

低ランク最適化のための高速特異値閾値処理の数理

Mathematical Derivation of Fast Singular Value Thresholding for Low-rank Optimization

佐々木崇元[†]
Takayuki Sasaki

北原正樹[†]
Masaki Kitahara

清水淳[†]
Atsushi Shimizu

1. はじめに

データや物理現象に潜在する低ランク性¹に基づくデータ解析手法(以降, 低ランクモデリング)が, コンピュータビジョンや画像処理, ゲノムデータ解析などの多くの分野で近年活発に研究されている[1–10]. 低ランクモデリングでは着目した行列の階数(rank)を最小化する数理計画問題を立式し, この解を求めて所望の解析を実現する.

関数 rank は不連続で微分不能, 非凸であり, これに基づく計画問題は NP 困難な組合せ最適化となる. このため, rank 関数の代わりに核型ノルムを正則化する緩和アプローチが広く用いられる. 核型ノルムは rank 関数の凸包絡である[11]ため, 核型ノルムを最小化することで間接的に低ランク性を高められる. 核型ノルムを含む最適化問題は Alternating Direction Method of Multipliers (ADMM)[12]等の1次法の反復計算で解を求められ, 核型ノルムの近接写像である特異値閾値処理(Singular Value Thresholding, 以降 SVT)[3]が繰り返し実行される.

しかし, SVT 算出には計算量の大きい特異値分解(Singular Value Decomposition, 以降 SVD)の算出が必要なため, 解析結果を得るのに多くの時間を要する. ここで, 各解析法で立式される核型ノルム正則化問題を(1)少数の大型行列を正則化する問題と(2)多数の小型行列を正則化する問題に分類する(表1). 前者(1)の場合には, 計算量を抑えて SVT を高速化する手法がいくつか提案されている. Caiら[13]は行列を事前に Complete Orthogonal Decomposition (COD) した後ニュートン法で反復更新し, SVD を行わずに SVT を求める Fast SVT (FSVT) を提案している. また Ohら[14]は大型行列を直交行列と小型のコア行列の積に近似することで, SVD の入力サイズを小さくして高速化する Fast Randomized SVT (FRSVT) を提案している. いずれの手法も入力行列のサイズが大きい(行数, 列数=500~2000)ときに計算量を抑え, 大幅な速度改善結果を示している.

一方で後者(2)の場合, 上記手法による高速化の効果は限定的であると考えられる. FSVT は入力小型の場合, 直接の SVD 計算と比較して COD とニュートン法の計算量が多い. また FRSVT は入力小型の場合, コア行列の縮退効果が小さく高速化できず, また近似法のため計算誤差が多い. 加えてこれらの手法は, 多数の行列を同時に処理するデータ並列のアプ

表 1: 低ランクモデリングの類型と高速化法

	(1) 少数の大型行列	(2) 多数の小型行列
用途	ロバスト主成分分析 [1, 2] 欠損値推定・補間 [3, 4] オプティカルフロー推定 [5, 6] ダイナミック MRI 解析 [7] ゲノムデータ解析 [8]	グラフ単純化 [9] 偽色除去 [10]
高速化	FSVT [13], FRSVT [14]	FMSVT(2×2) [15] FMSVT(M×2, 本稿)

ローチが取れず, 近年の並列アーキテクチャの計算資源を活用できない.

そこで筆者らは, 計算量を抑えながらデータ並列に SVT を算出する Fast Multiple SVT (FMSVT) を提案している[15]. FMSVT は核型ノルムを特異値を用いずに $L_{\infty, 2}$ 混合ノルムで表現することで, SVD が不要な SVT 計算を実現し, 計算量を削減している. 加えてアルゴリズムを容易にデータ並列化でき, 多数の行列について同時処理が可能である. 本稿では[15]で述べた 2×2 行列の FMSVT を拡張して, $M \times 2$ あるいは $2 \times N$ 行列に対する SVT アルゴリズムを検討する. [15]では論じなかった数理的な導出を示し, 実験によりアルゴリズムの速度と精度を明らかにする.

2. 核型ノルム正則化に基づく低ランクモデリング

本稿が対象にする核型ノルム正則化問題について整理する. 行列 $\mathbf{X} \in \mathbb{R}^{M \times N}$ の特異値を $\sigma_i(\mathbf{X}), i=1, \dots, K = \min(M, N)$ と表すとき, 核型ノルムは特異値の和,

$$\|\mathbf{X}\|_* = \sum_{i=1}^K \sigma_i(\mathbf{X}) \quad (1)$$

と定義される. 核型ノルムは関数 $\text{rank}(\mathbf{X})$ の凸包絡である[11]ため, 核型ノルムを正則化することで低ランク性を推進できる.

本稿では多数の小型行列が核型ノルムで正則化される最適化問題を取り扱う. 典型的にはヒルベルト空間 \mathcal{X} の変数 $\mathbf{x} \in \mathcal{X}$ に関する下記形式の最適化問題である:

$$\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) + \lambda \sum_{i=1}^L \|\Phi_i(\mathbf{x})\|_* \quad (2)$$

ここで目的関数第2項は核型ノルムによる低ランク性の正則化関数で, 第1項はその他の正則化や忠実化関数である. 関数 $\Phi_i: \mathcal{X} \rightarrow \mathbb{R}^{M \times N}$ は低ランク性を規約する行列の生成関数で, 典型的には線形写像である.

例えばグラフ単純化[9]では, M 次元空間に埋め込まれたグラフ $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ の頂点座標 $\mathbf{x} \in \mathbb{R}^{M \times |\mathcal{V}|}$ を変数とし, 隣接辺のベクトルを2つ1組で行列に押し込め, $\Phi_i(\mathbf{x}) \in \mathbb{R}^{M \times 2}$ を作成する. この行列の階数はグラ

[†]日本電信電話株式会社 メディアインテリジェンス研究所
NTT Media Intelligence Laboratories

¹データや物理現象から導かれる行列 $\mathbf{X} \in \mathbb{R}^{M \times N}$ について, その階数 $\text{rank}(\mathbf{X})$ が $\min(M, N)$ と比較して小さい性質.

フの曲折回数に比例することに着眼し、核型ノルムの和 $\sum_{i=1}^L \|\phi_i(\mathbf{x})\|_*$ を正則化することで各頂点を局所線形に整列させている。

問題 (2) を解くためには関数 f と線形写像 ϕ_i に応じたアルゴリズムを選択するが、[9,10] では f が微分可能あるいは近接写像計算可能² であるため、ADMM [12] や Primal Dual Splitting [16] 等の 1 次法を用いている。1 次法は最適解への収束列を生成する反復計算手続きで、各反復では関数の勾配、近接写像などの 1 次情報に基づき解を更新し、目的関数を最小化する。

問題 (2) の核型ノルム正則化関数については、補助変数 $\mathbf{Y}_i = \phi_i(\mathbf{x})$, $i = 1, \dots, L$ を導入し、 $g(\mathbf{Y}_1, \dots, \mathbf{Y}_L) = \sum_{i=1}^L \|\mathbf{Y}_i\|_*$ の近接写像を計算すれば良い。各 $\|\mathbf{Y}_i\|_*$ 独立に計算可能で、

$$\text{prox}_{\mu g}(\mathbf{Y}_1, \dots, \mathbf{Y}_L) = \left(\text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y}_1), \dots, \text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y}_L) \right) \quad (3)$$

が成立する。右辺各要素は核型ノルム $\|\cdot\|_*$ の近接写像で、特異値閾値処理 (SVT) に等しい [3] ことが知られている。SVT は入力行列を \mathbf{Y} 、閾値を $\mu > 0$ として以下の式で算出する。

$$\text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y}) = \mathbf{U}(\boldsymbol{\Sigma} - \mu \mathbf{I})_+ \mathbf{V}^T \quad (4)$$

$$\text{where } (\mathbf{U}, \boldsymbol{\Sigma}, \mathbf{V}) = \text{SVD}(\mathbf{Y}) \quad (5)$$

ここで関数 $\text{SVD}(\cdot)$ は特異値分解 (SVD)³ で、 $(\cdot)_+$ は入力の各要素を非負値にクリッピングするランプ関数である。即ち、関数 $g(\mathbf{Y}_1, \dots, \mathbf{Y}_L)$ の近接写像 (3) は L 回の SVT (式 (4)) により算出でき、それぞれの SVT では小型行列を入力とする SVD 算出 (式 (5)) が必要である。1 次法で反復演算する際、計算時間の多くは計算量の大きい SVD 算出に費やされる。

