行列平方根計算のためのニュートン法の高速化に関する研究

水野　匠　　　指導教員：臼田　毅

1　はじめに

行列平方根とは N 次正方行列 A に対して，$X^2 = A$ を満たす行列 X のことである．行列平方根を求める計算は量子情報科学や物性物理の分野において現れ，その高速かつ高精度な数値計算法の需要がある．これまでに，行列平方根を求めるための手法はいくつか提案されており，それらの手法はニュートン法 [1], [2] に基づいていることが多く，ニュートン法は基本的で重要な行列平方根を求める方法として見られている．ニュートン法は適当な条件下で局所的 2 次収束，すなわち解に近い初期値を与えることで効率良く収束する．しかし，初期値は元の行列 A と可換という制約条件があるため，従来の初期値は単位行列や元の行列 A とされていた．

先行研究 [3] では，先述の制約条件を満たし，従来の初期値を用いたニュートン法の反復回数よりも少ない回数で収束する初期値候補が導出されている．しかし，初期値候補を導出するための計算時間は無視できず，従来の初期値を用いたときよりも総計算時間が遅くなる場合がある．

そこで本研究では，解に近い初期値を効率良く与える試みとして，可換という制約条件を満たす行列多項式の集合の中から解との距離を近似した別の目的関数を最小化する初期値を提案する．そして，数値実験によりその有効性を検証する．

本稿では，$\Lambda(A)$ を A の固有値の集合，$\mathbb{R}^- := \{x \in \mathbb{R} \,|\, x \le 0\}$ とする．

2　ニュートン法

行列平方根を求めるためのニュートン法を Algorithm 1 に示す．

Algorithm 1 行列平方根を求めるニュートン法

$AX_0 = X_0 A$ となるような X_0 とする．
for $k = 0, 1, \ldots,$ until convergence **do**
　$X_{k+1} = \frac{1}{2}(X_k + X_k^{-1}A)$
end for

また，行列の平方根とニュートン法について以下の定理が知られている．

定理 1（行列の主平方根 [2, p.20]）
$\forall\lambda \in \Lambda(A)$，$\lambda \notin \mathbb{R}^-$ となるような $A \in \mathbb{C}^{N\times N}$ とする．そのとき，すべての固有値が右半平面にある行列 A の平方根が一意に存在する．この平方根を $A^{1/2}$ と書き，A の主平方根と呼ぶ．

定理 2（Algorithm 1 の収束性 [2, p.140]）
$\forall\lambda \in \Lambda(A)$，$\lambda \notin \mathbb{R}^-$ となるような $A \in \mathbb{C}^{N\times N}$ とする．もし，$A^{-1/2}X_0$ の固有値が右半平面にあるならば，そのとき X_k は大域的収束性 ($\lim_{k\to\infty} X_k = A^{1/2}$) を有し，$A^{1/2}$ に 2 次収束する．つまり任意の劣乗法性を満たす行列ノルムに対して以下のような関係が成り立つ．

$$\left\| X_{k+1} - A^{\frac{1}{2}} \right\| \le \frac{1}{2} \left\| X_k^{-1} \right\| \cdot \left\| X_k - A^{\frac{1}{2}} \right\|^2. \tag{1}$$

3　ニュートン法の初期値

ニュートン法の初期値 X_0 は元の行列 A と可換であるという制約条件を満たすために，以下に示す行列多項式の集合を用いる．

$$\mathcal{P}_n(A) := \left\{ \sum_{i=0}^{n} c_i A^i \,\middle|\, c_i \in \mathbb{R} \right\}.$$

