

欠損のある低ランク行列分解アルゴリズムの性能比較

吉田孝弘^{†1} 岡谷貴之^{†1} 出口光一郎^{†1}

欠損成分のある低ランク行列の分解を行う数値アルゴリズムが、最近新たに、複数提案されている。本稿では、古くからある Wiberg 法の新しい効率的な実装法を与え、さらにそれにダンピングファクタを加えた計算法を示す。そして、Buchanan らの Damped-Newton 法と、われわれの方法を含むそれ以降に提案された新しい計算法の計 6 つの方法の性能を、実データおよびシミュレーションデータを用いた実験により評価する。

Performance Comparisons Among Algorithms for Low-rank Matrix Factorization with Missing Components

TAKAHIRO YOSHIDA,^{†1} TAKAYUKI OKATANI^{†1}
and KOICHIRO DEGUCHI^{†1}

Several numerical methods for low-rank matrix factorization with missing components have been proposed recently. This paper presents a novel efficient implementation of the Wiberg method, which was developed in 70's, and then introduces its simple extension incorporating a damping factor. We evaluate six methods via several experiments using real data as well as synthetic data, which consist of the Damped-Newton method proposed by Buchanan et al., and the newer methods that have been proposed including ours.

1. はじめに

欠損成分のある行列 Y をそれより小さな 2 つの行列 U と V の積に

$$Y \rightarrow UV^T \quad (1)$$

^{†1} 東北大学
Tohoku University

と分解する問題は、コンピュータビジョンでは SFM (Structure from Motion) や、異方向複数照明画像の解析 (ホトメトリックステレオや IBR (Image Based Rendering)) などで現れ、それ以外の工学の分野でも普遍的に現れる。行列 Y のサイズを $m \times n$ 、ランクを (観測ノイズがなければ) r とするとき、 Y を説明する 2 つの因子、 $m \times r$ 行列 U と $n \times r$ 行列 V に分解したい。

最も一般的な定式化は、観測成分についての再現誤差を L_2 ノルムで測り、これを最小化するように U と V を決定することである：

$$f(U, V) = \|W \odot (Y - UV^T)\|_F^2 \quad (2)$$

ここで W は、 Y と同じサイズの行列で、 (i, j) 成分 y_{ij} が観測されているとき $w_{ij} = 1$ 、そうでないとき $w_{ij} = 0$ となる行列である。 \odot は行列の成分ごとの積を表す。用途に応じて上の式のノルムを変更するが (L_1 とした場合の方法が例えば文献¹⁾にある)、本稿では L_2 ノルムの最小化を考える。この場合、欠損成分がなければ確立された数値計算法があるが、欠損成分のある場合、(2) 式は非線形の最適化問題となり、反復計算を要し、大域的収束が一般に保証されなくなる。

この問題に対し、Shum ら²⁾ が Wiberg の論文³⁾ をコンピュータビジョンの分野で紹介したことを契機に、様々な方法が提案されてきた。その方法は多数あるが (文献⁴⁾ に、その時点でのサーベイがある)、それらは全体を同時に最適化する方法⁴⁾⁻⁷⁾ と、それ以外の方法⁸⁾⁻¹¹⁾ に分けられる。前者は、ニュートン法を始めとする反復解法である。後者は、部分行列で利用できる拘束を繰り返し適用し、徐々に対象とする行列の範囲を拡大し、最終的に一つの近似解あるいは最適解を得ようとする方法が多い。その結果を、前者の初期値に利用する考え方もある。本稿では前者を考える。

本稿の目的は、それらの方法の性能比較を行うことである。この目的の背後には 2 つの動機がある。一つは、近年いくつかの新しい方法⁵⁾⁻⁷⁾ が提案されており、それぞれの文献で既存の方法との性能比較がなされているものの、すべての方法を対象としていなかったり、評価の条件が必ずしも現実の問題を反映していないなど、いくつかの問題があり、正しい評価を下しているとはいえない難かったことである。もう一つは、われわれが、最近改めて顧みられている^{1),12)} Wiberg 法をベースに効率的な実装法を新たに考案したので、その性能を実証することである。

Buchanan らは論文⁴⁾ で、その当時提案されていた数多くのアルゴリズムをサーベイし、そして Damped-Newton 法 (以下 DN 法) を提案している。論文の結論は、DN 法が従来

手法に勝るといふものであった。そして DN 法は、それ以降の研究でほとんど必ず引用されているベンチマーク的方法である。本稿は、この Buchanan らの論文以降に提案された方法、すなわち Damped-Newton⁴⁾、Zhao らの SALS⁷⁾、Chen の LM_M/LM_S⁵⁾、Mittra らの MFLRSDP⁶⁾ に、われわれによる Wiberg 法の新たな実装 (DW) を加えた、計 6 つの方法の性能を比較する。

評価は、特に大域収束性と計算速度の 2 点に焦点を当てる。考えている最適化は上述のように非線形なものであり、理論的には大域収束性は保証されず (局所解に陥る可能性がある) 初期値依存性があることになる。しかしながら、Wiberg 法や Chen の方法のように、(2) 式の L_2 における U と V の双線形性を利用してどちらか一方の変数を消去する方法は、後に示すようにかなり高い大域収束性能を持つ。つまり、初期値をランダムに与えても高確率で大域解に到達できる。