さて式 (3) は各 L 個の行列 $\mathbf{Y}_1, \dots, \mathbf{Y}_L$ について依存関係が無いため、タスク並列処理が可能である。しかし本稿が扱う問題では各行列のサイズが小さいため、処理割当やメモリロードのオーバーヘッドが相対的に大きく、改善の効果は限定的である。

そこでオーバーヘッドが少ない、データ並列処理について考える。アルゴリズムをデータ並列化できれば、Single Instruction Multiple Data (SIMD) 等のデータ並列アーキテクチャを用いる実装で高速化できる。しかし SVD の算出は逐次的かつ行列要素の参照位置や処理内容が入力 \mathbf{Y}_i に依存する [17] ため、行列間で共通に計算できる処理が少なく、データ並列化は本質的に困難である。

以上より、多数の小型行列を低ランクに正則化するモデルは最適解を高速に得るのが困難で、その原因は

² 関数 g と正実数 $\mu > 0$ に対し、点 \mathbf{y} での近接写像とは $\text{prox}_{\mu g}(\mathbf{y}) = \arg \min_{\mathbf{z}} g(\mathbf{z}) + \frac{1}{2\mu} \|\mathbf{z} - \mathbf{y}\|^2$ と定義される。核型ノルムを含むいくつかの関数は近接写像の効率的な計算法が知られており、そのような関数を近接写像計算可能 (proximable) であると呼ぶ。

³ 本稿での特異値分解は "thin SVD" とする。即ち $\mathbf{Y} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ と分解した時、行列 \mathbf{U}, \mathbf{V} は $K = \min(M, N)$ 個の正規直交なベクトルから成り、 $K+1$ 番目以降の特異ベクトルの算出は省略する ($\mathbf{U} \in \mathbb{R}^{M \times K}$, $\mathbf{V} \in \mathbb{R}^{N \times K}$)。また行列 $\boldsymbol{\Sigma}$ は $K \times K$ 対角行列である。MATLAB では $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\mathbf{Y}, 'econ');$ 。

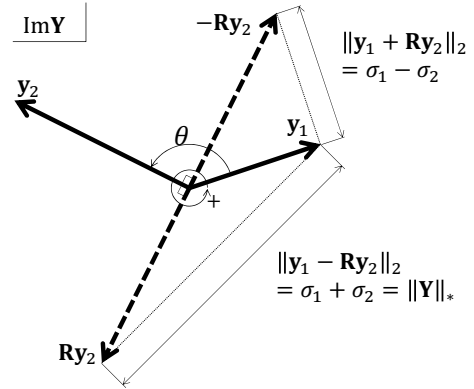


図 1: 核型ノルムの幾何的性質

SVD の多大な計算量と並列化の難しさにある。

3. Fast Multiple SVT

本章では式 (3) の多数の SVT 計算を高速に算出する、FMSVT [15] を導出する。この手法は核型ノルムがある部分空間上のベクトルの距離で特徴付けられるという、幾何的性質の発見に基づいて導かれる。この性質により、核型ノルムの特異値を用いずに表現 (命題 3.1, 系 3.2) でき、計算量が大きい SVD を使わない新しい SVT の表現 (命題 3.3) を獲得する。加えてこの SVT はほとんどが線形変換で記述 (命題 3.4) できるため、データ並列なアルゴリズムを導ける。

続く 3.1 節では核型ノルムの幾何的性質と新しい SVT 表現に関する命題を導き、3.2 節で式を展開して SVT 算出式 (定理 3.5, 3.6) を示す。最後に 3.3 節でデータ並列化したアルゴリズムを示し、これを FMSVT とする。以上より得られる FMSVT は式 (4), (5) による算出と比較して計算量が削減されており、かつデータ並列なアルゴリズムである。

本稿では [15] で 2×2 に限定していた入力行列のサイズを、 $M \times 2$ 及び $2 \times N$ に拡張する。なお $\text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y}) = (\text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y}^T))^T$ であるから、入力が $2 \times N$ 行列の SVT は $N \times 2$ サイズの SVT 算出の前後に転置処理を施すのみで実現できる。従って以降は、縦長の $M \times 2$ 行列の SVT について検討する。

3.1. 核型ノルムの幾何的性質と SVT の新しい表現

FMSVT 導出の中核となる、核型ノルムの幾何的性質と SVD 不要な SVT の表現について述べる。本節では結果のみを提示し、証明は付録に記す。以降、単一の入力行列を $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2] \in \mathbb{R}^{M \times 2}$, $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^M$ とし、その特異値を $\sigma_1, \sigma_2 (\sigma_1 \geq \sigma_2 \geq 0)$ とする。

命題 3.1. 特異値和 $\sigma_1 + \sigma_2$ と特異値差 $\sigma_1 - \sigma_2$ は、

$$\sigma_1 \pm \sigma_2 = \sqrt{\text{tr} \mathbf{Y}^T \mathbf{Y} \pm 2\sqrt{\det \mathbf{Y}^T \mathbf{Y}}} = \|\mathbf{y}_1 \mp \mathbf{R}\mathbf{y}_2\|_2 \quad (6)$$

である。ここで $\mathbf{R} \in \mathbb{R}^{M \times M}$ は回転行列で、像 $\text{Im} \mathbf{Y}$ ⁴ 上のベクトルを原点回りに、 $\text{Im} \mathbf{Y}$ に沿って $\pi/2$ [rad] 回転させる⁵。ただし回転方向の正負については、図 1 のように \mathbf{y}_1 から \mathbf{y}_2 に最短で辿り着く方向を、回転の正方向とする。

⁴ 行列 $\mathbf{A} \in \mathbb{R}^{M \times N}$ に対し部分空間 $\text{Im} \mathbf{A} = \{\mathbf{A}\mathbf{x} | \mathbf{x} \in \mathbb{R}^N\} \subset \mathbb{R}^M$ を \mathbf{A} の像という。

証明は付録 A に記す. 命題 3.1 では特異値の和や差が, ベクトル \mathbf{y}_1 と $\pm \mathbf{R}\mathbf{y}_2$ との間の, L_2 距離であると主張している (図 1). またトレースや行列式で構成される中間式は, $\mathbf{Y}^T \mathbf{Y}$ が 2×2 行列であるため, 算出が容易である.

命題 3.1 と核型ノルムの定義式 (1) から, 直ちに以下の系を求められる.

系 3.2. 核型ノルム $\|\mathbf{Y}\|_*$ は,

$$\|\mathbf{Y}\|_* = \|\mathbf{B}\text{vec}(\mathbf{Y})\|_{\infty,2} \quad (7)$$

と表現できる. ただし行列 $\mathbf{B} \in \mathbb{R}^{2M \times 2M}$ は

$$\mathbf{B} = \begin{bmatrix} \mathbf{I} & -\mathbf{R} \\ \mathbf{I} & \mathbf{R} \end{bmatrix} \quad (8)$$

とする. また $\|\cdot\|_{\infty,2}$ は $L_{\infty,2}$ 混合ノルム⁶ で, 関数 vec は行列をベクトル化⁷ する関数である.

系 3.2 は, 核型ノルムを特異値を使わずに線形変換 $\mathbf{B}\text{vec}(\cdot)$ と $L_{\infty,2}$ 混合ノルムの合成で表現できると主張している. これは核型ノルムの近接写像 (即ち SVT) における SVD 排除の可能性を示唆している. 事実, 系 3.2 と付録の補題 B.1, B.2 を組み合わせ, 次の命題が得られる.

命題 3.3. 行列 \mathbf{Y} の SVT について,

$$\text{prox}_{\mu, \|\cdot\|_*}(\mathbf{Y}) = \text{vec}^T \left(\mathbf{B}^T \text{prox}_{\mu, \|\cdot\|_{\infty,2}} \left(\frac{1}{2} \mathbf{B}\text{vec}(\mathbf{Y}) \right) \right) \quad (9)$$

が成立する.

証明は付録 B に記す. これより, 線形変換 $\mathbf{B}\text{vec}(\cdot)$ と $L_{\infty,2}$ ノルムの近接写像 $\text{prox}_{\mu, \|\cdot\|_{\infty,2}}(\cdot)$ の合成という, SVT の新しい表現が得られた. ここで近接写像 $\text{prox}_{\mu, \|\cdot\|_{\infty,2}}(\cdot)$ は非線形であるが, 入力依存の対角行列による線形変換として記述ができる.

