

分離可能想定下での非負行列分解に対する楕円丸め法

Ellipsoidal Rounding for Nonnegative Matrix Factorization Under Noisy Separability

水谷友彦

Tomohiko Mizutani

神奈川大学 工学部 情報システム創成学科

Faculty of Engineering, Department of Information Systems Creation, Kanagawa University

This manuscript provides an explanation of the algorithm proposed in the paper [9] for a nonnegative matrix factorization under noisy separability. The idea behind is shown from the geometric point of view. Then, the theoretical properties are described. Finally, numerical experiments are reported to show the practical performance.

1. はじめに

分離可能性という仮定の下での非負行列分解 (NMF) は、データの凸包の頂点を求める問題に読み替えることができる。本稿では、この問題に対して [9] で提案した手法について説明する。本稿は [10] における記述を参考にしている。

NMF とは与えられた非負行列を二つの非負行列の積の形に分解するという問題である。ここで、非負行列とは全ての要素が非負な実数となる行列である。この問題の正確な記述は以下の通りである。大きさが $d \times m$ の非負行列の集合を $\mathbb{R}_+^{d \times m}$ と書く。今、 $\mathbf{A} \in \mathbb{R}_+^{d \times m}$ と自然数 r が与えられているとする。このとき、 \mathbf{A} を

$$\mathbf{A} = \mathbf{F}\mathbf{W} + \mathbf{N}$$

となる $\mathbf{F} \in \mathbb{R}_+^{d \times r}$ と $\mathbf{W} \in \mathbb{R}_+^{r \times m}$ の積の形で分解することを考える。ここで、 \mathbf{N} は大きさが $d \times m$ の実行列で分解の残差を表す。 \mathbf{A} の NMF とは、残差を表す行列 \mathbf{N} をできるだけ零行列に近づけるような \mathbf{F} と \mathbf{W} を求める問題である。NMF は画像分析、クラスタリング、トピックモデリングなどに応用され、その有効性が報告されている。

NMF は NP 困難な問題であると知られている。この困難さに対して、Arora ら [1] は分離可能性という仮定を利用することを提案した。分離可能性を仮定すると NMF は扱い易い問題になる。 $\mathbf{A} \in \mathbb{R}_+^{d \times m}$ が

$$\mathbf{A} = \mathbf{F}\mathbf{W} \text{ for } \mathbf{F} \in \mathbb{R}_+^{d \times r} \text{ and } \mathbf{W} = (\mathbf{I}, \mathbf{K})\mathbf{\Pi} \in \mathbb{R}_+^{r \times m} \quad (1)$$

と表すことができるとき \mathbf{A} は**分離可能性**を満たすと言う。ここで、 \mathbf{I} は $r \times r$ の単位行列、 $\mathbf{\Pi}$ は $m \times m$ の置換行列を表す。 \mathbf{A} が分離可能性を満たすとは、 \mathbf{A} はその列ベクトルとして \mathbf{F} の列ベクトルを有していることを意味する。本稿では分離可能性を満たす行列を**分離可能な行列**と呼ぶ。(1) の形の分離可能な行列 \mathbf{A} を構成する \mathbf{F} を**基底行列**、 \mathbf{W} あるいはその部分行列 \mathbf{K} を**重み行列**と呼ぶ。分離可能な行列 \mathbf{A} にノイズを表す行列 \mathbf{N} を加えた行列 $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{N}$ を**ノイズを含む分離可能な行列**と呼ぶ。ノイズを含む分離可能な行列 $\tilde{\mathbf{A}}$ は

$$\begin{aligned} \tilde{\mathbf{A}} &= \mathbf{F}(\mathbf{I}, \mathbf{K})\mathbf{\Pi} + \mathbf{N} \\ &= (\mathbf{F} + \mathbf{N}^{(1)}, \mathbf{F}\mathbf{K} + \mathbf{N}^{(2)})\mathbf{\Pi} \end{aligned} \quad (2)$$

と表すことができる。ここで $\mathbf{N}^{(1)}$ と $\mathbf{N}^{(2)}$ は $\mathbf{N}\mathbf{\Pi}^{-1} = (\mathbf{N}^{(1)}, \mathbf{N}^{(2)})$ を満たす \mathbf{N} の部分行列でそれぞれの大きさは

連絡先: mizutani@kanagawa-u.ac.jp

$d \times r$ と $d \times (m - r)$ である。本稿では $\mathbf{F} + \mathbf{N}^{(1)}$ を**ノイズを含む基底行列**と呼ぶ。

