

分離可能想定下での非負行列分解に対する楕円丸め法

神奈川大学 工学部 情報システム創成学科
水谷友彦

Department of Information Systems Creation, Kanagawa University
Tomohiko Mizutani

(Email: mizutani@kanagawa-u.ac.jp)

概要

分離可能性という仮定の下での非負行列分解について考える。この行列分解は与えられたデータに対して、その凸包の全ての頂点を求める問題として解釈できる。[16] では、データにノイズが加わった状況を考え、それに対するアルゴリズムを提案した。本稿では主にその理論的な性能に関して解説を行う。

1 はじめに

非負行列分解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。**非負行列**とは全ての要素が非負であるような行列のことを言う。この問題の正確な記述は次のようになる。大きさが $d \times m$ の非負行列の集合を $\mathbb{R}_+^{d \times m}$ と書こう。今、 $A \in \mathbb{R}_+^{d \times m}$ と自然数 r が与えられているとする。このとき、 A を

$$A = FW + N$$

なる $d \times r$ の非負行列 F と $r \times m$ の非負行列 W の積の形で分解することを考える。ここで、 N は $d \times m$ の行列で分解の残差を表している。 $A \in \mathbb{R}_+^{d \times m}$ の非負行列分解とは、残差行列 N をできるだけ零行列に近づけるような $F \in \mathbb{R}_+^{d \times r}$ と $W \in \mathbb{R}_+^{r \times m}$ を求める問題である。非負行列分解は多くの応用を持っている。例えば顔画像分析 [15]、トピックモデリング [2, 3, 5]、文章クラスタリング [24, 18]、ハイパースペクトラル画像分析 [17, 9] などに应用され、その有効性が報告されている。

非負行列分解は NP 困難な問題である [23]。この問題の難しさに対して、Arora ら [2] は分離可能性という仮定を利用することを提案した。この仮定は Donoho-Stodden [6] が非負行列分解における分解の唯一性を議論する際に導入した三つの仮定の内の一つである。非負行列分解に分離可能性を仮定すると扱い易い問題になる。

非負行列 $A \in \mathbb{R}_+^{d \times m}$ が

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

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

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

と表すことができる。ここで $N^{(1)}$ と $N^{(2)}$ は $N\Pi^{-1} = (N^{(1)}, N^{(2)})$ を満たす N の部分行列でそれぞれの大きさは $d \times r$ と $d \times (m-r)$ である。本稿では $F + N^{(1)}$ をノイズを含む基底行列と呼ぶ。

本稿では分離可能性の下での非負行列分解を考える。これはノイズを含む分離可能な行列に対して、基底行列 (正確にはノイズを含む基底行列) に対応する列ベクトルを全て見つけるという問題である。正確には次のように記述される。

問題 1 データ行列 M はノイズを含む分離可能な $d \times m$ の行列とする。 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、 $M(I)$ ができるだけ基底 F に近くなるような I を求めよ。

ここで、 $M(I)$ は M の部分行列でその列ベクトルの添字集合が 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の列ベクトルをデータ点と呼ぶ。この問題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しい。

- **(正当性)** データ行列 M はノイズを含まない分離可能な行列であるとする。つまり 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 $M(I) = F$ となる I を出力する。
- **(頑強性)** データ行列 M はノイズを含む分離可能な行列であるとする。つまり M は分離可能な行列とノイズ行列 N の和で与えられているとする。もし $\|N\|_p < \epsilon$ なら、アルゴリズムは $\|M(I) - F\|_p < \tau\epsilon$ なる 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。