この集合は，従来の初期値の選び方である $X_0 = I$ や $X_0 = A$ を含む．実際，$n = 0$，$c_0 = 1$ のとき $X_0 = I$ に，$n = 1$，$c_0 = 0$，$c_1 = 1$ のとき $X_0 = A$ となる．したがって，集合 $\mathcal{P}_n(A)$ の中によりよい初期値があると予想し，以下の 2 ノルム最小化問題を解く．

$$\min_{X\in\mathcal{P}_n} \|A - X^2\|_2^2 \Leftrightarrow \min_{X\in\mathcal{P}_n} \max_{\|\boldsymbol{x}\|_2=1} \|(A - X^2)\boldsymbol{x}\|_2^2. \tag{2}$$

ただし，計算コストの観点からこの最適化問題を解くのは現実的ではないと思われる．そこで，本研究では $\boldsymbol{x}$ の属する集合 ($N-1$ 次元球面) を $m-1$ 次元球面 S に制限した以下の最適化問題を考える．

$$\min_{X\in\mathcal{P}_n} \max_{\boldsymbol{x}\in S} \|(A - X^2)\boldsymbol{x}\|_2^2. \tag{3}$$

すなわち，$\mathbb{R}^N$ の m 次元部分空間の正規直交基底を並べた行列を $V_m := [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m]$ とすると式 (3) は次のようになる．

$$\min_{X\in\mathcal{P}_n} \max_{\|\boldsymbol{y}_m\|_2=1} \|(A - X^2)V_m\boldsymbol{y}_m\|_2^2. \tag{4}$$

簡単のため $n = 1$，$m = 1$，$\|\boldsymbol{v}_1\|_2 = 1$ となるような乱数ベクトル $\boldsymbol{v}_1$ とすると式 (4) は

$$\begin{aligned}
&\min_{X\in\mathcal{P}_1} \max_{|y|=1} \|(A - X^2)\boldsymbol{v}_1 y\|_2^2 \\
&\quad\Leftrightarrow \min_{X\in\mathcal{P}_1} \|(A - X^2)\boldsymbol{v}_1\|_2^2 \\
&\quad\Leftrightarrow \min_{c_0,c_1\in\mathbb{R}} \|(A - (c_0 I + c_1 A)^2)\boldsymbol{v}_1\|_2^2.
\end{aligned}$$

この問題は 2 変数関数 $f(c_0, c_1) := \|(A - (c_0 I + c_1 A)^2)\ \boldsymbol{v}_1\|_2^2$ の無制約最適化問題

$$\hat{c}_0, \hat{c}_1 = \arg\min_{c_0,c_1\in\mathbb{R}} f(c_0, c_1) \tag{5}$$

に帰着され，初期値 X_0 は $X_0 = \hat{c}_0 I + \hat{c}_1 A$ で与えられる．この無制約最適化問題 (5) の解き方に関して，$\boldsymbol{a} := A\boldsymbol{v}_1$，$\boldsymbol{b} := \boldsymbol{v}_1$，$\boldsymbol{d} := A(A\boldsymbol{v}_1)$ とすると

$$\begin{aligned}
f(c_0, c_1) &= \|\boldsymbol{a} - 2c_0c_1\boldsymbol{a} - c_0^2\boldsymbol{b} - c_1^2\boldsymbol{d}\|_2^2 \\
&= \|\boldsymbol{b}\|_2^2 c_0^4 + 4(\boldsymbol{a},\boldsymbol{b})c_0^3c_1 + 2\left(2\|\boldsymbol{a}\|_2^2 + (\boldsymbol{b},\boldsymbol{d})\right)c_0^2c_1^2 \\
&\quad - 2(\boldsymbol{a},\boldsymbol{b})c_0^2 - 4\|\boldsymbol{a}\|_2^2 c_0c_1 + 4(\boldsymbol{a},\boldsymbol{d})c_0c_1^3 \\
&\quad + \|\boldsymbol{d}\|_2^2 c_1^4 - 2(\boldsymbol{a},\boldsymbol{d})c_1^2 + \|\boldsymbol{a}\|_2^2,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial f}{\partial c_0} &= 4\|\boldsymbol{b}\|_2^2 c_0^3 + 12(\boldsymbol{a},\boldsymbol{b})c_0^2c_1 + 4\left(2\|\boldsymbol{a}\|_2^2 + (\boldsymbol{b},\boldsymbol{d})\right)c_0c_1^2 \\
&\quad - 4(\boldsymbol{a},\boldsymbol{b})c_0 - 4\|\boldsymbol{a}\|_2^2 c_1 + 4(\boldsymbol{a},\boldsymbol{d})c_1^3, \\
\frac{\partial f}{\partial c_1} &= 4(\boldsymbol{a},\boldsymbol{b})c_0^3 + 4\left(2\|\boldsymbol{a}\|_2^2 + (\boldsymbol{b},\boldsymbol{d})\right)c_0^2c_1 - 4\|\boldsymbol{a}\|_2^2 c_0 \\
&\quad + 12(\boldsymbol{a},\boldsymbol{d})c_0c_1^2 + 4\|\boldsymbol{d}\|_2^2 c_1^3 - 4(\boldsymbol{a},\boldsymbol{d})c_1.
\end{aligned}$$