本稿ではノイズを含む分離可能性の下での NMF を考える。この問題は次のように記述される。

問題 1. データ行列 \mathbf{M} はノイズを含む分離可能な $d \times m$ の行列とする。 \mathcal{I} は集合 $\{1, \dots, m\}$ の部分集合で r 個の要素を持つとする。このとき、 $\mathbf{M}(\mathcal{I})$ ができるだけ基底 \mathbf{F} に近くなるような \mathcal{I} を求めよ。

ここで、 $\mathbf{M}(\mathcal{I})$ は \mathbf{M} の部分行列でその列ベクトルの添字集合が \mathcal{I} となるようなものを表す。本稿では問題 1 におけるデータ行列 \mathbf{M} の列ベクトルを**データ点**と呼ぶ。問題 1 に対するアルゴリズムは以下のような性質を持つことが望まれる。

- (正当性) データ行列 \mathbf{M} は分離可能な行列であるとする。このとき、アルゴリズムは $\mathbf{M}(\mathcal{I}) = \mathbf{F}$ となる \mathcal{I} を出力する。
- (頑強性) データ行列 \mathbf{M} はノイズを含む分離可能な行列であるとする。つまり \mathbf{M} は分離可能な行列とノイズを表す行列 \mathbf{N} の和で与えられているとする。もし $\|\mathbf{N}\|_p < \epsilon$ なら、アルゴリズムは $\|\mathbf{M}(\mathcal{I}) - \mathbf{F}\|_p < \tau\epsilon$ なる \mathcal{I} を出力する。ここで τ は実定数、 $\|\cdot\|_p$ は行列の p ノルムを表す。

頑強性において、アルゴリズムがノイズを含む基底を特定できると $\tau = 1$ となる。

[9] の提案手法は正当性と $\tau = 1$ の頑強性を理論的に保証することができる。提案手法のアルゴリズムは分離可能な行列の幾何的な構造に基いて設計されている。分離可能な行列に比較的妥当な仮定を置くと、その列ベクトルの凸包は単体となり、頂点は基底行列の列ベクトルに対応する。したがって、単体の頂点を全て見つけることができると分離可能な行列の基底行列が得られる。提案手法は以下のような二つの幾何的な性質を利用している。(2.2 節の性質 1) 単体に対して体積最小閉包楕円を描くとその頂点を見つけることができる。(2.2 節の性質 2) 多少のノイズが加わったとしても、体積最小閉包楕円を描くことで単体の頂点に対応している点を見つけることができる。

本稿は次のように構成されている。2 節では、まず幾何的な観点から提案手法が背景としているアイデアを説明する。そして、提案手法の概要を説明した後、理論的に保証されている手法の性質を述べる。3 節では既存手法との比較について述べる。4 節ではノイズへの頑強性に関する実験とトピック抽出に関する実験の結果を報告する。

2. 提案手法の概要と性質

2.1 仮定と問題 1 の幾何的な解釈

提案手法は分離可能な行列に対して以下のような仮定を置く。

仮定 1. 分離可能な行列 $A \in \mathbb{R}_+^{d \times m}$ は以下のような基底行列 $F \in \mathbb{R}_+^{d \times r}$ と重み行列 $W \in \mathbb{R}_+^{r \times m}$ で構成される。

- 1-a) W のどんな列ベクトルもその 1 ノルムは 1 となる。
- 1-b) F の r 個の列ベクトルは線形独立である。

1-a は一般性を失わずに仮定できる。この仮定は W の列ベクトルは超平面上 $\{x : e^\top x = 1\}$ に載っていることを意味する。ここで e はすべての要素が 1 であるベクトルを表している。1-b は比較的強い仮定である。しかし、実際の応用から生じる NMF においては、基底行列 F の列ベクトルが線形従属となることは稀である。実際、既存手法 [7] ではこれら 2 つの仮定の下で、正当性と頑強性を示している。

仮定 1 の下での問題 1 を幾何的な観点から考察する。簡単のために、データ行列は (1) の形の分離可能な行列 A とする。すると、仮定から $\text{conv}(A)$ は d 次元空間上の $(r-1)$ 次元単体となり、頂点が基底ベクトルに対応する。ここで、行列に対する $\text{conv}(\cdot)$ でその列ベクトルの凸包を表す。従って、 $\text{conv}(A)$ の全ての頂点を見つけることができると、基底行列 F が得られることが分かる。 $\text{conv}(A)$ の全ての頂点を計算することは容易であり、実際、効率的なアルゴリズムが存在する。次に、データ行列はノイズを含む分離可能な行列 \tilde{A} とする。この場合、 $\text{conv}(\tilde{A})$ の頂点は必ずしも基底ベクトルに対応するとは限らない。そのためノイズを含まない場合と比べて問題が難しくなることが分かる。図 1 の左はノイズを含まない場合、右はノイズを含む場合のデータ行列の列ベクトルの凸包を表している。