特に、頑強性において $\tau = 1$ となる場合、アルゴリズムはノイズを含む基底を特定できることを意味する。

[16] 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。提案手法では、単体に対して閉包楕円を描くとその頂点を見つけることができるという幾何的な性質を利用している。より正確には、単体に対してその体積最小閉包楕円を計算すると、単体の頂点はその境界上に載るという性質である。この性質の詳細は 2.2 節の性質 1 で述べる。提案手法のアルゴリズムの概要は次の通りである。まず、データ行列のデータ点に対して体積最小閉包楕円を計算する。次に、楕円の境界上に載っているデータ点の添字集合を出力する。上記の性質より、提案手法で分離可能な行列の基底行列が得られることが分かる。さらに、多少のノイズがデータ行列に加わったとしても、データ点に対する体積最小閉包楕円を計算することでノイズを含む基底行列を得ることができる。このことを支えている性質の詳細は 2.2 節の性質 2 で述べる。

本稿は次のように構成されている。2 節で提案手法のアルゴリズムを説明する。アルゴリズムの正当性、頑強性については 3 節で述べる。4 節で既存手法との比較について述べる。[16] に関して本稿では説明を省略した事柄について 5 節で触れる。

2 提案手法のアルゴリズム

まず、提案手法の正当性、頑強性を示すのに必要となる仮定について述べる。その後、手法の概要と詳細について述べる。

2.1 仮定

分離可能な行列に対して以下のような仮定を置く。この仮定の下で提案手法は正当性と頑強性を示すことができる。

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を選んでも、そのベクトルの 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。もし W の i 番目の列ベクトルがゼロベクトルであるなら、 A の i 番目の列ベクトルもゼロベクトルとなる。したがって、 A と W からゼロベクトルとなっている全ての列ベクトルを除く。すると、対角行列 D を用いて

$$A = FW \Leftrightarrow AD = FWD$$

とできる。ここで D の対角成分 d_{ii} は W の列ベクトル w_i の 1 ノルムの逆数 $1/\|w_i\|_1$ である。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非負行列分解においては、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。既存研究 [9] のアルゴリズムはこれら 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見ていこう。簡単のために、データ行列はノイズを含まず、ただの分離可能な行列 A であるとする。すると、仮定から $\text{conv}(A)$ は d 次元空間上の $(r-1)$ 次元単体となる。そして、この単体の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とする。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。したがって、ノイズを含まない場合と比べて、問題が難しくなっていることが分かる。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。これ以降では、簡単のために行列の列ベクトルの凸包のことを単に行列の凸包と呼ぶことにする。

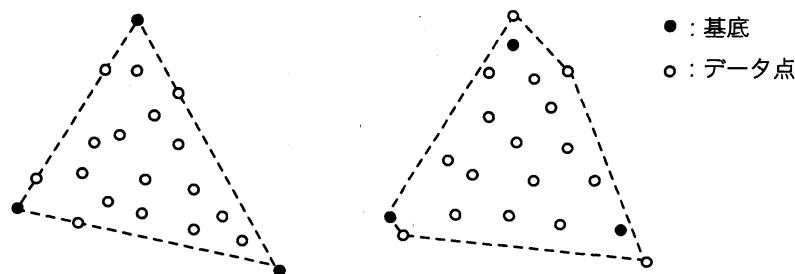


図 1: 仮定 1 の下での $(d, r) = (3, 3)$ における分離可能なデータ行列の凸包。左図はノイズが無い場合、右図はノイズが有る場合。

2.2 手法の概要

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。

性質 1 ([16] の Proposition 1) $S = \{\pm g_1, \dots, \pm g_r, \pm b_1, \dots, \pm b_\ell\}$ を含むような原点を中心に持つ体積最小な楕円を描くと、 Δ の頂点に対応している $\pm g_1, \dots, \pm g_r$ 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、 $(d, r) = (3, 3)$ の場合は八面体となる。図 2 は $(d, r) = (3, 3)$ の場合の 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。この性質は r 次元空間上の $(r-1)$ 次元単体の頂点はその閉包楕円を計算することで求められることを保証している。