なお、(2) 式の解は一意的に定まらない。任意の正則な 3×3 行列に対して、 $UV^T = UA^{-1}AV^T = (UA^{-1})(AV^T) = U'V'^T$ が成立するからである。したがって、この不定性を何らかのやり方で拘束する必要がある。後に示す実験結果は、この拘束のやり方が反復計算の収束性や計算量の多寡を左右することを示唆する。本稿では、個々の問題ごとに特定の拘束が利用できる場合—例えば SFM の場合、カメラが校正済みであれば U に制約が生じ、これが使える—は考えない。

2. 各方法の概要

2.1 Damped-Newton (DN) 法

Buchanan らの DN 法⁴⁾ では、目的関数 f を正則化した

$$f_R(U, V) = \|W \odot (Y - UV^T)\|_F^2 + \mu_1 \|U\|_F^2 + \mu_2 \|V\|_F^2 \quad (3)$$

を最小化することで分解不定性を除去している。なお f_R を最小化して得られる解は、元の f を最小化する解と異なる。なるべく両者が一致すべきだと考えるならば、 Y の値の大きさに比して小さな値を μ_1 および μ_2 に選ぶことになる。

DN 法はこの f_R を、一般的なダンピングファクタありのニュートン法で解く。行列 U, V の成分を順に並べて、長さ $(m+n)r$ のベクトル x を定義したとき、適当な初期値 x_0 から開始し、次の式を解いて更新量 δx を求める。

$$(H + \lambda I)\delta x = d \quad (4)$$

ここで $H = d^2 f_R / dx^2$ 、 $d = df_R / dx$ である。 λI がダンピングファクタで、 λ を f_R が確実に減少するように調節し大域収束性を高める ($f_R(x + \delta x) < f_R(x)$) のとき λ を $1/10$ し、そうでないときは 10 倍する方法が論文⁴⁾にある)。

DN 法の問題の一つは、後述する方法に比して変数が多く、扱う問題の規模が大きくなる ((4) 式の規模すなわち H のサイズが $(m+n)r \times (m+n)r$ となる) ことである。非観測成分が多ければ、 f のヘシアン $d^2 f / dx^2$ は疎行列となるが、 f_R のヘシアンは (3) 式の正則化項の存在から疎にならない。

なお、 μ_1 と μ_2 をともに 0 にセットすると f を最小化するのと同じになる。この場合、分解不定性のせいで (どのような Y および W に対しても) H はランク落ちする。しかしそれでも、ダンピングファクタ λI の存在により (4) 式は縮退せず、一応動作はする。その場合、 H が疎になり得る利点があるが、ダンピングを使ったニュートン法本来の特長—極小点まわりで λ が小さくなり、ニュートン法本来の 2 次収束性が発揮される—が棄損される可能性がある ($\lambda \rightarrow 0$ とすると (4) 式が縮退に向かうため、数値的不安定性が問題となる)。

2.2 Zhao-Zhang の Successively Alternate Least Square (SALS)

Zhao らは、 Y_{true} の各要素がある値の範囲内に収まる、すなわち $a \leq y_{ij} \leq b$ となることが事前に分かっている場合に、この不等式制約を利用した行列分解の計算法を示した⁷⁾。例えば、ランバート性反射の仮定の下でのホトメトリックステレオを考える場合、画像の濃淡値が $[0 : 255]$ の範囲内に収まることが利用できる。回転物体に対する SFM (例えば後の実験でも用いる Dinosaur) では、特徴点の画像座標 (x, y) が一定の範囲に収まることが利用できる。本稿では問題固有の事前知識を利用する除外して考えているが、SALS の場合は事前知識となる上記制約の与え方に、一定の一般性があることから取り上げた。

この方法は、テンプレート行列と呼ぶ $m \times n$ の X を、初期値 x_0 から初めて更新し、最終的に Y_{true} に接近するように反復計算を行う。この反復を外部ループと呼び、その反復一回あたり、現在のテンプレート行列 X を固定し、

$$f_{\text{SALS}}(U, V) = \|W \odot (Y - UV^T)\|_F^2 + \lambda^2 \|W^c (X - UV^T)\|_F^2 \quad (5)$$

を ALS によって最小化するプロセスを、 f_{SALS} の値が閾値 ε を下回るか、反復回数が K 回を超えるまで繰り返す (内部ループ)。ただし W^c は W の 1 と 0 を反転させた $m \times n$ 行列である。この内部ループでの U ないし V に与える初期値は、 X を特異値分解し、 r 個の最大の特異値に対応するベクトルを使って与えるとしている。内部ループ終了後、得られた U, V

を使って x を復元するが、その際不等式制約を各成分に強制する：

$$(x)_{ij} \leftarrow \begin{cases} (UV)_{ij}, & \text{if } (UV)_{ij} \in [a, b], \\ a + \delta, & \text{if } (UV)_{ij} < a, \\ b - \delta, & \text{if } (UV)_{ij} > b. \end{cases} \quad (6)$$