命題 3.4. $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^M$ に対し,

$$\text{prox}_{\mu, \|\cdot\|_{\infty,2}}(\mathbf{x}) = \begin{bmatrix} k_1 \mathbf{x}_1 \\ k_2 \mathbf{x}_2 \end{bmatrix} = \text{diag}(k_1, k_1, k_2, k_2) \mathbf{x} \quad (10)$$

where

$$k_i = \begin{cases} \frac{(\|\mathbf{x}_1\|_2 + \|\mathbf{x}_2\|_2 - \mu)_+}{2\|\mathbf{x}_1\|_2} & \text{if } \|\mathbf{x}_1\|_2 - \|\mathbf{x}_2\|_2 \leq \mu \\ 1 - \frac{\mu}{\|\mathbf{x}_1\|_2} & \text{if } \|\mathbf{x}_1\|_2 - \|\mathbf{x}_2\|_2 > \mu \\ 1 & \\ 1 - \frac{\mu}{\|\mathbf{x}_2\|_2} & \text{if } \|\mathbf{x}_1\|_2 - \|\mathbf{x}_2\|_2 < -\mu \end{cases} \quad (11)$$

⁵ 任意の $\mathbf{y} \in \text{Im} \mathbf{Y}$ について, (1) $\mathbf{R}\mathbf{y} \in \text{Im} \mathbf{Y}$ (2) $\|\mathbf{R}\mathbf{y}\|_2 = \|\mathbf{y}\|_2$ (3) $\mathbf{y}^T \mathbf{R}\mathbf{y} = 0$ (4) $\mathbf{R}^2 \mathbf{y} = -\mathbf{y}$ (5) $\mathbf{R}^T \mathbf{R}\mathbf{y} = \mathbf{R}\mathbf{R}^T \mathbf{y} = \mathbf{y}$ が成立する.

⁶ $L_{\infty,2}$ 混合ノルムは L_{∞} ノルムと L_2 ノルムの合成関数で, 入力 $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T$ に対し, $\|\mathbf{x}\|_{\infty,2} = \max(\|\mathbf{x}_1\|_2, \|\mathbf{x}_2\|_2)$ である.

⁷ 関数 vec は入力行列を $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2] \in \mathbb{R}^{M \times 2}$ を並べ替えて列ベクトル $[\mathbf{y}_1^T, \mathbf{y}_2^T]^T \in \mathbb{R}^{2M}$ を出力する線形変換である. 後の vec^T はこの逆変換で, $\mathbf{Y} = \text{vec}^T(\text{vec}(\mathbf{Y}))$ とする.

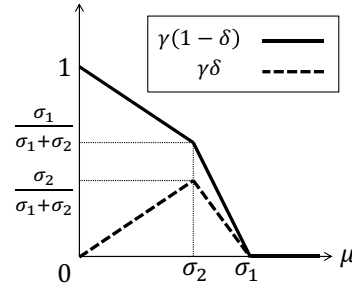


図 2: SVT の線形結合表現の係数

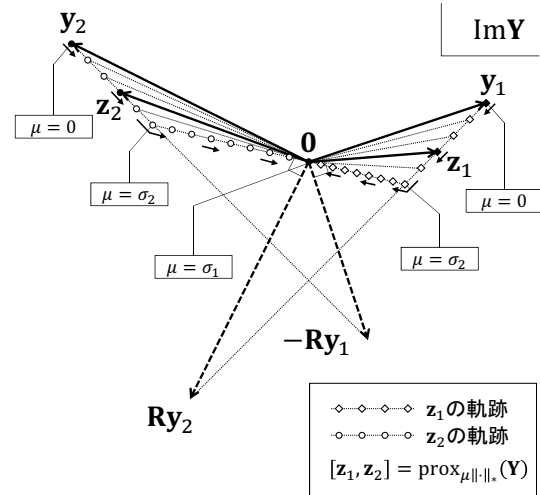


図 3: 閾値 μ の変化による SVT の軌跡

が成立する⁸.

この命題の証明は Moreau 直交分解 [18] と $L_{1,2}$ 混合ノルム球への射影 [19] に基づく (付録 C).

以上より係数 k_1, k_2 の算出 (11) を除けば, SVT は全て線形変換で表現できることが明らかとなった. 続く節では上記式を展開し, SVT 算出式とアルゴリズムを提示する.

3.2. SVT 算出法

前節の命題から, SVT を算出する下記の定理を得る.

定理 3.5. $\text{rank} \mathbf{Y} = 2$ かつ $\sigma_1 \neq \sigma_2$ のとき,

$$\text{prox}_{\mu, \|\cdot\|_*}(\mathbf{Y}) = \gamma(1 - \delta)\mathbf{Y} + \gamma\delta\mathbf{R}\bar{\mathbf{Y}} \quad (12)$$

where

$$\gamma = \left(1 - \frac{(\mu - \sigma_2)_+}{\sigma_1 - \sigma_2}\right)_+, \delta = \frac{\min(\mu, \sigma_2)}{\sigma_1 + \sigma_2}, \bar{\mathbf{Y}} = [\mathbf{y}_2, -\mathbf{y}_1] \quad (13)$$

が成立する.

上記定理は, 式 (9) の prox 計算を式 (10) で展開し, さらに式 (6) と (11) を適用して得られる. 単純な変形と場合分けで確認できるため, 証明を省略する.

定理 3.5 は, SVT を入力行列 \mathbf{Y} と行列 $\mathbf{R}\bar{\mathbf{Y}}$ の線形結合から算出できると主張している. 結合係数 $\gamma(1 - \delta), \gamma\delta$ は振幅パラメータ γ と内分比パラメータ δ から構成され, いずれも SVT の閾値 μ の関数である. 係数 $\gamma(1 - \delta), \gamma\delta$ の関数形を図 2 にプロットする.

⁸ $\mathbf{x}_i = \mathbf{0}$ のとき $k_i \mathbf{x}_i = (0/0)\mathbf{0}$ となるが, 例外的に $k_i \mathbf{x}_i = \mathbf{0}$ とする.

この係数に基づき μ の増加による $\text{prox}_{\mu\|\cdot\|_*}(\mathbf{Y})$ の軌跡を図 3 に示す. $[\mathbf{z}_1, \mathbf{z}_2] = \text{prox}_{\mu\|\cdot\|_*}(\mathbf{Y})$ とすることで, $\text{Im}\mathbf{Y}$ 上のベクトル $\mathbf{z}_1, \mathbf{z}_2$ の軌跡を表している. $0 < \mu \leq \sigma_2$ では \mathbf{Y} と $\mathbf{R}\bar{\mathbf{Y}}$ の内分となり, $\sigma_2 < \mu < \sigma_1$ では振幅が線形に減少し, $\mu \geq \sigma_1$ ではゼロ行列 \mathbf{O} となる.

定理 3.5 を適用するには, $\pi/2[\text{rad}]$ 回転行列 \mathbf{R} による変換 $\mathbf{R}\bar{\mathbf{Y}}$ の具体的な値を求める必要がある. この求め方を, $M \geq 3$ の場合と $M=2$ の場合の 2 通りに分けて述べる. まず $M \geq 3$ の場合, 行列 \mathbf{R} は脚注 5 の全ての条件を満たし, 命題 3.1 に述べた回転方向である必要がある. これらを満たす行列 \mathbf{R} は

$$\mathbf{R} = \frac{1}{\sqrt{\det \mathbf{Y}^T \mathbf{Y}}} (\mathbf{y}_2 \mathbf{y}_1^T - \mathbf{y}_1 \mathbf{y}_2^T) \quad (14)$$

であることを見出した. しかし式 (14) で \mathbf{R} を算出してから行列積 $\mathbf{R}\bar{\mathbf{Y}}$ を求める手順では, $\mathcal{O}(M^2)$ オーダーの計算量が必要である. そこで内積を先に展開し, 計算順序を入れ替えて,

$$\mathbf{R}\bar{\mathbf{Y}} = \frac{1}{\sqrt{\det \mathbf{Y}^T \mathbf{Y}}} \mathbf{Y} \begin{bmatrix} -\mathbf{y}_2^T \mathbf{y}_2 & \mathbf{y}_1^T \mathbf{y}_2 \\ \mathbf{y}_1^T \mathbf{y}_2 & -\mathbf{y}_1^T \mathbf{y}_1 \end{bmatrix} \quad (15)$$

と計算する. これにより, 計算量オーダーを $\mathcal{O}(M)$ に削減できる. 加えて式 (12) については \mathbf{Y} の積で括り,

$$(12) = \mathbf{Y} \left(\gamma(1-\delta)\mathbf{I}_2 + \frac{\gamma\delta}{\sqrt{\det \mathbf{Y}^T \mathbf{Y}}} \begin{bmatrix} -\mathbf{y}_2^T \mathbf{y}_2 & \mathbf{y}_1^T \mathbf{y}_2 \\ \mathbf{y}_1^T \mathbf{y}_2 & -\mathbf{y}_1^T \mathbf{y}_1 \end{bmatrix} \right) \quad (16)$$

とすれば, さらに計算量を削減できる.