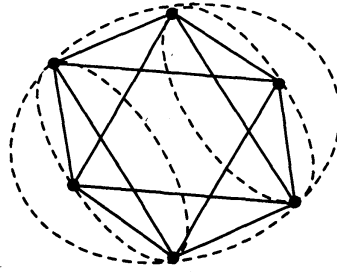


図 2: $(d, r) = (3, 3)$ の場合の S に対する体積最小閉包楕円。このとき S の凸包は八面体になる。

次に、 Δ の点 $b_1, \dots, b_\ell$ にノイズを加えて摂動させた点 $b'_1, \dots, b'_\ell$ を考えよう。 b_i は Δ 内の点なので、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 の要素は全て非負で足し合わせると 1 となる。つまり、 k_i は

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

を満たす。今、摂動させた点 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。このような $b'_1, \dots, b'_\ell$ に対しても、上と同様の性質が成り立つ。

性質 2 ([16] の Cororally 1) $S' = \{\pm g_1, \dots, \pm g_r, \pm b'_1, \dots, \pm b'_\ell\}$ を含むような原点を中心を持つ体積最小な楕円を描くと、 $\pm g_1, \dots, \pm g_r$ 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求められることを意味している。図 3 は $(d, r) = (3, 3)$ の場合の Δ とその閉包楕円の様子を図示したものである。左図はノイズが無い場合、右図は $\|k'\|_2 < 1$ を満たすようなノイズが加わった場合を表している。

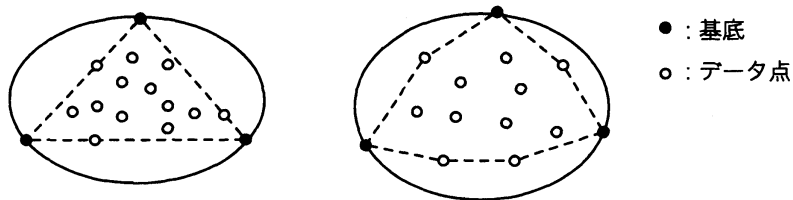


図 3: $(d, r) = (3, 3)$ の場合の Δ とその閉包楕円。(左図) ノイズが無い場合。(右図) $\|k'\|_2 < 1$ を満たすようなノイズが加わった場合。

次に提案手法のアルゴリズムの概要について述べる。前節では、仮定 1 の下では (1) の形の分離可能な行列 $A \in \mathbb{R}_+^{d \times m}$ の凸包は d 次元空間上の $(r-1)$ 次元単体となり、その頂点が基底に対応し

ていることを観察した。 A の凸包が存在する空間の次元は r 次元よりも高いが、特異値分解を利用すると、空間の次元を d から r に削減することができる。よって、性質1を利用することで、 A の基底行列 F を計算できることが分かる。また、性質2より、多少 A にノイズが加わったとしても、基底行列 F に近い行列(正確にはノイズを含む基底行列)が計算できる。

2.3 手法の詳細

提案手法の具体的なアルゴリズムを記述する。各ステップの詳細は以下の節で説明する。

Algorithm 1 楕円丸め法

入力: $M \in \mathbb{R}_+^{d \times m}$, $r \in \mathbb{N}$

出力: $\mathcal{I}$

- 1: M の特異値分解を計算し、サイズ r の縮小行列 $P \in \mathbb{R}^{r \times m}$ を構築する。
 - 2: P の列ベクトル $p_1, \dots, p_m \in \mathbb{R}^r$ に正負の符号を付けたベクトルから構成される集合 $\mathcal{S} = \{\pm p_1, \dots, \pm p_m\}$ を準備する。次に、 $\mathcal{S}$ 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 $\mathcal{I}$ を構築する。
-

2.3.1 ステップ1: 特異値分解の計算

このステップでは、特異値分解を経由して M のランク r 近似行列 M^r を計算し、それに対して直交変換を施すことで規模の小さい行列 P を構築する。 M を $d \times m$ の実行列とする。 M に特異値分解を施すと

$$M = U \Sigma V^\top$$

という形に分解することができる。ここで、 U と V は大きさがそれぞれ $d \times d$ と $m \times m$ の直交行列である。 Σ は以下のような形の $d \times m$ の対角行列である。