次に $f(c_0, c_1)$ の停留点を求める．ここでグレブナー基底を利用して，2 変数代数方程式と等価な 1 変数の 2 つの代数方程式にする．このグレブナー基底は先行研究 [3] で求められており，これを利用する．そして，求めた停留点ごとのヘッセ行列を求める．ヘッセ行列と停留点の関係性として，ヘッセ行列の固有値がすべて正なとき関数 f は極小である．よって，この関係が成り立つ c_0, c_1 の組み合わせを初期値とする．Algorithm 2 にここまでの初期値導出の流れを示す．

Algorithm 2 対蹠点上（0 次元球面上）の最適化問題に基づく初期値の与え方．

1: グレブナー基底を用い，目的関数 $f(c_0, c_1)$ の停留点の集合を求める．
2: 停留点の集合の中から，$f(c_0, c_1)$ が最小となる組 ($\hat{c}_0$, $\hat{c}_1$) を求める．
3: $X_0 = \hat{c}_0 I + \hat{c}_1 A$ とする．

また，このときの初期値を用いた Algorithm 1 の 2 次収束性に関して，以下の命題が成り立つ．

命題 3 行列 $A \in \mathbb{R}^{N\times N}$，$\forall\lambda \in \Lambda(A)$，$\lambda \notin \mathbb{R}^-$ とし，$X_0 = c_0 I + c_1 A$ $(c_0, c_1 \in \mathbb{R})$ とする．もし，$c_0 > 0$ かつ $c_1 \geq 0$，または $c_0 \geq 0$ かつ $c_1 > 0$ ならば，そのとき X_k は $A^{1/2}$ に 2 次収束，すなわち任意の劣乗法性を満たす行列ノルムに対して以下のようになる．

$$\|X_{k+1} - A^{\frac{1}{2}}\| \leq \frac{1}{2}\|X_k^{-1}\| \cdot \|X_k - A^{\frac{1}{2}}\|^2.$$

4 数値実験

数値実験では，Algorithm 1 の初期値に Algorithm 2 を用いて求めた本研究の初期値，フロベニウスノルムを用い最小化して求めた初期値，従来の初期値の 3 つの初期値を用い，それぞれの計算時間の比較に関する実験を行った．なお，Algorithm 1 より $X_0 = I$ または $X_0 = A$ のどちらを初期値としても $X_1 = 1/2(I + A)$ となるため，従来法の初期値は $X_0 = 1/2(I + A)$ とした．また，数値実験には Intel(R) Core(TM) i7-2600 CPU @3.40GHz，MATLAB 7.13.0.564(R2011b) を用い，物性物理に現れる重なり行列 [4] を用い，収束判定条件は $\|A - X_k^2\|_{\mathrm{F}}/\|A\|_{\mathrm{F}} \leq 10^{-12}$ の時とした．