続いて $M=2$ の場合, 符号関数 sgn を用いて

$$\mathbf{R} = \text{sgn}(\det \mathbf{Y}) \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad (17)$$

の成立は自明である. 即ち要素の入れ替えと符号反転で $\mathbf{R}\bar{\mathbf{Y}}$ を算出でき,

$$\mathbf{R}\bar{\mathbf{Y}} = \text{sgn}(\det \mathbf{Y}) \begin{bmatrix} -y_{22} & y_{21} \\ y_{12} & -y_{11} \end{bmatrix}, \text{ where } \mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \quad (18)$$

である.

以上により定理 3.5 を用いて SVT を求める方法を述べた. しかしこの定理は $\text{rank}\mathbf{Y}=2$ かつ $\sigma_1 \neq \sigma_2$ という条件が課されている. これ以外の場合は下記の通り, μ より定まる振幅を \mathbf{Y} に乗じて, 容易に算出できる.

定理 3.6. 「 $\text{rank}\mathbf{Y}=2$ かつ $\sigma_1 \neq \sigma_2$ 」以外のとき,

$$\text{prox}_{\mu\|\cdot\|_*}(\mathbf{Y}) = \begin{cases} \mathbf{O} & \text{if } \text{rank}\mathbf{Y}=0 \\ \left(1 - \frac{\mu}{\|\mathbf{Y}\|_F}\right)_+ \mathbf{Y} & \text{if } \text{rank}\mathbf{Y}=1 \\ \left(1 - \frac{\sqrt{2}\mu}{\|\mathbf{Y}\|_F}\right)_+ \mathbf{Y} & \text{if } \text{rank}\mathbf{Y}=2 \text{ かつ } \sigma_1 = \sigma_2 \end{cases} \quad (19)$$

ここで $\|\mathbf{Y}\|_F$ は行列 \mathbf{Y} のフロベニウスノルムで, 行列要素の 2 乗和平方根である.

証明は付録 D に記す. 定理 3.5 および定理 3.6 を用い

表 2: $\text{rank}\mathbf{Y}$ の判別法

	$\det \mathbf{Y}^T \mathbf{Y} = 0$	$\det \mathbf{Y}^T \mathbf{Y} \neq 0$
$\mathbf{Y} = \mathbf{O}$	$\text{rank}\mathbf{Y} = 0$	—
$\mathbf{Y} \neq \mathbf{O}$	$\text{rank}\mathbf{Y} = 1$	$\text{rank}\mathbf{Y} = 2$

アルゴリズム 1: SVT 算出法 ($M \geq 3$)

```

input :  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2] \in \mathbb{R}^{M \times 2}$ ,  $\mu > 0$ 
output:  $\mathbf{Z} = \text{prox}_{\mu\|\cdot\|_*}(\mathbf{Y})$ 
1 if  $\mathbf{Y} = \mathbf{O}_{M \times 2}$  ( $\leftrightarrow \text{rank}\mathbf{Y} = 0$ ) then
2  $\mathbf{Z} \leftarrow \mathbf{O}_{M \times 2}$ ;
3 else
4  $a \leftarrow \mathbf{y}_1^T \mathbf{y}_1, b \leftarrow \mathbf{y}_1^T \mathbf{y}_2, c \leftarrow \mathbf{y}_2^T \mathbf{y}_2$ ;
5  $d \leftarrow ac - b^2$ ,  $e \leftarrow \sqrt{d}$ ; %  $d = \det \mathbf{Y}^T \mathbf{Y}$ 
6  $f \leftarrow a + c$ ; %  $f = \text{tr} \mathbf{Y}^T \mathbf{Y} = \|\mathbf{Y}\|_F^2$ 
7 if  $d = 0$  ( $\leftrightarrow \text{rank}\mathbf{Y} = 1$ ) then
8  $\mathbf{Z} \leftarrow \left(1 - \frac{\mu}{\sqrt{f}}\right)_+ \mathbf{Y}$ ;
9 else
10  $w_2 \leftarrow \sqrt{f - 2e}$ ; %  $w_2 = \sigma_1 - \sigma_2$ 
11 if  $w_2 = 0$  ( $\leftrightarrow \sigma_1 = \sigma_2$ ) then
12  $\mathbf{Z} \leftarrow \left(1 - \frac{\sqrt{2}\mu}{\sqrt{f}}\right)_+ \mathbf{Y}$ ;
13 else
14  $w_1 \leftarrow \sqrt{f + 2e}$ ; %  $w_1 = \sigma_1 + \sigma_2$ 
15  $\sigma_2 \leftarrow \frac{1}{2}(w_1 - w_2)$ ;
16  $\gamma \leftarrow \left(1 - \frac{(\mu - \sigma_2)_+}{w_2}\right)_+$ ,  $\delta \leftarrow \frac{\min(\mu, \sigma_2)}{w_1}$ ;
17  $\mathbf{Z} \leftarrow \mathbf{Y} \left( \gamma(1-\delta)\mathbf{I}_2 + \frac{\gamma\delta}{e} \begin{bmatrix} -c & b \\ b & -a \end{bmatrix} \right)$ ;

```

て SVT 算出するためには, 階数 $\text{rank}\mathbf{Y}$ と特異値 σ_1, σ_2 から適切に場合分けする必要がある. 階数については, \mathbf{Y} がゼロ行列 \mathbf{O} か否か, $\det \mathbf{Y}^T \mathbf{Y}$ が 0 か否かにより, 表 2 の通り判別できる. また特異値については, 式 (6) の中間式より $\sigma_1 \pm \sigma_2$ を算出すれば良い.

以上の検討をまとめ, SVT を算出するアルゴリズム 1, 2 を提案する. アルゴリズム 1 は $M \geq 3$ の場合, アルゴリズム 2 は $M=2$ の場合の手順を示している. 計算量を削減するため, 場合分けに必要な $\det \mathbf{Y}^T \mathbf{Y}$, $\sigma_1 - \sigma_2$ や係数 γ, δ は全て内積 $\mathbf{y}_i^T \mathbf{y}_j$, ($i, j = 1, 2$) を用いて算出している.

アルゴリズムの計算量について考察する. 式 (4), (5) の通り SVD を用いる SVT 算出法では, SVD に $24M + 160$ [flops]⁹, 閾値処理に 2 [flops], 行列積に $6M + 4$ [flops] で合計 $30M + 166$ [flops] の浮動小数点演算が必要である. 一方でアルゴリズム 1 は $12M + 26$ [flops], アルゴリズム 2 は 36 [flops] であり, SVD に基づく方法と比較して 60~84% の浮動小数点演算を削減できる.

3.3. 行列間並列化

続いて前節のアルゴリズムを用いて, L 個の行列入力 $\mathbf{Y}_1, \dots, \mathbf{Y}_L$ の SVT を同時に求める方法を述べる.

アルゴリズム 1, 2 の処理の大半は, ベクトルや行列に関する基本演算 (和, 積, 定数倍) より構成され, 並列化の効果が高いことを示唆している. そこで L 個の行列 $\mathbf{Y}_1, \dots, \mathbf{Y}_L$ に対してデータ並列に SVT を算出す

⁹サイズ $M \times N$ の入力行列に対して R-SVD アルゴリズム [17] を用いて thin SVD を求める時の浮動小数点演算回数 $6MN^2 + 20N^3$ [flops] に, $N=2$ を代入して算出している.

