sum_i w_{ik} \sum_j \frac{x_{ij} h_{kj}}{\sum_{k'} w_{ik'} h_{k'j}}} \quad (20)$$

結局、 $w_{ik} \geq 0$ において、式(20)は全てのインデクスで $\partial \mathcal{L} / \partial w_{ik} = 0$ を満たす上、非負制約及びノルム制約を必ず満たす更新式となる。

同様に、KKT 条件を用いることで、 h_{kj} については $\partial \mathcal{L} / \partial h_{kj} = 0$ から次式が得られる。

$$h_{kj} = \frac{\sum_i x_{ij} \delta_{ijk}}{\sum_i w_{ik} + \mu} \quad (21)$$

また、補助変数の等式条件(10)を式(21)に代入すると、 h_{kj} は次式となる。

$$h_{kj} = h_{kj} \frac{\sum_i x_{ij} \frac{w_{ik}}{\sum_{k'} w_{ik'} h_{k'j}}}{\sum_i w_{ik} + \mu} \quad (22)$$

これより、式(22)も $h_{kj} \geq 0$ で非負制約を必ず満たす。なお、式(22)は、 L_1 ノルム正則化を用いる多くの SNMF の \mathbf{H} の更新式と同様である [8, 10–12]。

式(18)としてラグランジュ乗数の解析解が得られたが、このような手続きは SNMF の最適化問題が一般化 KL 擬距離以外や L_1 ノルム制約以外を用いた場合は困難である。例えば、二乗 Euclid 距離や Itakura–Saito 擬距離の損失関数を用いる場合、式(17)の分母に x_{ij} を含む項が現れる。その結果、 λ_k の解は文献 [11] の通り有利関数の根を求める問題となってしまう、数値解の計算が必要となる。

3 実験

3.1 実験条件

本章では、前章で導出した norm-constrained SNMF の MM アルゴリズムに基づく乗算更新式 (MM-based update) が、目的関数値の単調非増加性を保ちながら収束することを確認する。また、比較対象として、反復毎に変数の正規化を伴う simple SNMF (MM-based update with rescaling) [8] 及び正規化済み行列を直接目的関数に導入する normalized SNMF (heuristic update) [10] を採用した。提案法と合わせたこれら 3 つの手法は同じ目的関数のため、その挙動を比較可能である。

実験には、SiSEC 2011 [15] データセットのギター音源 dev1_wdrums_src_1.wav の冒頭 10 s を用い、短時間フーリエ変換から振幅スペクトログラムを得た。サンプリング周波数 16 kHz の音源に対し、窓長 1024 点 (ハン窓)、ハーフシフトで変換し、 $(I, J) = (513, 313)$ の観測行列 \mathbf{X} を構成した。

実験として、 \mathbf{X} を各 SNMF で近似分解し、反復毎の目的関数値を算出して収束挙動を評価した。基底数 K は 10 から 100 までを等間隔に 10 通り、正則化項の重み係数 μ は 10^{-6} から 1 までを対数スケールで等間隔に 20 通り設定し、全組合せについて網羅的に実験を実施した。また、 \mathbf{W} 及び \mathbf{H} の初期化は区間 $(0, 1)$ の一様分布から生成し、各 K, μ の条件に対して 50 個の乱数シードを用いて目的関数値の平均的な収束の様子を調査した。反復更新回数は 10000 回とし、途中の収束条件は設けていない。反復毎に正規化を必要とする simple SNMF については、各反復において変数更新直後と正規化処理後の目的関数値を併せて記録した。

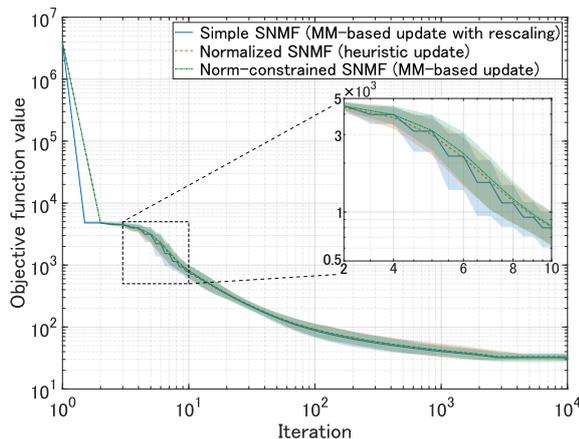


Fig. 1 Convergence behaviors when $K = 100$ and $\mu = 10^{-6}$. Lines and colored areas respectively show mean values and ranges of standard deviations.