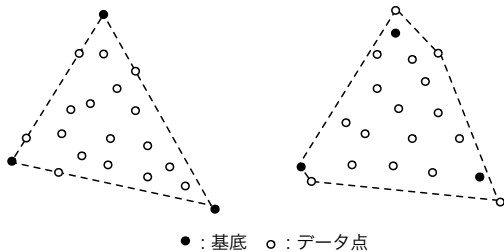


図 1: 仮定 1 の下での $(d, r) = (3, 3)$ における分離可能なデータ行列の凸包。(左) ノイズが無い場合、(右) ノイズが有る場合。

2.2 基盤となる性質

提案手法は次のような二つの幾何的な性質に基づいている。 r 次元空間上の $(r-1)$ 次元単体 Δ を考える。 Δ の r 個の頂点を $g_1, \dots, g_r \in \mathbb{R}^r$ 、 Δ 内の ℓ 個の点を $b_1, \dots, b_\ell \in \mathbb{R}^r$ と書く。これらに対して、正負の符号を付けたベクトルの集合 $\{\pm g_1, \dots, \pm g_r, \pm b_1, \dots, \pm b_\ell\}$ を S とする。

性質 1 ([9] の Proposition 1). S を含むような原点を中心を持つ体積最小閉包楕円を描くと、 $\pm g_1, \dots, \pm g_r$ のみが楕円の境界上に載る。

ここで、 S の凸包は r 次元の多面体で、四角形を赤道面に持つ双角錐となる。特に、 $r=3$ の場合、つまり、三次元の場合は八面体となる。図 2 は $r=3$ の S に対する原点中心な体

積最小閉包楕円の様子を表している。性質 1 は r 次元空間上の $(r-1)$ 次元単体の頂点は閉包楕円を利用することで求められることを意味している。

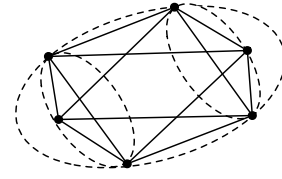


図 2: $r=3$ の S に対する体積最小閉包楕円。このとき S の凸包は八面体になる。

単体 Δ 内の点 b_i はその頂点 g_1, \dots, g_r の凸結合として $b_i = Gk_i$ のように書ける。ここで G は g_1, \dots, g_r を列ベクトルとして持つ $r \times r$ の行列で、 k_i は非負ベクトルでその要素を足し合わせると 1 となる。つまり、

$$\|k_i\|_1 = 1 \text{ and } k_i \geq 0, \quad i = 1, \dots, \ell$$

を満たす。今、点 b_i を摂動させた点 b'_i を考える。点 b'_i は

$$\|k'_i\|_2 < 1, \quad i = 1, \dots, \ell$$

なるベクトル k'_i を用いて、 $b'_i = Gk'_i$ と表すことができるとする。このような b'_1, \dots, b'_ℓ に対しても、性質 1 と同様の性質が成り立つ。集合 $\{\pm g_1, \dots, \pm g_r, \pm b'_1, \dots, \pm b'_\ell\}$ を S' と書く。

性質 2 ([9] の Cororally 1). S' を含むような原点を中心を持つ体積最小楕円を描くと、 $\pm g_1, \dots, \pm g_r$ のみが楕円の境界上に載る。

性質 2 は r 次元空間上の $(r-1)$ 次元単体 Δ の点にノイズが多少加わったとしても、単体の頂点はその閉包楕円を計算することで見つけることができることを意味している。

2.3 アルゴリズムとその性質

提案手法のアルゴリズムの概要を述べる。2.1 節において、仮定 1 の下では分離可能な行列 A の凸包は d 次元空間上の $(r-1)$ 次元単体となり、その頂点が基底ベクトルに対応していることを考察した。性質 1、2 を利用するためには空間の次元を d から r に削減する必要があるが、これは A に対して特異値分解 (SVD) を適用することで実現することができる。提案手法の概要は下記のアロリズム 1 のようになる。