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_t) \in \mathbb{R}^{d \times m} \text{ with } \sigma_1 \geq \dots \geq \sigma_t \geq 0$$

幾つかの実数に対して $\text{diag}(\cdot)$ はそれを対角要素として持つ対角行列を表す。 $\sigma_1, \dots, \sigma_t$ は M の特異値で、 t は $\min\{d, m\}$ を表す。今、 Σ の t 個の特異値の内、大きいものから順に r 個の特異値 $\sigma_1, \dots, \sigma_r$ を選択し、残りのものは0と置いたものを Σ^r と書く。

$$\Sigma^r = \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) \in \mathbb{R}^{d \times m}$$

すると、 M のランク r 近似行列 M^r は

$$M^r = U \Sigma^r V^\top$$

として得られる。実際、 M^r は行列2ノルムの下での最適なランク r 近似行列となっていることが知られている。詳しくは例えば[10]のTheorem 2.5.3を参照せよ。直交行列 $U^\top$ を右から M^r に掛けると

$$U^\top M^r = \begin{pmatrix} P \\ 0 \end{pmatrix} \in \mathbb{R}^{d \times m}$$

となる。 P は $r \times m$ の行列でランクは r となっている。本稿ではこのような P を M に対するサイズ r の縮小行列と呼ぶ。

2.3.2 ステップ 2: 原点中心な体積最小閉包楕円の計算

まず楕円の定義とその体積について復習し、次に原点中心な体積最小閉包楕円の定式化と解法について説明する。 L は $r \times r$ の正定値行列、 z は r 次元ベクトルとする。 r 次元空間上の楕円は L と z を用いて、 $\{x \in \mathbb{R}^r : (x - z)^\top L (x - z) \leq 1\}$ と定義される。ここで、 L の固有ベクトルと固有値は楕円の形を規定する軸に対応し、 z は楕円の中心に対応する。この楕円の体積は $c(r)/\sqrt{\det L}$ となる。ここで $c(r)$ は r 次元の単位球の体積を表しており、次元 r に依存した実数である。

原点中心な体積最小閉包楕円の定式化について述べる。 m 個の r 次元ベクトル $p_1, \dots, p_m \in \mathbb{R}^r$ に対して正負の符号を付けたベクトルの集合 $\{\pm p_1, \dots, \pm p_m\}$ を S とする。この S に対する原点中心な体積最小閉包楕円は以下のような凸計画問題として定式化することができる。

$$\begin{aligned} \mathbb{Q}: \quad & \text{minimize} \quad -\log \det L, \\ & \text{subject to} \quad \langle p_i p_i^\top, L \rangle \leq 1, \quad i = 1, \dots, m, \\ & \quad \quad \quad L \succ 0 \end{aligned}$$

ここで、二つの行列に対する $\langle \cdot, \cdot \rangle$ はそれらのフロベニウス内積を表している。問題 $\mathbb{Q}$ の変数は行列 L である。 $\mathbb{Q}$ の最適解 L^* に対して、 $\{x \in \mathbb{R}^r : x^\top L^* x \leq 1\}$ が S に対する原点中心な体積最小閉包楕円となる。特に、 $\{x \in \mathbb{R}^r : x^\top L^* x = 1\}$ が楕円の境界である。ステップ 2 では $\mathbb{Q}$ の最適解 L^* を計算し、集合 S 中の要素 p_i に対して、 $p_i^\top L^* p_i = 1$ が成立するものを全て見つけて、その添字集合を $\mathcal{I}$ とする。

問題 $\mathbb{Q}$ を解くための手法としては二つの手法がよく知られている。一つは $\mathbb{Q}$ の双対問題に対して Frank-Wolfe 法 [12, 14, 20, 1] を利用するというもので、もう一つは内点法 [22, 11, 19, 21] である。理論的には、二つの手法はともに最適値に十分近い値を得るまでの反復回数は入力データの多項式で抑えられることが知られている。実用的には、内点法と切除平面法を組み合わせると $\mathbb{Q}$ を効率的に解けることが知られている [19, 1]。したがって、[16] におけるアルゴリズム 1 の実装としては、内点法と切除平面法を組み合わせた手法を利用することを提案している。