以上の外部ループを繰り返し、 x の変化量が十分小さくなるまでこれを行う。

SALS では行列要素の値の範囲 $[a : b]$ の他に、内部ループの最大反復数 K 、(5) 式の λ 、(6) 式の δ の計 3 つ目のパラメータを決める必要がある。

2.3 Wiberg の方法 (線形性に基づく変数消去)

Wiberg の方法^{3),12)} は、 $f(U, V)$ をガウス・ニュートン法で最小化する方法である。その際、 f の双線形性を利用し、 U, V のいずれかを消去し、変数の数を減じ問題のサイズを小さくする。 Y の縦横長さ (m と n) に差があるときこの方法は特に有効である。これは現実の問題でも良く起こり、長い方の変数を消去すればよい。

例えば U を消去することにすれば、 V を固定したときの f を最小化する U を、 $U = \hat{U}(V)$ と表現する。これを $f(U, V)$ に差し込んだ

$$g(V) \equiv f(\hat{U}(V), V) = \min_U f(U, V) \quad (7)$$

を V について最小化する。

Y の観測成分のみを適当な順でベクトル化したものを y と書き、 U, V をそれぞれベクトル化したものを u, v と書いたとき、 f は次のようにあらわされる。

$$f = \frac{1}{2} \|y - Fu\|^2 = \frac{1}{2} \|y - Gv\|^2 \quad (8)$$

Y の観測成分の総数を $p (< mn)$ と書くと、 $F = F(v)$ 、 $G = G(u)$ はそれぞれ、 $p \times mr$ 、 $p \times nr$ の行列である。

これらの記号を使うと、 g をガウス・ニュートン法で最小化する過程は、次の交互反復を実行することに帰着される。

$$u \leftarrow \underset{u}{\operatorname{argmin}} f(u, v), \quad (9a)$$

$$v \leftarrow v + \delta v, \quad \text{ただし } \delta v = \underset{\delta v}{\operatorname{argmin}} \|Q_F G \delta v - Q_F y\|^2 \quad (9b)$$

$f(u, v)$ は $f(U, V)$ と同じものを表し、 $Q_F = I - F(F^T F)^{-1} F^T$ である。

2.4 Chen の方法 (LM_S/LM_M)

Chen は、上の Wiberg 法同様、 f から U を消去し V のみの最小化問題へと変換した上で、ダンピング付きのニュートン法を適用するアルゴリズムを示している⁵⁾ (論文では V を消去しているが対称性からどちらを消去しても同じ。) 論文では、Wiberg 法が U を陽に計算する必要があるのに対し、Chen の方法はそれを要しない点で、消去の仕方が Wiberg 法と異なると主張している。しかしながら、 f から U を消去する方法は (7) 式のやり方以外にあり得ないことを考えれば、最小化しているもの自体が同一という意味で、基本的に同じ方法であると分かる。

文献⁵⁾ では LM_S, LM_M という 2 つの方法が示されている。LM_S は分解不定性を拘束しない方法であり、DN で $\mu_1 = \mu_2 = 0$ としたとき同様、ヘッセ行列が縮退するが、ダンピングファクタ λI の加算によって、更新量が定まらないことを免れる構造になっている。LM_M は解空間をグラスマン多様体上に限定して分解不定性を拘束する。その方法は文献¹³⁾ を下敷きにしてある。

LM_M は具体的には次のようなアルゴリズムである。現在の推定値 v は、直交行列 ($vv^T = I$) である。 v の列空間の直交補空間を列空間にもつ、ある $n \times (n - r)$ 行列を v_\perp としたとき、 $v \leftarrow v + v_\perp \Delta k$ と更新することとし、この Δk をダンピング付きニュートン法で計算する。そのため Δk (の成分) に対する g の勾配 b およびヘッセ行列 H が計算され、

$$(H + \lambda I) \delta k = b \quad (10)$$

を解いて更新量を決める (δk は Δk の成分を並べたベクトル)。更新された v は直交行列に直され、以後収束するまでこれを繰り返す。

文献⁵⁾ では、Wiberg 法は (9b) 式の規模 ($p \times nr$) が大きすぎて比較実験が不可能であるとし、他の手法との比較を行っていない (なお、LM_S ではその規模は $nr \times nr$ 、LM_M で $(n - r)r \times (n - r)r$ である。) しかしながら次節に述べるように、(9b) 式を正規方程式に直せば規模は同一 ($nr \times nr$ となる) であり、これは本質的な問題ではない。より本質的な違いは、Chen の方法がヘッセ行列を近似なしに求めること (Wiberg 法はガウス・ニュートン近似をする)、ダンピングを用いること、LM_M で分解不定性を拘束するために、グラスマン多様体上の解探索を行うことの 3 つである。

2.5 Damped-Wiberg 法 (DW)

上述のように Wiberg 法の更新式 (9b) を、等価な正規方程式に直せば、LM_M/LM_S の

それと同規模となる．さらにダンピングを加えて大域収束性を向上させることを考える．われわれは，さらに分解不定性の拘束を効率良く行う実装方法を考案したので，それらを述べる．

(9b) 式の $\|Q_F G \delta v - Q_{FY}\|^2$ のノルムの内部に，左から $(Q_F G)^T$ を掛けると，この最小化問題と等価な正規方程式

$$G^T Q_F G \delta v = G^T Q_{FY} \quad (11)$$

を得る．ただし 2.3 節で述べたように， $Q_F G$ したがって $G^T Q_F G$ がフルランクでないため，この方程式では δv は一意に定まらない．