Algorithm 1 楕円丸め法の概要

入力: $M \in \mathbb{R}_+^{d \times m}$ 、自然数 r ; **出力:** \mathcal{I}

- 1: M の SVD を利用して冗長な次元を削減した行列 $P \in \mathbb{R}^{r \times m}$ を求める。
- 2: P の列ベクトル $p_1, \dots, p_m \in \mathbb{R}^r$ に対して $S = \{\pm p_1, \dots, \pm p_m\}$ を準備する。 S に対して原点中心な体積最小閉包楕円を計算し、楕円の境界上に載っている点の添字集合 \mathcal{I} を出力する。

アルゴリズム 1 の主要な計算は、Step 1 における行列 M の SVD の計算と、Step 2 における S に対する体積最小閉包

楕円の計算である。 S に対する体積閉包楕円の計算は m 変数の凸計画問題として定式化することができ、Frank-Wolfe 法や内点法を利用することで多項式時間で大域的な最適解が得られることが知られている。データ数が多い場合は、切除平面法を内点法と組み合わせると有効に動作することが知られている。したがって、提案手法では内点法と切除平面法を組み合わせるという手法を使用している。

アルゴリズム 1 は以下のような性質を持つことを保証できる。

- (正当性; [9] の Theorem 1) 分離可能な行列 A は仮定 1 を満たすとする。 (A, r) に対してアルゴリズム 1 を実行して得られた添字集合を \mathcal{I} は $A(\mathcal{I}) = F$ を満たす。
- (頑強性; [9] の Theorem 2) ノイズを含む分離可能な行列 $\tilde{A} = A + N$ を考える。ここで A は分離可能な行列、 N はノイズを表す行列とする。 A は仮定 1 を満たすとする。 ϵ を

$$\epsilon = \frac{1}{4}\sigma(1 - \mu) \quad (3)$$

となるように取る。 (\tilde{A}, r) に対してアルゴリズム 1 を実行して得られた添字集合を \mathcal{I} とする。すると、 $\|N\|_2 < \epsilon$ なら、 $\|\tilde{A}(\mathcal{I}) - F\|_2 < \epsilon$ を満たす。

これらはそれぞれ性質 1, 2 を基盤として導くことができる。頑強性の記述における σ と μ は \tilde{A} を構成してる分離可能な行列 A によって定まる値である。 σ は A の基底行列 F の最小特異値である。つまり、データの凸包の潰れ具合を表す値である。 μ は A の重み行列 K の列ベクトル k_i によって定まる値で、

$$\mu = \max_{i=1, \dots, m-r} \|k_i\|_2$$

を表している。仮定 1-a と k_i の非負性から $\mu \leq 1$ を満たす。 μ は基底とデータ点との離れ具合を表す値であり、特に、基底とデータ点が一致する場合に $\mu = 1$ となり、基底とデータ点が離れるに従って μ の値は小さくなる。

2.4 実践的なアルゴリズム

アルゴリズム 1 において、入力データに含まれるノイズの大きさが (3) の ϵ より大きい場合、楕円の境界上には r 個以上のデータ点が載る可能性がある。その場合、SPA[7] や XRAY[8] などの既存手法を利用して境界上の点から r 個の点を選択する。アルゴリズム 2 ではアルゴリズム 1 は既存手法の性能を

Algorithm 2 実践的な楕円丸め法の概要

入力: $M \in \mathbb{R}_+^{d \times m}$ 、自然数 r ; 出力: \mathcal{I}

- 1: アルゴリズム 1 を実行して得られた集合を \mathcal{J} とする。
- 2: $|\mathcal{J}| > r$ ならば、既存手法を利用して \mathcal{J} から r 個の要素を選択し、それらの集合を \mathcal{I} として出力する。

高めるための前処理として働くものとみなせる。

3. 既存手法

分離可能性の下での NMF に対する手法としては主に AGKM[1]、Hottopixx[4, 5, 6]、SPA[7]、XRAY[8] が知られている。表 1 はそれらの正当性、頑強性に関する理論的な保証についてまとめたものである。表中の“○”は対応している性質が理論的に保証されていることを、“×”は現在のところ必ず

しも理論的には保証はされていないことを表している。“ τ ”は 1 節の頑強性に関する記述における τ を意味しており、“ ≥ 1 ”は対応するアルゴリズムの τ の値は 1 以上であることを、“1”は τ の値が 1 であることを表している。“計算コスト”という欄は各アルゴリズムの主要な計算コストについての記述をまとめている。LP とは線形計画問題のことを表している。例えば、AGKM における記述は m 個の $m-1$ 変数の LP を解く必要があることを意味している。