アルゴリズム 2: SVT 算出法 ($M = 2$)

```

input :  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2] = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \in \mathbb{R}^{2 \times 2}, \quad \mu > 0$ 
output:  $\mathbf{Z} = \text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y})$ 
1 if  $\mathbf{Y} = \mathbf{O}_{2 \times 2}$  ( $\leftrightarrow \text{rank} \mathbf{Y} = 0$ ) then
2  $|\mathbf{Z} \leftarrow \mathbf{O}_{2 \times 2}$ ;
3 else
4  $a \leftarrow \mathbf{y}_1^T \mathbf{y}_1, c \leftarrow \mathbf{y}_2^T \mathbf{y}_2$ ;
5  $d \leftarrow \det \mathbf{Y}, \quad e \leftarrow |d|$ ;
6  $f \leftarrow a + c; \quad \% f = \text{tr} \mathbf{Y}^T \mathbf{Y} = \|\mathbf{Y}\|_F^2$ 
7 if  $d = 0$  ( $\leftrightarrow \text{rank} \mathbf{Y} = 1$ ) then
8  $|\mathbf{Z} \leftarrow \left(1 - \frac{\mu}{\sqrt{f}}\right)_+ \mathbf{Y}$ ;
9 else
10  $w_2 \leftarrow \sqrt{f - 2e}; \quad \% w_2 = \sigma_1 - \sigma_2$ 
11 if  $w_2 = 0$  ( $\leftrightarrow \sigma_1 = \sigma_2$ ) then
12  $|\mathbf{Z} \leftarrow \left(1 - \frac{\sqrt{2}\mu}{\sqrt{f}}\right)_+ \mathbf{Y}$ ;
13 else
14  $w_1 \leftarrow \sqrt{f + 2e}; \quad \% w_1 = \sigma_1 + \sigma_2$ 
15  $\sigma_2 \leftarrow \frac{1}{2}(w_1 - w_2)$ ;
16  $\gamma \leftarrow \left(1 - \frac{(\mu - \sigma_2)_+}{w_2}\right)_+, \quad \delta \leftarrow \frac{\min(\mu, \sigma_2)}{w_1}$ ;
17  $|\mathbf{Z} \leftarrow (\gamma - \gamma\delta)\mathbf{Y} + \text{sgn}(d)\gamma\delta \begin{bmatrix} -y_{22} & y_{21} \\ y_{12} & -y_{11} \end{bmatrix}$ ;

```

るアルゴリズム 3 を Fast Multiple SVT (FMSVT) として提案する. FMSVT の計算量は, 単純にアルゴリズム 1,2 の L 倍である.

アルゴリズム 3 では係数 $\gamma^{(i)}, \delta^{(i)}$ の算出のみ条件分岐させ, その他の処理をインデックス $i = 1, \dots, L$ 間で共通化する. これにより, 係数 $\gamma^{(i)}, \delta^{(i)}$ 算出に費やす $4L$ [flops] 以外を全てデータ並列化でき, $M \geq 3$ では 93.5%以上, $M=2$ では 88.9%の浮動小数点演算をデータ並列に処理する.

加えてアルゴリズム 3 は SIMD 型アーキテクチャとの親和性が特に高い. SIMD では条件分岐をマスク演算で実現でき, 係数 $\gamma^{(i)}, \delta^{(i)}$ の算出もデータ並列化できる.

4. 実験

提案手法による SVT 算出性能を評価するために, 乱数で生成した行列に各アルゴリズムで SVT を施し, 実行時間と計算精度を観測する. 以下の実験は, 全て MATLAB R2016b (Ver. 9.1.0.441655) を用いて実装し, CPU: Intel Core i7-4770@3.40GHz (4 コア 8 スレッド), RAM: 16.0GB の計算機により実行している.

4.1. シミュレーション条件

シミュレーションに用いるテスト行列は, ランダムな入力行列 $\mathbf{Y}_i (i = 1, \dots, L)$ と, その SVT 結果である出力行列 $\mathbf{Z}_i (i = 1, \dots, L)$ の組である. 出力行列は各アルゴリズムで算出された SVT 結果の精度を確かめるのに用いる.

テスト行列の生成法を以下に述べる. 文献 [20] で言及されている方法に倣い, ランダムな行列から真値に近い特異ベクトルを得る方法を採用する. まず正規乱数を要素に持つ $M \times 2$ 行列 \mathbf{X}_i を生成する. 続いて \mathbf{X}_i を SVD し, ランダムな特異ベクトルの組 $\mathbf{U}_i, \mathbf{V}_i$ を得る. 次にランダムな特異値 $\sigma_1 \in [0.5, 1], \sigma_2 \in$

アルゴリズム 3: Fast Multiple SVT

```

input :  $(\mathbf{Y}_1, \dots, \mathbf{Y}_L), \quad \mu > 0$ 
output:  $(\mathbf{Z}_1, \dots, \mathbf{Z}_L) = (\text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y}_1), \dots, \text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y}_L))$ 
1 各  $i = 1, \dots, L$  についてアルゴリズム 1,2 に従って
    $a^{(i)}, b^{(i)}, c^{(i)}, d^{(i)}, e^{(i)}, f^{(i)}, w_1^{(i)}, w_2^{(i)}, \sigma_2^{(i)}$  を算出;
2 各  $i = 1, \dots, L$  について以下の条件分岐を実行;
3 if  $\mathbf{Y}_i = \mathbf{O}$  then
4  $|\gamma^{(i)} \leftarrow 0, \quad \delta^{(i)} \leftarrow 0$ ;
5 else if  $d^{(i)} = 0$  then
6  $|\gamma^{(i)} \leftarrow \left(1 - \frac{\mu}{\sqrt{f^{(i)}}}\right)_+, \quad \delta^{(i)} \leftarrow 0$ ;
7 else if  $w_2^{(i)} = 0$  then
8  $|\gamma^{(i)} \leftarrow \left(1 - \frac{\sqrt{2}\mu}{\sqrt{f^{(i)}}}\right)_+, \quad \delta^{(i)} \leftarrow 0$ ;
9 else
10  $|\gamma^{(i)} \leftarrow \left(1 - \frac{(\mu - \sigma_2^{(i)})_+}{w_2^{(i)}}\right)_+, \quad \delta^{(i)} \leftarrow \frac{\min(\mu, \sigma_2^{(i)})}{w_1^{(i)}}$ ;
11 各  $i = 1, \dots, L$  についてアルゴリズム 1,2 の 17 行目に従って  $\mathbf{Z}_i$  を算出;

```

表 3: 計算時間
時間単位は [msec]

(M, L)	SVDSVT [3]	FRSVT [14]	FMSVT-np	FMSVT-p
	時間	時間 比	時間 比	時間 比
(2, 10)	0.099	0.383 $\times 0.26$	0.057 $\times 1.74$	0.042 $\times 2.37$
(2, 100)	0.958	3.760 $\times 0.25$	0.553 $\times 1.73$	0.054 $\times 17.80$
(2, 1000)	9.637	38.008 $\times 0.25$	5.512 $\times 1.75$	0.173 $\times 55.75$
(2, 10000)	97.195	374.000 $\times 0.26$	55.059 $\times 1.77$	1.014 $\times 95.82$
(3, 10)	0.110	0.362 $\times 0.30$	0.079 $\times 1.39$	0.050 $\times 2.19$
(3, 100)	1.051	3.510 $\times 0.30$	0.772 $\times 1.36$	0.068 $\times 15.50$
(3, 1000)	10.660	35.209 $\times 0.30$	7.639 $\times 1.40$	0.221 $\times 48.34$
(3, 10000)	111.440	352.130 $\times 0.32$	75.787 $\times 1.47$	1.312 $\times 84.93$
(10, 10)	0.112	0.362 $\times 0.31$	0.081 $\times 1.39$	0.051 $\times 2.20$
(10, 100)	1.095	3.543 $\times 0.31$	0.770 $\times 1.42$	0.074 $\times 14.84$
(10, 1000)	10.713	35.508 $\times 0.30$	7.655 $\times 1.40$	0.402 $\times 26.64$
(10, 10000)	108.430	356.510 $\times 0.30$	76.764 $\times 1.41$	2.246 $\times 48.27$
(50, 10)	0.120	0.377 $\times 0.32$	0.085 $\times 1.42$	0.055 $\times 2.20$
(50, 100)	1.171	3.697 $\times 0.32$	0.823 $\times 1.42$	0.117 $\times 10.04$
(50, 1000)	11.430	37.166 $\times 0.31$	7.989 $\times 1.43$	0.850 $\times 13.45$
(50, 10000)	117.490	375.610 $\times 0.31$	82.119 $\times 1.43$	16.219 $\times 7.24$
(100, 10)	0.247	0.429 $\times 0.58$	0.086 $\times 2.86$	0.061 $\times 4.05$
(100, 100)	2.079	4.195 $\times 0.50$	0.833 $\times 2.50$	0.236 $\times 8.82$
(100, 1000)	19.841	42.190 $\times 0.47$	8.670 $\times 2.29$	1.452 $\times 13.66$
(100, 10000)	191.050	420.990 $\times 0.45$	87.749 $\times 2.18$	31.169 $\times 6.13$

$[0, 0.5]$ を一様乱数で生成する. 最後に, 入力テスト行列 $\mathbf{Y}_i = \mathbf{U}_i \text{diag}(\sigma_1, \sigma_2) \mathbf{V}_i^T$ と, 出力テスト行列 $\mathbf{Z}_i = \mathbf{U}_i \text{diag}((\sigma_1 - \mu)_+, (\sigma_2 - \mu)_+) \mathbf{V}_i^T$ を算出する. 以上の手順は全て倍精度の浮動小数点演算により算出した. 行列の列サイズは $M = 2, 3, 10, 50, 100$ の 5 パターン, 行列の数は 10, 100, 1000, 10000 の 4 パターンを用いた.