最も簡単な対策は， $\|\delta v\|^2$ が最小になる解を選ぶことである．このように選んだ更新量は， V の分解不定性 ($V \leftrightarrow VA^T$) に寄与しない． $G^T Q_F G$ のランクが (データ自身の縮退がなければ) ちょうど $nr - r^2$ になると分かっているので¹²⁾，この知識を使えば特異値分解によりそのような解を求められる．

しかしこれよりも効率の良い方法がある．それは， $G^T Q_F G$ の零空間が陽に表現できることを利用することである．この零空間は，現在の v の成分を並べて作った $nr \times r^2$ 行列 M の列空間になる．このことから， δv をこの空間に直交するように

$$G^T Q_F G \delta v = G^T Q_{FY}, \quad (12a)$$

$$M \delta v = 0 \quad (12b)$$

を連立させて δv を決めることができる．(12a) の両辺のベクトルはともに同一の $nr - r^2$ 次元部分空間にあるので，この連立方程式は $(G^T Q_F G + MM^T) \delta v = G^T Q_{FY}$ を解くのと同じで，しかも左辺の行列はフルランクになる．そこでこの方程式にダンピング項を加えた

$$(G^T Q_F G + MM^T + \lambda I) \delta v = G^T Q_{FY} \quad (13)$$

を更新式とする．

なお， $Q_F = I - F(F^T F)^{-1} F^T$ は $p \times p$ 行列 (p は観測成分数) だが， $r \times r$ の小行列をブロック成分とするブロック対角行列なので，そのことを使えば計算量と記憶域を大幅に節約できる．また F もブロック対角行列となるので，その各ブロック部分のみを QR 分解したその直交行列 Q の成分を使うと Q_F の計算が著しく効率化できる．以下の実験ではこのような効率的な実装を施したものをを用いている．

2.6 Mitra-Sheorey-Chellappa の方法 (MF-LRSDP)

Mitra らは⁶⁾ で，行列分解の問題を低ランク半正定値計画問題 (LRSDP/low-rank semidefinite program) として定式化することを提案した．LRSDP の効率のよい解法として拡張ラグランジュ法 (augmented Lagrangian method) が知られており¹⁴⁾，準ニュートン法である BFGS 法に基づいて計算が行われる．また，事前知識の導入が簡単であることも利点であるという．DN 同様に，この方法でも (3) 式の正則化項付きの $f_R(U, V)$ を最小化する．紙面の都合で方法の詳細な説明は文献⁶⁾ に譲る．

3. 実験

3.1 実画像を用いた実験

まず表 1 に示す 3 つのデータを用いて，各方法を同一の条件で比較した．なお，各方法について述べた文献それぞれに次のような問題があることを指摘しておく．Chen の LMLS/M の論文⁵⁾ では先述の通り Wiberg の方法が扱われていない．Zhao らの SALS の論文⁷⁾ では，ランダムな初期値 (あるいは同等の公平な初期条件) から計算を始めたときの結果が示されていない．また，Mitra らの MFLRSDP の論文⁶⁾ では，ニュートン法系の方法が DN のみで，LMLS, LLM, Wiberg 各方法が比較対象となっていない．

表 1 3 つの実画像データの詳細

	Dinosaur	Giraffe	Face
観測行列サイズ	319 × 72	240 × 166	2596 × 20
ランク	4	6	4
観測成分数	5302	27794	33702
欠損成分の割合	76.92%	30.24%	35.09%

3 つのデータはそれぞれ，Dinosaur は SFM，Giraffe は非剛体 SFM，Face は異方向照明下の複数画像を使った IBR に相当する．Dinosaur は固定カメラ前でおもちゃの恐竜をターンテーブル上で回転させた 36 枚の画像の特徴点を追跡したものである．Giraffe は歩くキリンを並走するカメラで撮影した動画画像上で特徴点を追跡したものである．Face は人の顔を，照明方向を変えながら 20 フレーム撮影した画像の輝度を，[0:1] に正規化し，列ベクトルとして行列に格納したものである．

3.1.1 実験条件

大域収束の判定に重要となる終了条件は，次のように定めた．MFLRSDP 以外は，目的関数の変化量が $|J_{k-1} - J_k| < 10^{-9} J_k$ の条件を満たすか，反復回数が 500 回を超えた場合

に計算を終了した．MFLRSDP は，終了条件は文献⁶⁾に従い，事前に指定する計算時間の許容最大値を，Dinosaur および Face では 10 分，Giraffe は 1 時間とした．

各データに対し，初期値をランダムに変えて何度も試行を繰り返した． v の初期値を MATLAB のコード `randn(m,r)` によって与え， U の初期値が必要な方法では， v を固定したときの最小 2 乗問題を解いて (ALS の 1 反復の半分に相当)，それを与えた．このとき，全方法で同一の初期値を与えるようにした．Dinosaur と Face では試行回数を 100 回，ただし Face では計算時間がかかりすぎるため DN のみ除外した．Giraffe は 25 回試行した．