表 1: 既存手法の正当性と頑強性

	正当性	頑強性	τ	計算コスト
AGKM	○	○	≥ 1	m 個の $m-1$ 変数の LP
Hottopixx	○	○	1	m^2 変数の LP
SPA	○	○	≥ 1	軽い
XRAY	○	×	—	比較的軽い

正当性に関しては四つのアルゴリズムとも保証されている。頑強性に関しては XRAY 以外の三つのアルゴリズムは保証されている。Hottopixx は提案手法と同様に正当性と $\tau=1$ の頑強性が保証されている。しかし、 m^2 変数の LP を解く必要がある。ここで m はデータの個数に対応している。一般に LP は解き易い問題であるが、規模が大きくなると問題を解くために必要な記憶容量が大きくなり解くことが困難になる。そのため論文 [4] では大規模な LP を解くための並列計算手法を提案している。SPA と XRAY は実用的な手法で、実際に実験を通して大規模なデータに対しても有効であることが報告されている。

4. 数値実験

4.1 ノイズに対する頑強性の比較

人工データを用いて提案手法であるアルゴリズム 2 と既存手法である SPA[7] と XRAY[8] の 3 つのバリエーション “max”, “dist”, “greedy” のノイズに対する頑強性の比較を行った。本稿における実験で使用したプログラムは MATLAB 上で実装した。以下では、XRAY において、例えば “max” というバリエーションを用いたものを XRAY(max) と記述する。提案手法であるアルゴリズム 2 の Step 2 において、例えば既存手法として XRAY(max) を利用したものを ER-XRAY(max) と記述する。

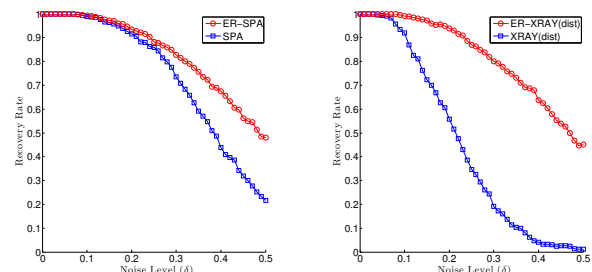


図 3: 復元率の比較

実験データは (2) の形式の \tilde{A} を人工的に生成した。データの大きさは $(d, m, r) = (250, 5000, 10)$ とした。 W の要素は $[0, 1]$ 上の一様分布、 K の列ベクトルはパラメータが $[0, 1]$ 上

表 2: NIPS データセットから得られたトピック

アンカーワード	1	2	3	4	5
spike	neurons	neuron	firing	synaptic	cell
motion	visual	eye	cells	image	object
policy	reinforcement	state	action	actor	critic
chip	circuit	analog	voltage	vlsi	synapse
speech	word	recognition	hmm	speaker	training

のディリクレ分布、 \mathbf{N} の要素は平均 0、分散 δ^2 の正規分布にしたがって生成した。 δ はノイズの大きさを表すパラメータである。アルゴリズムの評価は以下のような復元率を用いた。

$$\text{Recovery Rate} = |\mathcal{I}_* \cap \mathcal{I}| / |\mathcal{I}_*|$$

\mathcal{I}_* は実験データ $\tilde{\mathbf{A}}$ において、基底ベクトルの添字集合、 \mathcal{I} はアルゴリズムで得られた添字集合を表す。アルゴリズムによって得られた基底と人工的に生成した真の基底が一致したとき復元率は 1 となる。

実験では、 δ の値を 0 から 0.5 までの 0.01 刻みで各 δ 毎に 50 個のデータ行列を生成し、その 50 個のデータに対するアルゴリズムの復元率の平均値を求めた。図 3 はそれをまとめたものである。ER-XRAY(max)、ER-XRAY(greedy) の実験結果はそれぞれ ER-SPA、ER-XRAY(dist) と似ているので本稿では省略した。この図から、提案手法である ER は既存手法である SPA、XRAY と比較してよりノイズに対して頑強であることが読み取れる。

4.2 トピック抽出

最後に提案手法をトピック抽出に適用して得られた結果を報告する。 d 個の文章を考える。ここでの文章とは、例えばニュース記事や論文などに対応する。文章を bag-of-words という形式でベクトルを用いて表現する。今、 d 個の文章中の単語の総数を m とする。すると、文章 i は m 次元ベクトル \mathbf{a}_i を用いて表される。この文章ベクトル \mathbf{a}_i の j 番目の要素は非負で、文章 i における j 番目の単語の出現頻度を表す。