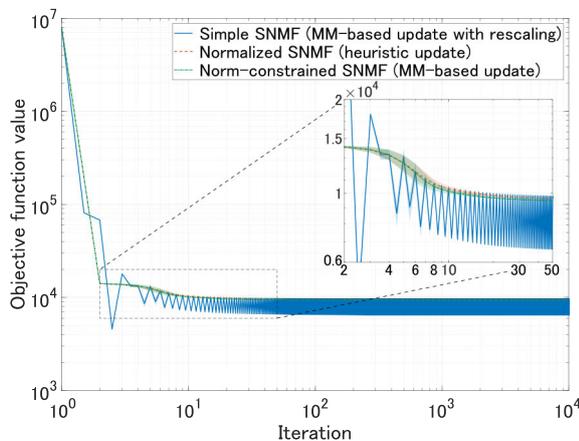


Fig. 2 Convergence behaviors when $K = 100$ and $\mu = 1$. Lines and colored areas respectively show mean values and ranges of standard deviations.

3.2 実験結果

基底数 $K = 100$, 重み係数 $\mu = 10^{-6}$ における各手法の収束カーブを Fig. 1 に, また, $\mu = 1$ の場合を Fig. 2 に示す. いずれの図も, 乱数シードに対する目的関数値の平均をプロットしており, 標準偏差を透過帯で表現している. なお, simple SNMF については, $n + 1$ 回目の更新における正規化処理前の値を, 便宜上 $n + 0.5$ 回目の更新の値としてプロットしている.

結果より, simple SNMF は小さい重み係数 μ においては優れた収束を示す一方で, μ の増加によって目的関数値の単調非増加性が失われることが確認できる. これに対して, 提案する norm-constrained SNMF 及び normalized SNMF はいずれの重み係数 μ に対しても安定した収束挙動を示し, 目的関数が常に単調非増加性を維持することが確認された. さらに, 異なる K においても, 提案法は他の手法と同等かそれ以下の目的関数値に収束しつつ, 常に安定した挙動を示した.

4 おわりに

本稿では, ノルム制約を導入した KL 基準 SNMF に対し, スケール任意性問題の解消と, 目的関数の単調非増加性の保証を両立する新たな最適化手法を提案した. 補助関数法に基づき KKT 条件を厳密に満たす形で更新式を導出し, 理論的に単調非増加性を担保した上で, 従来法との比較実験によりその有効性を実証した.

謝辞 本研究の一部は JSPS 科研費 23K24908 の助成を受けたものである.

参考文献

- [1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [2] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. in Neural Info. Process. Syst.*, 2000.
- [3] X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 36, pp. 59–80, 2019.
- [4] D. R. Hunter and K. Lange, "Quantile regression via an MM algorithm," *J. Comput. Graph. Statist.*, vol. 9, no. 1, pp. 60–77, 2000.
- [5] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, 2017.
- [6] P. O. Hoyer, "Non-negative sparse coding," in *Proc. IEEE Workshop Neural Netw. Signal Process.*, pp. 557–565, 2002.
- [7] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, pp. 1457–1469, 2004.
- [8] W. Liu, N. Zheng, and X. Lu, "Non-negative matrix factorization for visual coding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 3, 2003.
- [9] J. Eggert and E. Körner, "Sparse coding and NMF," in *Proc. Int. Joint Conf. Neur. Netw.*, pp. 2529–2533, 2004.
- [10] J. L. Roux, F. J. Weninger, and J. R. Hershey, "Sparse NMF – half-baked or well done?," *Mitsubishi Electric Research Laboratories, Tech. Rep. TR2015-023*, 2015.
- [11] V. Leplat, N. Gillis, and J. Idier, "Multiplicative updates for NMF with β -divergences under disjoint equality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 42, no. 2, pp. 730–752, 2021.
- [12] A. Marmin, J. H. d. M. Goulart, and C. Févotte, "Majorization-minimization for sparse nonnegative matrix factorization with the β -Divergence," *IEEE Trans. Signal Process.*, vol. 71, pp. 1435–1447, 2023.
- [13] Y. Iwase and D. Kitamura, "Supervised audio source separation based on nonnegative matrix factorization with cosine similarity penalty," *IEICE Trans. Fundamentals*, vol. E105-A, no. 6, pp. 906–913, 2022.
- [14] C. Févotte and J. Idier, "Algorithms for Nonnegative Matrix Factorization with the β -Divergence," in *Neural Comput.*, vol. 23, no. 9, pp. 2421–2456, 2011.
- [15] S. Araki, F. Nesta, E. Vincent, Z. Koldovsky, G. Nolte, A. Ziehe, and A. Benichoux, "The 2011 Signal Separation Evaluation Campaign (SiSEC2011): - Audio source separation -," in *Proc. Int. Conf. Latent Variable Anal. Signal Sep.*, pp. 414–422, 2012.