3 提案手法の正当性と頑強性

提案手法は正当性と頑強性を有する。これらの性質は厳密には以下のように記述される。

- (正当性; [16] の Theorem 1) 分離可能な行列 A は仮定 1 を満たすとする。 $(A, \text{rank}(A))$ 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、 $A(\mathcal{I}) = F$ を満たす。
- (頑強性; [16] の Theorem 2) 分離可能な行列 A とノイズ行列 N に対して、 $\tilde{A} = A + N$ と表されるノイズを含む分離可能な行列 $\tilde{A}$ を考える。 A は仮定 1 を満たすとする。 ϵ を

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

となるように取る。 $(\tilde{A}, \text{rank}(A))$ 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、もし $\|N\|_2 < \epsilon$ なら、 $\|\tilde{A}(\mathcal{I}) - F\|_2 < \epsilon$ を満たす。

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、

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

を表す。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。

この解析結果はアルゴリズムの頑強性に関する幾何的な直感とよく合っている。解析結果の ϵ はデータの凸包の潰れ具合を表す σ と基底とデータ点との離れ具合を表す $1 - \mu$ の積として与えられる。そして、これらの値が小さい場合、ノイズは小さくなる必要があることを主張している。 σ が小さいとは、図 4 の中図のように、データの凸包は平坦に近い形状になっていることを意味する。 $1 - \mu$ が小さいとは、図 4 の右図のように、データの凸包の頂点とデータ点への距離が近いことを意味する。これらの場合、ノイズが加わると、凸包の形状は大きく変わると思われる。一方で、図 4 の左図のように、凸包の形状は等方的に膨らみを持ち、かつ、凸包の頂点とデータ点への距離が十分離れている場合、多少ノイズが加わったとしても、凸包の形状は大きくは変わらないと思われる。したがって、凸包の頂点はノイズが加わったとしても、頂点であり続けられる可能性がある。

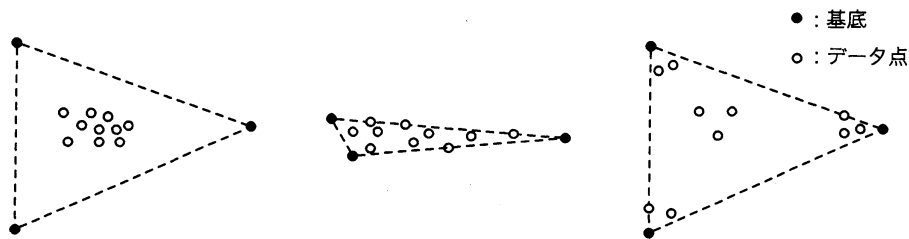


図 4: データの凸包の形状の例。(左図) ノイズの影響を受けにくい形状。(中図、右図) ノイズの影響を受けやすい形状。

4 既存手法との比較

分離可能性の下での非負行列分解に対するアルゴリズムとしては主に AGKM[2]、Hottopixx[4, 7, 8]、SPA[9]、XRAY[13] がよく知られている。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。四つのアルゴリズムとも正当性は保証されている。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。

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 [2]	○	○	≥ 1	m 個の $m - 1$ 変数の LP
Hottopixx [4, 7, 8]	○	○	1	m^2 変数の LP
SPA [9]	○	○	≥ 1	軽い
XRAY [13]	○	×	—	比較的軽い