文章ベクトル \mathbf{a}_i は文章のトピックを表す r 個の m 次元非負ベクトル $\mathbf{w}_1, \dots, \mathbf{w}_r$ の凸結合で生成されるというモデルを考える。 $\mathbf{w}_1^\top, \dots, \mathbf{w}_r^\top$ を上から下へ並べて得られる大きさが $r \times m$ の非負行列を \mathbf{W} と書く。すると、 \mathbf{a}_i は $\mathbf{a}_i^\top = \mathbf{f}_i^\top \mathbf{W}$ と表すことができる。ここで \mathbf{f}_i は r 次元の非負ベクトルで、その要素を足し合わせると 1 となる。つまり、 $\mathbf{A} = \mathbf{F}\mathbf{W}$ となる。 \mathbf{A} は $\mathbf{a}_1^\top, \dots, \mathbf{a}_d^\top$ を上から下へ並べた $d \times m$ の非負行列、 \mathbf{F} は $\mathbf{f}_1^\top, \dots, \mathbf{f}_d^\top$ を同様に並べた $d \times r$ の非負行列としている。

いま、この行列 \mathbf{A} が分離可能であると仮定する。つまり、あるトピックベクトル \mathbf{w}_{j^*} には含まれるが、他のトピックベクトル \mathbf{w}_j , $j \in \{1, \dots, r\} \setminus \{j^*\}$ には含まれないような単語が存在することを仮定する。このような単語をアンカーワードと呼ぶ。アンカーワードは対応しているトピックの意味を良く表すような単語であると考えられる。トピックベクトルがアンカーワードを持つという仮定は比較的妥当であるとされている [2, 3, 8]。

表 2 は提案手法であるアルゴリズム 2 を NIPS データセットに適用して得られたトピックをまとめている。このデータセットは 1987 年から 1999 年までの NIPS という機械学習に関する会議で発表された 1740 本の論文によって構成されている文章データである。トピックのアンカーワードは“アンカーワード”、出現頻度が高かった上位 5 個の単語をその順番に従って“1”から“5”の欄に記述している。実験では $r = 5$ としてア

ルゴリズム 2 を実行した。この表から提案手法を用いて意味のあるトピックが抽出できたことが分かる。

5. おわりに

提案手法では行列の SVD とデータ点に対する体積最小閉包楕円の計算を行う必要がある。そのため、並列計算機へ実装を考えた場合、良い並列効率が得られない可能性がある。今後の課題として、SVD を経由せずに同様な手法が設計できるかを検討したい。

参考文献

- [1] S. Arora, R. Ge, R. Kannan, and A. Moitra: Computing a nonnegative matrix factorization – provably, STOC, pp. 145–162, 2012.
- [2] S. Arora, R. Ge, and A. Moitra: Learning topic models – going beyond SVD, FOCS, pp. 1–10, 2012.
- [3] S. Arora, R. Ge, Y. Halpern, D. Mimno, and A. Moitra: A practical algorithm for topic modeling with provable guarantees, ICML, pp. 280–288, 2013.
- [4] V. Bittorf, B. Recht, C. Re, and J.A. Tropp: Factoring nonnegative matrices with linear programs, NIPS, pp. 1223–1231, 2012.
- [5] N. Gillis: Robustness analysis of Hottopixx, a linear programming model for factoring nonnegative matrices: SIAM Journal on Matrix Analysis and Applications, 34(3):1189–1212, 2013.
- [6] N. Gillis and R. Luce: Robust near-separable nonnegative matrix factorization using linear optimization. arXiv:1302.4385v1, 2013.
- [7] N. Gillis and S.A. Vavasis: Fast and robust recursive algorithms for separable nonnegative matrix factorization. Available on Early Access Articles at IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013.
- [8] A. Kumar, V. Sindhwani, and P. Kambadur: Fast conical hull algorithms for near-separable non-negative matrix factorization, ICML, pp. 231–239, 2013.
- [9] T. Mizutani: Ellipsoidal rounding for nonnegative matrix factorization under noisy separability, arXiv:1309.5701, Accepted with minor revisions in Journal of Machine Learning Research, 2013.
- [10] 水谷: 分離可能想定下での非負行列分解に対する楕円丸め法, 京都大学 数理解析研究所 講究録, 2014