実験に用いる SVT のアルゴリズムは, 式 (4), (5) の通り SVD に基づいて算出する方法 [3](SVDSVT), Oh らの手法 [14](FRSVT), 本稿で提案したアルゴリズム 1,2(FMSVT-np), およびこれをデータ並列化したアルゴリズム 3(FMSVT-p) である. SVDSVT は MATLAB 標準の SVD 関数, および行列積を用いて実装し, また FRSVT については Oh らが作成したオリジナルの MATLAB コードを用いた. FRSVT のコア行列サイズは 1 とし, 他のパラメータは [14] と同じ設定にした. また SVDSVT, FRSVT, FMSVT-np は, 行列間の処理 (3) をループで逐次的に算出している.

4.2. シミュレーション結果

各アルゴリズムで SVT を算出するのに要した実行時間を表 3 に示す. 各計算は倍精度演算で実行してい

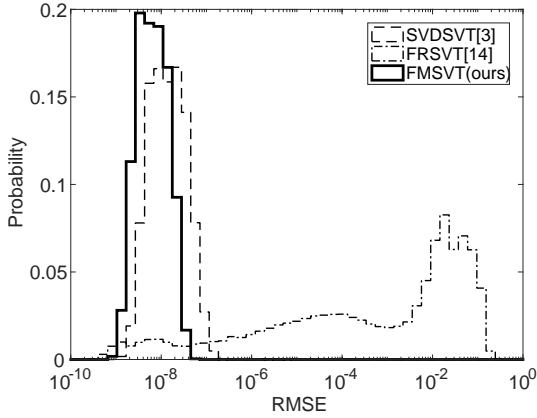


図 4: 計算精度の分布

RMSE の平均値は SVDSVT, FRSVT, FMSVT の順に,
 2.11×10^{-8} , 1.86×10^{-2} , 8.77×10^{-9} .

る。どの条件においても提案した FMSVT-p が最も速く、SVDSVT と比較して最大 95.82 倍速度向上している。また並列化していない FMSVT-np の効果が 1.3~2.8 倍程度であることから、アルゴリズム 3 のデータ並列化効果が著しく高いことが確認できる。なお FRSVT の速度が SVDSVT に劣るのは、入力行列の列数が 2 と小さく、コア行列縮退の効果を十分に得られないためである。

続いて計算精度について図 4 を用いて述べる。図 4 は、単精度で実行した各 SVT の出力結果と出力テスト行列との計算誤差を、二乗平均平方根誤差 (Root Mean Square Error ;RMSE) の分布で示している。ただし FRSVT に関しては、Oh らのコードが単精度演算できない仕様のため、倍精度演算の結果を示している。また FMSVT-np と FMSVT-p は同一の計算結果であるので、まとめて FMSVT と記載している。提案した FMSVT は SVDSVT と比較して計算誤差が半分程度で、RMSE は 8.77×10^{-9} であった。一方で FRSVT は他の手法と比較して誤差が約 10^6 倍大きい。これは FRSVT が行列近似に基づいて高速化する手法であるため、本質的に回避できない。

以上より FMSVT が計算速度に優れ、計算誤差の小さい SVT 計算手法であることを確認した。

5. おわりに

本稿では、多数の小型行列を低ランク正則化する最適化問題の解を高速に求めることを目的として、特異値閾値処理 (SVT) の高速算出アルゴリズム FMSVT を提案した。核型ノルムを部分空間上のベクトル距離で表現できる発見に基づき、特異値分解 (SVD) が不要でデータ並列な SVT 算出法を導いた。評価実験の結果、従来手法以上の計算精度を持ちつつ、最大 95.82 倍高速に SVT を算出できることを確認した。

付録

A. 命題 3.1 の証明

証明. 行列 $\mathbf{Y}^T \mathbf{Y}$ の固有値を $\lambda_1, \lambda_2, (\lambda_1 \geq \lambda_2 \geq 0)$ とする。 $\lambda_1 = \sigma_1^2, \lambda_2 = \sigma_2^2, \lambda_1 + \lambda_2 = \text{tr} \mathbf{Y}^T \mathbf{Y} = \|\mathbf{y}_1\|_2^2 + \|\mathbf{y}_2\|_2^2,$

$$\lambda_1 \lambda_2 = \det \mathbf{Y}^T \mathbf{Y} = \|\mathbf{y}_1\|_2^2 \|\mathbf{y}_2\|_2^2 - (\mathbf{y}_1^T \mathbf{y}_2)^2 \text{ より,}$$

$$\begin{aligned} \sigma_1 \pm \sigma_2 &= \sqrt{\sigma_1^2 + \sigma_2^2 \pm 2\sigma_1 \sigma_2} = \sqrt{\lambda_1 + \lambda_2 \pm 2\sqrt{\lambda_1 \lambda_2}} \\ &= \sqrt{\text{tr} \mathbf{Y}^T \mathbf{Y} \pm 2\sqrt{\det \mathbf{Y}^T \mathbf{Y}}} \\ &= \sqrt{\|\mathbf{y}_1\|_2^2 + \|\mathbf{y}_2\|_2^2 \pm 2\sqrt{\|\mathbf{y}_1\|_2^2 \|\mathbf{y}_2\|_2^2 - (\mathbf{y}_1^T \mathbf{y}_2)^2}} \end{aligned} \quad (20)$$

である。

ここで \mathbf{Y} を表現行列とする線形写像の像 $\text{Im} \mathbf{Y}$ を考えると、高々 2 次元の部分空間である ($\dim(\text{Im} \mathbf{Y}) = \text{rank} \mathbf{Y} \leq 2$)。そこで図 1 の通り $\text{Im} \mathbf{Y}$ を本稿紙面に埋め込み、ベクトル $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^M$ の幾何的性質に着目する。 $\mathbf{y}_1, \mathbf{y}_2$ の成す角を $\theta \in [0, \pi]$ [rad] とすると、 $(\mathbf{y}_1^T \mathbf{y}_2)^2 = \|\mathbf{y}_1\|_2^2 \|\mathbf{y}_2\|_2^2 \cos^2 \theta$ であるから、

$$(20) = \sqrt{\|\mathbf{y}_1\|_2^2 + \|\mathbf{y}_2\|_2^2 \pm 2\|\mathbf{y}_1\|_2 \|\mathbf{y}_2\|_2 \sin \theta} \quad (21)$$

が得られる。

また図 1 に示すように、ベクトル \mathbf{y}_2 を $\pi/2$ [rad] 回転させた $\mathbf{R} \mathbf{y}_2$ を考えると、 \mathbf{y}_1 と $\mathbf{R} \mathbf{y}_2$ の成す角は $\theta + \pi/2$ [rad] である。 $\|\mathbf{R} \mathbf{y}_2\|_2 = \|\mathbf{y}_2\|_2, \sin \theta = -\cos(\theta + \pi/2)$ であるから、

$$\begin{aligned} \|\mathbf{y}_1\|_2 \|\mathbf{y}_2\|_2 \sin \theta &= -\|\mathbf{y}_1\|_2 \|\mathbf{R} \mathbf{y}_2\|_2 \cos(\theta + \pi/2) \\ &= -\mathbf{y}_1^T \mathbf{R} \mathbf{y}_2 \end{aligned} \quad (22)$$

が成立する。式 (21) に代入して、

$$\begin{aligned} (21) &= \sqrt{\|\mathbf{y}_1\|_2^2 + \|\mathbf{R} \mathbf{y}_2\|_2^2 \mp 2\mathbf{y}_1^T \mathbf{R} \mathbf{y}_2} \\ &= \|\mathbf{y}_1 \mp \mathbf{R} \mathbf{y}_2\|_2 \end{aligned} \quad (23)$$

が得られる。従って題意は示された。□

B. 命題 3.3 の証明

命題 3.3 の証明に先立って、次の 2 つの補題を示す。

補題 B.1. 式 (8) の行列 \mathbf{B} は、任意の $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T]^T, \mathbf{x}_1, \mathbf{x}_2 \in \text{Im} \mathbf{Y}$ に対し、

$$\mathbf{B} \mathbf{B}^T \mathbf{x} = \mathbf{B}^T \mathbf{B} \mathbf{x} = 2\mathbf{x} \quad (24)$$