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。[4] では、 $p = 1$ とし

て、 ϵ を

$$\epsilon = \frac{\alpha \min\{d_0, \alpha\}}{9(r+1)}$$

とすると、ある仮定の下で Hottopixx は $\tau = 1$ の頑強性を持つことを証明している。ここで α はデータの凸包の潰れ具合を表す値である。具体的には、基底 f_i とそれ以外の基底 $f_j, j \in \{1, \dots, r\} \setminus \{i\}$ の凸包との ℓ_1 距離の中で最小のものを表している。 d_0 は基底とデータ点との離れ具合を表す値である。具体的には基底とデータ点との ℓ_1 距離の中で最小のものを表している。図5は $(d, r) = (3, 3)$ の場合の α と d_0 の様子を表している。

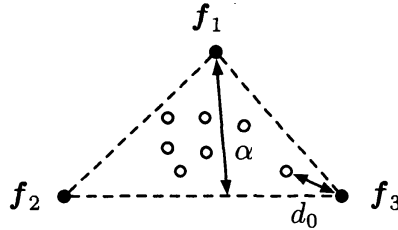


図 5: $(d, r) = (3, 3)$ の場合の α と d_0 の様子

提案手法と Hottopixx の頑強性解析の結果を比較する。二つの手法の ϵ は共に「データの凸包の潰れ具合を表す値」と「基底とデータ点との離れ具合を表す値」の積の形で表されており、特に、Hottopixx はこの積を分離可能な行列のランク r で割った形になっている。Hottopixx の ϵ は $1/r$ という項を持っているので、 r が大きくなると ϵ の値は小さくなる。一方で、提案手法の ϵ においては $1/r$ は含まれていない。したがって、提案手法の頑強性は分離可能な行列のランクには依存しないことを意味している。しかし、提案手法は Hottopixx よりも多少強い仮定を置く必要がある。提案手法は仮定 1 を満たす必要がある。特に、仮定 1-b はデータ行列の凸包が単体になることを要求する。一方で、Hottopixx は 1-b の代わりにより弱い仮定を置く。具体的には、ある基底は残りの基底の凸結合で書けないという仮定である。つまり、データ行列の凸包が多面体になることを要求する仮定である。

次に提案手法と Hottopixx の計算コストを比較する。Hottopixx は次のような事実に基づいてアルゴリズムを設計している。(1) の形の分離可能な行列 A に対して、以下のような $m \times m$ の行列 C を考える。

$$C = \Pi^{-1} \begin{pmatrix} I & K \\ 0 & 0 \end{pmatrix} \Pi \in \mathbb{R}^{m \times m}$$

この C は $AC = C$ を満たす。そして、対角要素が 1 となる (i, i) 成分が A における基底ベクトルの添字 i に対応している。Hottopixx はこのような C を LP の変数として定式化する。したがって、Hottopixx は m^2 変数を持つ LP を解く必要がある。一般的に LP は解きやすい問題であるが、変数の個数が多くなると解くことが困難になる。実問題では m が数万程度、できれば数十万程度のデータ行列を扱う必要がある。このような場合、Hottopixx を実行することは難しくなる。一方で、提案手法の主な計算コストは $d \times m$ 行列に対する特異値分解の計算と体積最小閉包楕円の計算である。2.3.2 節で述べたように、体積最小閉包楕円の計算は m 変数の凸計画問題として定式化される。その解法として切除平面法を利用すると、実際には m 変数よりかなり少ない変数の凸計画問題を複数回解くことで、効率的に体積最小閉包楕円の計算を行うことができることが知られている [19, 1]。実際、 m が数万、数十万程度のデータ行列に対して閉包楕円の計算を実行することができる。

AGKM と SPA は $\tau \geq 1$ の頑強性が保証されている。実際の性能としては、SPA と XRAY が優れていると考えられる。[9, 13] において、これらの手法は大規模なデータ行列に対しても動作し、また、優れた頑強性を示すことが実験的に報告されている。