方法ごとのパラメータは以下のように，文献で推奨されているものを用いた．SALS で指定する成分の範囲 $[a, b]$ は，観測成分の最小値および最大値を用い， $\lambda = 0.01$ ，内部ループ最大反復数 $K = 60$ ， $\delta = (b - a)/100$ ， $x_0 = W \odot Y$ とした．なお以下で示す SALS の反復回数は外部ループの回数である．DN および MFLRSDP の目的関数の正則化項のパラメータは 3 つのデータすべてで $\mu_1 = \mu_2 = 10^{-3}$ とした．

以上の条件の下で，各試行終了後の残差，反復回数および実行時間を記録した．残差は観測成分についての総和を観測成分数 p で除した

$$\text{rms} = \sqrt{\frac{\|W \odot (Y - UV^T)\|_F^2}{p}} \quad (14)$$

を用いる．実行時間は，それぞれの方法を MATLAB 上で実行したときの CPU 時間である．同じアルゴリズムであっても，実装の並列計算と相性や，キャッシュの影響があり得るため，厳密な比較には適さないが，参考にはなる．すべての実験は，64 ビットの Windows OS と 32GB メモリ，3.16GHz の Xeon CPU を持つ PC 上の 64 ビット版の MATLAB(7.9.0.529/R2009b) で行った．

3.1.2 Dinosaur

Dinosaur の結果を図 1 に示す．図はすべて，100 回行った試行に関する累積ヒストグラムである．左から残差，反復回数，実行時間である．MFLRSDP のみ，反復回数の意味が異なるため図から省略した．

同図の残差のヒストグラムから (ランダムな初期値から開始した場合の) 大域収束性は，DW が圧倒的によいことがわかる．DW は 100 回中 98 回大域解に収束しており，下位 3 つの MFLRSDP, SALS, DN はもちろん，2,3 番手の LM_M および LM_S の 50-60 回と比べてもはつきり差をつけている．なお，DN, LM_S/M, MFLRSDP の各方法の結果は，文献⁴⁾⁻⁶⁾ に示されているものとほぼ一致している．SALS は，文献⁷⁾ にランダムな初期値か

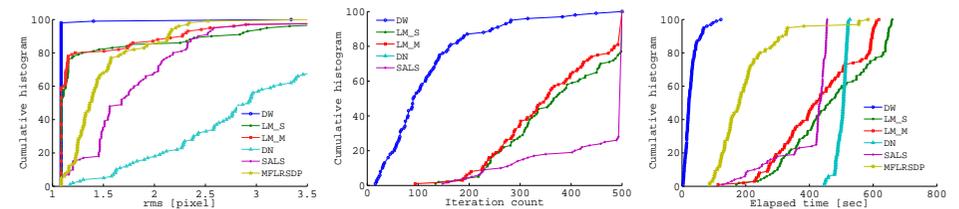


図 1 Dinosaur の結果．左から残差，反復回数，実行時間の累積ヒストグラム．

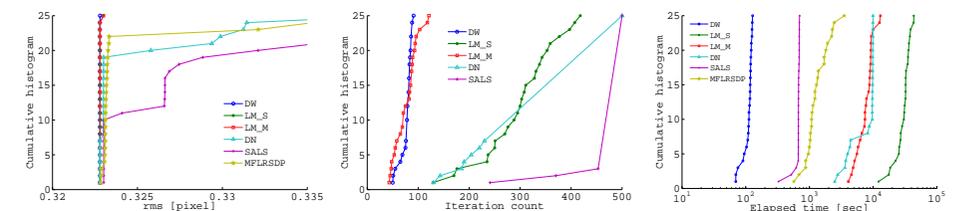


図 2 Giraffe の結果．左から残差，反復回数，実行時間の累積ヒストグラム．

ら開始した結果が報告されておらず，不明である．

また反復回数のヒストグラムから，反復回数でも DW は他手法を圧倒していることがわかる．その結果，計算時間も同図からわかるように，非常に低くなっている．

3.1.3 Giraffe

Giraffe の結果を図 2 に示す．図の配置は上と同様である．大域収束性は DW, LM_M, LM_S の 3 つの方法がほぼ 100%大域解に収束している．DN, MFLRSDP, SALS の順で試行の大半で大域解近くに収束しているが，厳密には多くが大域解には収束していないことも分かる．反復回数で見ると，DW, LM_M がほぼ同等，LM_S がこれに劣り，DN と SALS は試行の多くで上限回数に設定した 500 回に到達してしまっている．計算時間は DW がもっとも少なく，これに SALS, MFLRSDP が続くが，この 2 つはほとんど大域解に収束していないのであまり意味はなからう．LM_M は DW と反復回数こそ同じだが，反復一回の計算時間が大幅に長いことから，70 倍もの計算時間がかかっている．