が成立する。

証明. 脚注 5 の性質 (5) より、

$$\begin{aligned} \mathbf{B} \mathbf{B}^T \mathbf{x} &= \begin{bmatrix} \mathbf{I} + \mathbf{R} \mathbf{R}^T & \mathbf{I} - \mathbf{R} \mathbf{R}^T \\ \mathbf{I} - \mathbf{R} \mathbf{R}^T & \mathbf{I} + \mathbf{R} \mathbf{R}^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{x}_1 + \mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_2 \\ \mathbf{x}_1 - \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_2 \end{bmatrix} = 2\mathbf{x} \end{aligned} \quad (25)$$

$$\mathbf{B}^T \mathbf{B} \mathbf{x} = \begin{bmatrix} 2\mathbf{I} & \mathbf{O} \\ \mathbf{O} & 2\mathbf{R}^T \mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = 2\mathbf{x} \quad (26)$$

である。□

補題 B.2. ベクトル $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^M$ を $[\mathbf{z}_1, \mathbf{z}_2] = \text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y})$ とする。このとき $\mathbf{z}_1, \mathbf{z}_2 \in \text{Im} \mathbf{Y}$ である。

証明. $(\mathbf{U}, \Sigma, \mathbf{V}) = \text{SVD}(\mathbf{Y})$ とし,

$$\begin{aligned}\mathbf{U} &= [\mathbf{u}_1, \mathbf{u}_2], \mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^M \\ \Sigma &= \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}, \sigma_1, \sigma_2 \in \mathbb{R} \\ \mathbf{V} &= \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}, v_{11}, v_{12}, v_{21}, v_{22} \in \mathbb{R}\end{aligned}$$

とする. SVD の定義 (脚注 3) より $\mathbf{y}_i = \sigma_1 v_{i1} \mathbf{u}_1 + \sigma_2 v_{i2} \mathbf{u}_2$ なので,

$$\text{Im} \mathbf{Y} = \begin{cases} \text{Im} \mathbf{U} & \text{if rank} \mathbf{Y} = 2 \\ \text{Im} \mathbf{u}_1 & \text{if rank} \mathbf{Y} = 1 \\ \{\mathbf{0}\} & \text{if rank} \mathbf{Y} = 0 \end{cases} \quad (27)$$

である.

同様に SVT の定義 (式 (4)) より, $\mathbf{z}_i = (\sigma_1 - \mu)_+ v_{i1} \mathbf{u}_1 + (\sigma_2 - \mu)_+ v_{i2} \mathbf{u}_2$ なので, (27) の場合分け毎に, $\mathbf{z}_i \in \text{Im} \mathbf{Y}$ が確認できる. \square

以上を踏まえて, 命題 3.3 を証明する.

証明. 系 3.2 と近接写像の定義 (脚注 2) より,

$$\begin{aligned}\text{prox}_{\mu \|\cdot\|_*}(\mathbf{Y}) &= \text{prox}_{\mu \|\mathbf{B} \text{vec}(\cdot)\|_{\infty, 2}}(\mathbf{Y}) \\ &= \arg \min_{\mathbf{Z} \in \mathbb{R}^{D \times 2}} \|\mathbf{B} \text{vec}(\mathbf{Z})\|_{\infty, 2} + \frac{1}{2\mu} \|\mathbf{Z} - \mathbf{Y}\|_F^2 \\ &= \text{vec}^T \left(\arg \min_{\mathbf{z} \in \mathbb{R}^{2D}} \|\mathbf{Bz}\|_{\infty, 2} + \frac{1}{2\mu} \|\mathbf{z} - \text{vec}(\mathbf{Y})\|_2^2 \right) \quad (28)\end{aligned}$$

が成立する. ここで $\mathbf{z} = \text{vec}(\mathbf{Z})$ と変数置換しているが, 補題 B.2 より探索範囲を $\mathbf{z} \in \text{Im} \mathbf{Y}^2$ に限定できる. そこで変数を再度置換して

$$\mathbf{x} = \frac{1}{2} \mathbf{Bz} \quad (29)$$

とする. 両辺に \mathbf{B}^T を掛けると, 補題 B.1 より $\mathbf{B}^T \mathbf{x} = \mathbf{z}$ となる. また再び補題 B.1 より, $\text{vec}(\mathbf{Y}) = \frac{1}{2} \mathbf{B}^T \mathbf{B} \text{vec}(\mathbf{Y})$ が成立する. 以上より,

$$\begin{aligned}\|\mathbf{z} - \text{vec}(\mathbf{Y})\|_2^2 &= \left\| \mathbf{B}^T \mathbf{x} - \frac{1}{2} \mathbf{B}^T \mathbf{B} \text{vec}(\mathbf{Y}) \right\|_2^2 \\ &= \left(\mathbf{x} - \frac{1}{2} \mathbf{B} \text{vec}(\mathbf{Y}) \right)^T \mathbf{B} \mathbf{B}^T \left(\mathbf{x} - \frac{1}{2} \mathbf{B} \text{vec}(\mathbf{Y}) \right) \\ &= 2 \left\| \mathbf{x} - \frac{1}{2} \mathbf{B} \text{vec}(\mathbf{Y}) \right\|_2^2 \quad (30)\end{aligned}$$

となる (最右辺への変形にも補題 B.1 を使用). 式 (28) に式 (29), (30) を適用すると,

$$\begin{aligned}(28) &= \text{vec}^T \left(\mathbf{B}^T \arg \min_{\mathbf{x} \in \text{Im} \mathbf{Y}^2} \|\mathbf{x}\|_{\infty, 2} + \frac{1}{2\mu} \left\| \mathbf{x} - \frac{1}{2} \mathbf{B} \text{vec}(\mathbf{Y}) \right\|_2^2 \right) \\ &= \text{vec}^T \left(\mathbf{B}^T \text{prox}_{\mu \|\cdot\|_{\infty, 2}} \left(\frac{1}{2} \mathbf{B} \text{vec}(\mathbf{Y}) \right) \right) \quad (31)\end{aligned}$$

となり, 題意を得る. \square

C. 命題 3.4 の証明

証明. $L_{\infty, 2}$ ノルムの双対ノルムは $L_{1, 2}$ ノルムなので, Moreau の直交分解¹⁰ [18] より,

$$\text{prox}_{\mu \|\cdot\|_{\infty, 2}}(\mathbf{x}) = \mathbf{x} - \text{proj}_{\mu B_{1, 2}}(\mathbf{x}) \quad (32)$$

である. ここで $\text{proj}_{\mu B_{1, 2}}(\mathbf{x})$ は半径 μ の $L_{1, 2}$ 球である $\mu B_{1, 2}$ への L_2 距離射影で, 下記の最適化問題の解である.

$$\begin{aligned}\text{proj}_{\mu B_{1, 2}}(\mathbf{x}) &= \arg \min_{\mathbf{z} \in \mu B_{1, 2}} \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 \\ \mu B_{1, 2} &= \{ \mathbf{z} = [\mathbf{z}_1^T, \mathbf{z}_2^T]^T \mid \|\mathbf{z}_1\|_2 + \|\mathbf{z}_2\|_2 \leq \mu \} \quad (33)\end{aligned}$$

van den Berg ら [19] によれば, 問題 (33) は L_1 球射影と L_2 球射影の問題に分離でき,

$$\text{proj}_{\mu B_{1, 2}}(\mathbf{x}) = \begin{bmatrix} \text{proj}_{\eta_1 B_2}(\mathbf{x}_1) \\ \text{proj}_{\eta_2 B_2}(\mathbf{x}_2) \end{bmatrix} \quad (34)$$

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \text{proj}_{\mu B_1} \left(\begin{bmatrix} \|\mathbf{x}_1\|_2 \\ \|\mathbf{x}_2\|_2 \end{bmatrix} \right) \quad (35)$$

の 2 つの問題と等価である. 式 (34) の L_2 ノルム射影は $\text{proj}_{\eta_i B_2}(\mathbf{x}_i) = (\eta_i / \|\mathbf{x}_i\|_2) \mathbf{x}_i$ と計算できる. 一方で式 (35) の L_1 ノルム射影は

$$(35) = \begin{cases} \begin{bmatrix} \frac{\|\mathbf{x}_1\|_2 - (\|\mathbf{x}_1\|_2 + \|\mathbf{x}_2\|_2 - \mu)_+}{2} \\ \frac{\|\mathbf{x}_2\|_2 - (\|\mathbf{x}_1\|_2 + \|\mathbf{x}_2\|_2 - \mu)_+}{2} \end{bmatrix} & \text{if } \|\mathbf{x}_1\|_2 - \|\mathbf{x}_2\|_2 \leq \mu \\ \begin{bmatrix} \mu \\ 0 \end{bmatrix} & \text{if } \|\mathbf{x}_1\|_2 - \|\mathbf{x}_2\|_2 > \mu \\ \begin{bmatrix} 0 \\ \mu \end{bmatrix} & \text{if } \|\mathbf{x}_1\|_2 - \|\mathbf{x}_2\|_2 < -\mu \end{cases} \quad (36)$$