5 おわりに

[16] に関して本稿で説明を省略した事柄について触れる。頑強性解析の結果から、提案手法はノイズの大きさがある値以下ならノイズを含む分離可能な行列から基底に対応している行列を特定できることが保証される。ここで、頑強性解析で保証されている値よりもノイズが大きくなる場合を考えてみよう。この場合、データ点に対する体積最小閉包楕円はその境界上に r 個よりも多くの個数のデータ点が載る可能性がある。したがって、楕円の境界上の複数のデータ点から r 個の点を適切に選択する必要がある。この選択に関して [16] では既存手法である SPA や XRAY を利用することを提案している。まず、アルゴリズム 1 を利用して、基底と成り得るデータ点の候補を限定し、その後、既存手法を利用して適切に r 個のデータ点を選択するというアルゴリズムを提案している。[16] では、この実践的なアルゴリズムと既存手法である SPA と XRAY の性能を実験を通して評価した結果を報告している。

参考文献

- [1] S.D. Ahipasaoglu, P. Sun, and M.J. Todd. Linear convergence of a modified Frank-Wolfe algorithm for computing minimum-volume enclosing ellipsoids. *Optimization Methods and Software*, 23(1):5–19, 2008.
- [2] S. Arora, R. Ge, R. Kannan, and A. Moitra. Computing a nonnegative matrix factorization – provably. In *Proceedings of the 44th symposium on Theory of Computing (STOC)*, pages 145–162, 2012.
- [3] S. Arora, R. Ge, and A. Moitra. Learning topic models – Going beyond SVD. In *Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1–10, 2012.
- [4] V. Bittorf, B. Recht, C. Re, and J.A. Tropp. Factoring nonnegative matrices with linear programs. In *Advances in Neural Information Processing Systems 25 (NIPS)*, pages 1223–1231, 2012.
- [5] W. Ding, M.H. Rohban, P. Ishwar, and V. Saligrama. Topic discovery through data dependent and random projections. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013.
- [6] D. Donoho and V. Stodden. When does non-negative matrix factorization give a correct decomposition into parts? In *Advances in Neural Information Processing Systems 16 (NIPS)*, pages 1141–1148, 2003.
- [7] 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.
- [8] N. Gillis and R. Luce. Robust near-separable nonnegative matrix factorization using linear optimization. arXiv:1302.4385v1, 2013.
- [9] 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.

- [10] G.H. Golub and C.F.V. Loan. *Matrix Computation*. The Johns Hopkins University Press, 3rd edition, 1996.
- [11] R.H. Tütüncü K.C. Toh, M.J. Todd. SDPT3 – a MATLAB software package for semidefinite programming. *Optimization Methods and Software*, 11:545–581, 1999.
- [12] L.G. Khachiyan. Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21(2):307–320, 1996.
- [13] A. Kumar, V. Sindhwani, and P. Kambadur. Fast conical hull algorithms for near-separable non-negative matrix factorization. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013.
- [14] P. Kumar and E.A. Yildirim. Minimum-volume enclosing ellipsoids and core sets. *Journal of Optimization Theory and Applications*, 126(1), 2005.
- [15] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
- [16] 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.
- [17] J.M.P. Nascimento and J.M.B. Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4):898–910, 2005.
- [18] F. Shahnaz, M.W. Berry, V.P. Pauca, and R.J. Plemmons. Document clustering using nonnegative matrix factorization. *Information Processing and Management*, 42(2):373–386, 2006.
- [19] P. Sun and R.M. Freund. Computation of minimum-volume covering ellipsoids. *Operations Research*, 52(5):690–706, 2004.
- [20] M.J. Todd and E.A. Yildirim. On Khachiyan’s algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics*, 155(13):1731–1744, 2007.
- [21] T. Tsuchiya and Y. Xia. An extension of the standard polynomial-time primal-dual path-following algorithm to the weighted determinant maximization problem with semidefinite constraints. *Pacific Journal of Optimization*, 3(1):165–182, 2007.
- [22] L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2):499–533, 1998.
- [23] S.A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal of Optimization*, 20(3):1364–1377, 2009.
- [24] W. Xu, X. Liu, and Y. Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR)*, pages 267–273, 2003.