3.1.4 Face

Face の結果を図 3 に示す．図の配置は上と同様である．先述の通り，DN は時間がかかりすぎるためこの実験からは除外している．

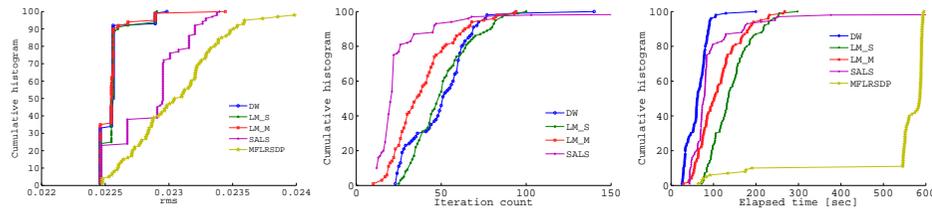


図 3 Face の結果．左から残差，反復回数，実行時間の累積ヒストグラム．

図から，大域収束性については DW と LM_M が同等で，LMS がこれに匹敵し，SALS がこれに続く．MFLRSDP は Dinosaur 同様，ほとんどの試行で大域解に収束し得ていないことがわかる．次に反復回数についてみると，SALS が少なく，LM_M, DW, LM_S と続く．ただし計算時間を見ると，DW がもっとも少ない．

なお SALS は，他の 2 つのデータと比べると，Face で特によい性能を示す．これは，成分値の範囲を制約するという SALS の事前情報の与え方が，Face の問題設定できわめて有効であることによると思われる．

3.2 シミュレーションデータ

欠損成分の割合やデータの性質を制御してその影響をみるため，シミュレーション実験を行った．以下では特に， 500×500 程度の大きなサイズの行列に対する性能を調べた．これは，MFLRSDP の論文⁶⁾で，ヘッセ行列を陽に求める必要のあるニュートン法系の方法は扱えないとあり，これを検証するためである．

3.2.1 データの生成方法と実験条件

実験はアフィンカメラの SFM を対象に，2 通りの画像系列と 2 種類の観測/欠損マスクの組み合わせによる，3 通りのデータを用いた．

1 つ目の系列は，空間に点群をランダムに発生させ，その周囲 360 度を正射影カメラが動いた場合である．点群は xyz 各軸につき $[-100, 100] \times [-100, 100] \times [0, 200]$ の範囲で一様乱数によって発生させた．以降，これを「回転運動」と呼ぶ．2 つめの系列は，点群を xyz 各軸について $[-100, 100] \times [0, 720] \times [0, 100]$ の範囲で一様乱数によって発生させ，これに対し正射影カメラを， xz 平面上から y 軸正方向に並進させた場合である．この場合を「並進運動」と呼ぶ．いずれの場合も，点の数を M とし視点の数を n とすると，行列のサイズは $2M \times n$ となる．いずれの場合も，画像上の投影点にガウス性のノイズを加算し，誤差をシミュレートした．ノイズの標準偏差の大きさは $[0.5 : 3.0]$ の範囲で変化させたが，傾向

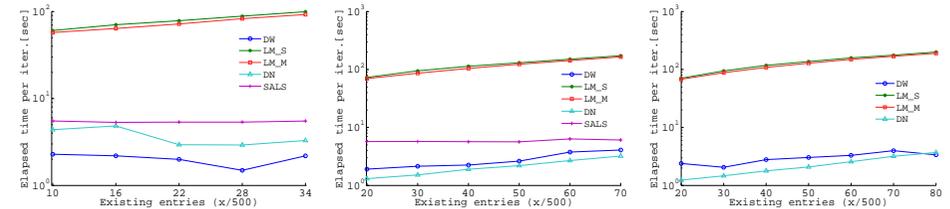


図 4 ニュートン法に属する方法の反復 1 回あたり計算時間．左から回転運動/ランダムマスク，回転運動/バンドマスク，並進運動/バンドマスク

がほとんど変わらなかったため，以下では 0.5 の場合のみを示す．

こうして得られるデータに，2 種類の観測/欠損マスク w を考えた．一つは，観測行列 Y の 1 行あたりの観測成分数を決め，ランダムにマスクを定めるもの．もう一つは，実データの Dinosaur のように， w の対角成分付近にバンド状に決定論的に定めたものである．以下では，回転運動とランダムマスク，バンドマスクの組み合わせおよび，並進運動とバンドマスクの組み合わせの計 3 通りに対して行った結果を示す．終了条件は上の実データを用いた実験と同じとした．

3.2.2 ニュートン法系の方法の実行速度

論文⁶⁾にあるように，DN の実行時間は遅く，統計情報を得るため何度も実行するのは難しかった．このことは LM_M, LM_S にも当てはまった．MFLRSDP 以外の方法について，反復一回あたりの計算時間を記録したものを図 4 に示す．SALS の指定パラメータは，値の範囲 $[a, b]$ には観測値の最大・最小を，内部ループ反復回数 $K = 60$ ， $\delta = (b - a)/100$ とそれぞれ設定し， $x_0 = w \odot Y$ とした．DN の指定パラメータは実行時間がかかり過ぎるため，高速に計算が可能な $\mu_1 = \mu_2 = 0$ とした．

図 4 は右から，回転運動/ランダムマスク，回転運動/バンドマスク，並進運動/バンドマスクの結果であるが，傾向はほぼ同じである．DW と DN が早く，これに SALS が続き，LM_M, LM_S がそれらより一桁遅い．DN の早さは， $\mu_1 = \mu_2 = 0$ とセットしたことで行列が疎になったことが原因であるが，反復回数が多いので結局のところ計算時間は長い．なお，並進運動の場合，成分値の有効な範囲を指定できないため，SALS は除外した．