となる. 以上より,

$$(32) = \begin{bmatrix} k_1 \mathbf{x}_1 \\ k_2 \mathbf{x}_2 \end{bmatrix} \quad (37)$$

where

$$\begin{aligned}\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} &= \begin{bmatrix} 1 - \frac{\eta_1}{\|\mathbf{x}_1\|_2} \\ 1 - \frac{\eta_2}{\|\mathbf{x}_2\|_2} \end{bmatrix} \\ &= \begin{cases} \begin{bmatrix} \frac{(\|\mathbf{x}_1\|_2 + \|\mathbf{x}_2\|_2 - \mu)_+}{2\|\mathbf{x}_1\|_2} \\ \frac{(\|\mathbf{x}_1\|_2 + \|\mathbf{x}_2\|_2 - \mu)_+}{2\|\mathbf{x}_2\|_2} \end{bmatrix} & \text{if } \|\mathbf{x}_1\|_2 - \|\mathbf{x}_2\|_2 \leq \mu \\ \begin{bmatrix} 1 - \frac{\mu}{\|\mathbf{x}_1\|_2} \\ 1 \end{bmatrix} & \text{if } \|\mathbf{x}_1\|_2 - \|\mathbf{x}_2\|_2 > \mu \\ \begin{bmatrix} 1 \\ 1 - \frac{\mu}{\|\mathbf{x}_2\|_2} \end{bmatrix} & \text{if } \|\mathbf{x}_1\|_2 - \|\mathbf{x}_2\|_2 < -\mu \end{cases} \quad (38)\end{aligned}$$

となり, 題意を満たす. \square

¹⁰ある凸関数 $f(\mathbf{x})$ とそのルジャンドル変換 $f^*(\mathbf{x})$ について, \mathbf{x} の分解 $\mathbf{x} = \text{prox}_f(\mathbf{x}) + \text{prox}_{f^*}(\mathbf{x})$ を Moreau の直交分解という. 特に凸関数がノルム $f(\mathbf{x}) = \mu \|\mathbf{x}\|$ のとき, 双対ノルム $\mu \|\mathbf{x}\|_d$ の球 B_d への射影 $\text{proj}_{\mu B_d}(\mathbf{x})$ を用いて, $\mathbf{x} = \text{prox}_{\mu \|\cdot\|}(\mathbf{x}) + \text{proj}_{\mu B_d}(\mathbf{x})$ が成立する.

D. 定理 3.6 の証明

証明. $\mathbf{Z} = \text{prox}_{\mu\|\cdot\|_*}(\mathbf{Y})$ とする. $\text{rank}\mathbf{Y} = 0$ の場合, $\mathbf{Y} = \mathbf{O}_{M \times 2}$ であるため自明.

$\text{rank}\mathbf{Y} = 1$ の場合, $\sigma_1 > 0, \sigma_2 = 0$ であるため,

$$\begin{aligned} \mathbf{Z} &= \mathbf{U} \begin{bmatrix} (\sigma_1 - \mu)_+ & 0 \\ 0 & 0 \end{bmatrix} \mathbf{V}^T = \frac{(\sigma_1 - \mu)_+}{\sigma_1} \mathbf{U} \begin{bmatrix} \sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{V}^T \\ &= \left(1 - \frac{\mu}{\sigma_1}\right)_+ \mathbf{Y} \end{aligned} \quad (39)$$

である. ここで,

$$\begin{aligned} \|\mathbf{Y}\|_F &= \sqrt{\text{tr}\mathbf{Y}^T\mathbf{Y}} = \sqrt{\text{tr}(\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T)^T (\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T)} \\ &= \sigma_1 \end{aligned} \quad (40)$$

であるため, 題意を満たす.

$\text{rank}\mathbf{Y} = 2$ かつ $\sigma_1 = \sigma_2$ の場合,

$$\begin{aligned} \mathbf{Z} &= \mathbf{U} \begin{bmatrix} (\sigma_1 - \mu)_+ & 0 \\ 0 & (\sigma_1 - \mu)_+ \end{bmatrix} \mathbf{V}^T \\ &= \frac{(\sigma_1 - \mu)_+}{\sigma_1} \mathbf{U} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_1 \end{bmatrix} \mathbf{V}^T \\ &= \left(1 - \frac{\mu}{\sigma_1}\right)_+ \mathbf{Y} \end{aligned} \quad (41)$$

である. ここで,

$$\begin{aligned} \|\mathbf{Y}\|_F &= \sqrt{\text{tr}\mathbf{Y}^T\mathbf{Y}} = \sqrt{\text{tr}(\sigma_1 \mathbf{U}\mathbf{V}^T)^T (\sigma_1 \mathbf{U}\mathbf{V}^T)} \\ &= \sqrt{2}\sigma_1 \end{aligned} \quad (42)$$

より題意を満たす. \square

参考文献

- [1] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," *Advances in Neural Information Processing Systems 22*, eds. by Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, and A. Culotta, pp.2080–2088, Curran Associates, Inc., 2009.
- [2] E.J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol.58, no.3, pp.11:1–11:37, June 2011.
- [3] J.-F. Cai, E.J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol.20, no.4, pp.1956–1982, 2010.
- [4] S. Gandy and I. Yamada, "Convex optimization techniques for the efficient recovery of a sparsely corrupted low-rank matrix," *Journal of Math-for-Industry (JMI)*, vol.2, no.B, pp.147–156, 2010.
- [5] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.35, no.3, pp.597–610, March 2013.
- [6] W. Dong, G. Shi, X. Hu, and Y. Ma, "Nonlocal sparse and low-rank regularization for optical flow estimation," *IEEE Transactions on Image Processing*, vol.23, no.10, pp.4527–4538, Oct. 2014.
- [7] R. Otazo, E.J. Candès, and D.K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol.73, no.3, pp.1125–1136, 2015.
- [8] G. Ye, M. Tang, J.-F. Cai, Q. Nie, and X. Xie, "Low-rank regularization for learning gene expression programs," *PloS one*, vol.8, no.12, p.e82146, 2013.
- [9] 佐々木崇元, 谷田隆一, 清水 淳, "グラフ信号の局所線形近似によるグラフ形状単純化," 第15回情報科学技術フォーラム (FIT2016), 第3分冊, pp.1–4, Sept. 2016.
- [10] S. Ono and I. Yamada, "Color-line regularization for color artifact removal," *IEEE Transactions on Computational Imaging*, vol.2, no.3, pp.204–217, 2016.
- [11] M. Fazel, "Matrix rank minimization with applications," PhD thesis, Stanford University, 2002.
- [12] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol.3, no.1, pp.1–122, 2011.
- [13] J.-F. Cai and O. Stanley, "Fast singular value thresholding without singular value decomposition," *Methods and Applications of Analysis*, vol.20, no.4, pp.335–352, Dec. 2013.
- [14] T.-H. Oh, Y. Matsushita, Y.-W. Tai, and I.S. Kweon, "Fast randomized singular value thresholding for nuclear norm minimization," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.4484–4493, June 2015.
- [15] 佐々木崇元, 北原正樹, 清水 淳, "領域情報符号化における核型ノルム最適化の高速計算法," 第31回画像符号化シンポジウム (PCSJ2016), pp.140–141, Nov. 2016.
- [16] L. Condat, "A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol.158, no.2, pp.460–479, 2013.
- [17] G.H. Golub and C.F. Van Loan, *Matrix computations*, 4th edition, John Hopkins University Press, 2012.
- [18] J.J. Moreau, "Décomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires," *Comptes Rendus de l'Académie des Sciences*, vol.225, pp.238–240, 1962.
- [19] E. van den Berg, M. Schmidt, M.P. Friedlander, and K. Murphy, "Group sparsity via linear-time projection," Technical report, Department of Computer Science, University of British Columbia, Vancouver, Canada, 2008.
- [20] 高田雅美, 木村欣司, 岩崎雅史, 中村佳正, "高速特異値分解のためのライブラリ開発," *情報処理学会論文誌コンピュータビジョンシステム*, vol.47, no.SIG7(ACS14), pp.91–104, May 2006.