4.1 数値実験 1

数値実験 1 では，行列 (S_au864)[4] を用いて数値実験を行った．そのニュートン法の収束履歴を図 1 に，各手法ごとの行列平方根を求めるための総計算時間を図 2 に示す．図 1 の縦軸は相対残差の常用対数，横軸は反復回数を示し，図 2 の縦軸は計算時間 [sec.] を示す．

図 1 より，すべての初期値を用いたニュートン法は 2 次収束していることがわかる．実際に，フロベニウスノルム最小化に基づく初期値は $c_0 = 0.5358$, $c_1 = 0.4497$，本研究の提案した初期値は $c_0 = 0.6139$, $c_1 = 0.3809$ と命題 3 の十分条件を満たしている．また，3 反復目において本手法とフロベニウスノルム最小化に基づく手法は従来法よりも高精度の近似解を生成した．

図 2 の計算時間に関しては，本手法は従来法よりも少ない反復回数で収束判定条件を満たしたこと，およびフロベニウスノルム最小化に基づく手法と同じ反復回数であったが，式 (5) の $f(c_0, c_1)$ を求めるために要した計算時間が短かったため，結果として提案法が最も高速であった．

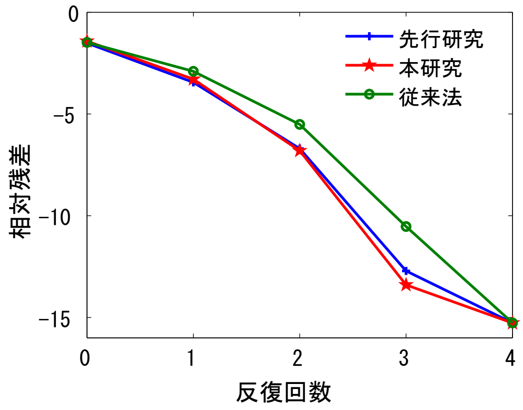

図 1 S_au864 に対する各手法ごとのニュートン法の収束履歴．

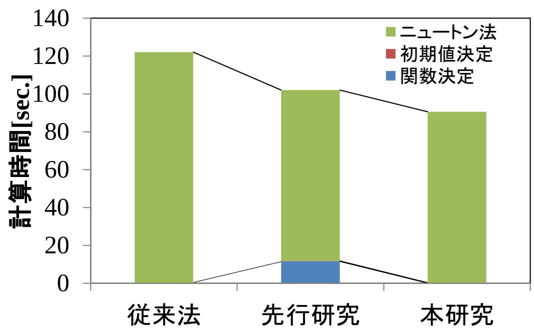

図 2 S_au864 に対する各手法ごとの計算時間．

5 まとめ

本研究ではニュートン法の初期値の与え方に着目し，球面制約上の最適化問題の解を初期値とする枠組みを提案した．さらにその枠組みの中から今回は簡単のため行列 1 次多項式の集合の中から，0 次元球面制約上の最適化問題に基づく初期値を与えた．今回は最も低次元な球面制約上の初期値の与え方であるものの，数値実験により本手法は従来法やフロベニウスノルム最小化に基づく手法よりも高速かつ高精度の近似解を生成する可能性が示された．

今後は Algorithm 2 で求めた c_0, c_1 が常に命題 1 を満たしているか等のさらなる数学的定理構築が課題となる．

参考文献

[1] N. J. Higham, Newton's method for the matrix square root, Math. Comp., **46** (1986), 537–549.

[2] N. J. Higham, Functions of Matrices: Theory and Computation, SIAM, Philadelpha, 2008.

[3] 森泉佳洋，行列平方根におけるニュートン法の初期値に関する研究，愛知県立大学大学院情報科学研究科修士論文，2012.

[4] ELSES, `http://www.elses.jp`.

公表論文

[A] **Sho Mizuno**, Yosuke Moriizumi, Tsuyoshi S. Usuda, Tomohiro Sogabe, "An initial guess of Newton's method for the matrix square root based on a sphere constrained optimization problem", JSIAM letters, (accepted).