3.2.3 DW と MFLRSDP の比較

この結果を踏まえて，DW と MFLRSDP のみの比較を行った． w の 1 行あたりの観測成分数を変えながら，大域収束性を調べた．具体的には，ランダムな初期値から開始し，20

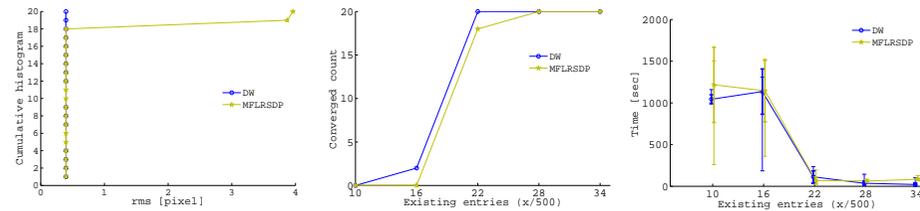


図 5 回転運動/ランダムマスクの結果．左：1 行 (500 個) につき観測数が 22 個の場合の残差の累積ヒストグラム．中：1 行あたり観測数に対する大域最小解に収束した回数．右：同観測数に対する実行時間 (太い線は標準偏差 σ として区間 $[-\sigma : \sigma]$ を示す)．

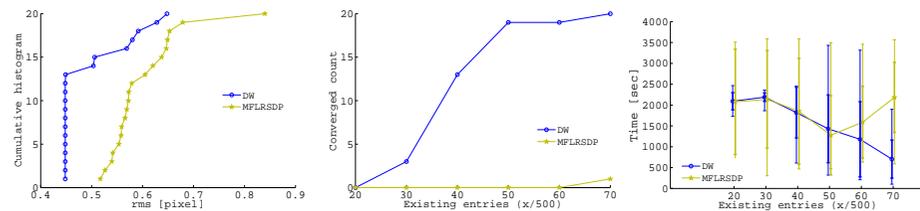


図 6 回転運動/バンドマスクの結果．左：1 行 (500 個) につき観測数が 40 個の場合の残差の累積ヒストグラム．中：1 行あたり観測数に対する大域最小解に収束した回数．右：同観測数に対する実行時間 (太い線は標準偏差 σ として区間 $[-\sigma : \sigma]$ を示す)．

度の試行を行って残差の度数分布を元に、大域解への収束の正否を調べた。MFLRSDP の最大許容時間はランダムマスクで 1500 秒、バンドマスクで 3500 秒に設定した。また、(3) 式の $\mu_1 = \mu_2 = 10^{-3}$ とした。

3.2.3.1 回転運動とランダムマスク

回転運動とランダムマスクの場合の結果を図 5 に示す。左図は、観測成分数が 1 行あたり 22 個の場合の残差の累積ヒストグラムである。DW が 20 回すべての試行で、MFLRSDP は 18 回の試行で大域解に収束していることがわかる。いくつかの観測成分数について、このように大域解への収束試行数を [10:34] の範囲で変えたのが中央の図である。DW が若干優れるものの、ほぼ同程度の性能と言える。また、図 5 の右は実行時間の平均および上下限を表示したものである。両手法がほぼすべて収束した、1 行あたり 22 以上の成分数の場合、計算時間も両者はほぼ同レベルであることがわかる。

3.2.3.2 回転運動とバンドマスク

回転運動とバンドマスクの場合の結果を図 6 に示す。同図左は行列の 1 行あたり観測成分

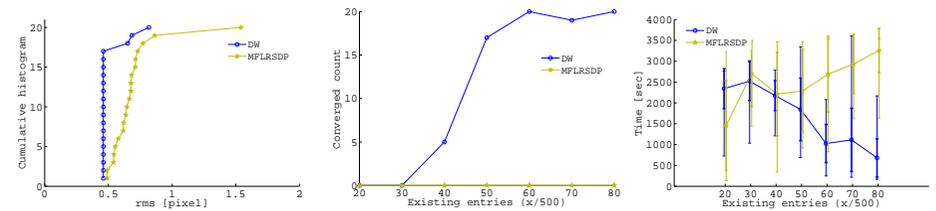


図 7 並進運動 (バンドマスク) の結果．左：1 行 (500 個) につき観測数が 50 個の場合の残差の累積ヒストグラム．中：1 行あたり観測数に対する大域最小解に収束した回数．右：同観測数に対する実行時間 (太い線は標準偏差 σ として区間 $[-\sigma : \sigma]$ を示す)．

数が 40 個の場合の残差を示し、同図中央はこれを [20:70] の範囲で変えたときの大域解への収束回数である。ランダムマスクの場合と異なり、MFLRSDP はまったく大域解に収束していないことが分かる。ただし多くの場合で収束条件を満たして終了したので、局所解には収束していると思われる。一方 DW は 40 個あたりから徐々に、50 個以上ではほとんどが大域解に収束している。同図右の計算時間をみると、DW は局所解に収束した MFLRSDP と同程度の時間で計算を終えていることが分かる。

3.2.3.3 並進運動とバンドマスク

並進運動とバンドマスクの場合の結果を図 7 に示す。1 行あたり観測成分数が 50 個の場合の残差が同図左、これを [20:70] の範囲で変えたときの大域解への収束回数が同図中央である。傾向は上の回転運動/バンドマスクの場合とほぼ同じである。MFLRSDP がまったく大域解に収束せず、DW は 40 個あたりから徐々に、50 個以上ではほとんどが大域解に収束している。同図右の計算時間も同様であった。

3.2.4 DW と MFLRSDP の比較のまとめ

上で示したように、マスクがランダムな場合 MFLRSDP は高い (DW に匹敵する) 性能を示す一方で、マスクがバンド状である 2 つのデータで MFLRSDP は全くふるわなかった。論文⁶⁾では、MFLRSDP は、欠損率の高い 500×500 の観測行列に対して高い収束性能を示したことが報告されている。ただしそこで実験に用いられたデータは、正規乱数で生成した 2 つの行列の積を観測行列とし、マスクを乱数によって決定したものであった。このデータは、上の実験では回転運動/ランダムマスクの場合に近く、ここでの結果は論文⁶⁾の結果と矛盾しない。

DW は、いずれの場合でも (相対的に) よい結果を与えた。大規模な問題では、記憶域および計算時間の点で、ニュートン法より準ニュートン法がより現実的な解になるというの

は一般的な見解である¹⁵⁾。しかし、行列分解の問題では、特に大域収束性を重視する場合、記憶域の問題がないという条件下では、DWをはじめとするニュートン法系の方法がむしろ有利かもしれない。

4. 結 論

欠損を含む行列分解の問題に対し、特に最近提案されている複数の手法を比較した。結果は、双線形性を用いて一方の変数を消去する Wiberg スタイルの方法が、大域収束性能および計算時間の両面で優れていることが分かった。特に、ガウスニュートン法である Wiberg 法を修正し、更新量の計算式を正規方程式に直し、さらにダンピングを加えたものももっとも優れていた。今後の課題は、なぜ大域収束性能が変数を消去することで向上するのかと、それが欠損成分の分布の仕方にどのように依存するかを調べることである。

参 考 文 献

- 1) Eriksson, A. and Hengel, A. V.D.: Efficient Computation of Robust Low-Rank Matrix Approximations in the Presence of Missing Data using the L1 Norm, *Proc. CVPR* (2010).
- 2) Shum, H., Ikeuchi, K. and Reddy, R.: Principal component analysis with missing data and its application to polyhedral object modeling, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.17, No.9, pp.855–867 (1995).
- 3) Wiberg, T.: Computation of principal components when data are missing, *Proc. Symposium of Comp. Stat.*, pp.229–326 (1976).
- 4) Buchanan, A.M. and Fitzgibbon, A.W.: Damped Newton Algorithms for Matrix Factorization with Missing Data, *Proc. CVPR* (2005).
- 5) Chen, P.: Optimization Algorithms on Subspaces: Revisiting Missing Data Problem in Low-Rank Matrix, *Int. J. Comput. Vision*, Vol.80, pp.125–142 (2008).
- 6) Mitra, K., Sheorey, S. and Chellappa, R.: Large-Scale Matrix Factorization with Missing Data under Additional Constraints, *Proc. Advances in Neural Information Processing Systems* (2010).
- 7) Zhao, K. and Zhang, Z.: Successively alternate least square for low-rank matrix factorization with bounded missing data, *Computer Vision and Image Understanding*, Vol.114, pp.1084–1096 (2010).
- 8) Jacobs, D.W.: Linear fitting with missing data for structure-from-motion, *Computer Vision and Image Understanding*, Vol.82, No.1, pp.57–81 (2001).
- 9) Martinec, D. and Martinec, D.: 3D Reconstruction by Fitting Low-rank Matrices with Missing Data, *Proc. CVPR*, pp.198–205 (2005).

- 10) Guilbert, N., Bartoli, A. and Heyden, A.: Affine Approximation for Direct Batch Recovery of Euclidian Structure and Motion from Sparse Data, *Int. J. Comput. Vision*, Vol.69, pp.317–333 (2006).
- 11) Tardif, J.-P., Bartoli, A., Trudeau, M., Guilbert, N. and Roy, S.: Algorithms for Batch Matrix Factorization with Application to Structure-from-Motion, *Proc. CVPR*, pp.1–8 (2007).
- 12) Okatani, T. and Deguchi, K.: On the Wiberg Algorithm for Matrix Factorization in the Presence of Missing Components, *Int. J. Comput. Vision*, Vol.72, pp.329–337 (2007).
- 13) Manton, J.H., Mahony, R. and Hua, Y.: The geometry of weighted low-rank approximations, *IEEE Trans. on Signal Processing*, Vol.51, No.2, pp.500–514 (2003).
- 14) Burer, S. and Choi, C.: Computational enhancements in low-rank semidefinite programming, *Optimization Methods and Software*, Vol.21, No.3, pp.493–512 (2006).
- 15) Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A.: Bundle Adjustment — A Modern Synthesis, *Vision Algorithms: Theory & Practice* (Triggs, B., Zisserman, A. and Szeliski, R., eds.), Springer-Verlag LNCS 1883 (2000).