Vol.2017-HPC-160 No.18
2017/7/27

実対称正定値行列を係数行列とする連立1次方程式の数値解に対する精度保証法

寺尾 剛史[1,a)]　尾崎 克久[2]　南畑 淳史[3,4]

概要：本稿では，実対称正定値行列を係数行列とする連立 1 次方程式の数値解に対する精度保証付き数値計算法を扱う．先行研究として，コレスキー分解 1 回の計算コストで高速精度保証可能な Rump-Ogita の方法がある．Rump-Ogita の方法では精度保証が成功しない悪条件な問題に対応するため，コレスキー分解で得られる上三角行列の近似逆行列を用いた精度保証法を提案する．

Fast Validated Solution of Linear Systems with Real Symmetric and Positive Definite Matrices

Terao Takeshi[1,a)]　Ozaki Katsuhisa[2]　Minamihata Atsushi [3,4]

1. はじめに

本論では連立 1 次方程式

$$Ax = b, \quad A = A^T \in \mathbb{R}^{n\times n}, \quad b \in \mathbb{R}^n$$

の数値解 $\hat{x}$ に対する誤差上限を与える精度保証付き数値計算法を扱う．ここで，A は正定値行列とする．先行研究として，シフト付きコレスキー分解を用いた方法 [1] が知られている（以後 Rump-Ogita 法と呼ぶ）．Rump-Ogita 法はコレスキー分解 1 回程度の計算量で精度保証が可能であり，計算コストにおいて非常に優れた手法である．ただし，行列のサイズが大きいとき，または行列が悪条件のときにこの方法は精度保証に失敗しやすい．

行列が対称正定値であることを仮定しない一般的な精度保証付き数値計算法が知られている．それらの多くは LU 分解を利用した精度保証法であり，LU 分解の議論をコレスキー分解に適応させることで実対称正定値の場合も精度保証を行うことが可能である．

[1] 芝浦工業大学大学院理工学研究科
[2] 芝浦工業大学システム理工学部
[3] 国立研究開発法人産業総合研究所生体システムビッグデータ解析オープンイノベーションラボラトリ
[4] 早稲田大学理工学術院
[a)] nb17105@shibaura-it.ac.jp

本論では，コレスキー分解を利用した新しい精度保証法を提案する．提案手法は Ogita-Rump 法よりも高コストであるが，より悪条件な問題まで精度保証できる．また，LU 分解を用いる精度保証法をコレスキー分解に適用させた手法よりは，計算コストの面で利点があることを示す．

2. 準備

この章では，本論で用いる記号の説明と数学的基礎の紹介を行う．$\mathbb{F}$ を IEEE 754 規格に基づく固定された精度における浮動小数点数 [2] の集合とし，u を $\mathbb{F}$ の精度における丸めの単位とする（例えば binary64 形式の場合 $u = 2^{-53}$ となる）．$fl(\cdot), fl_{\triangledown}(\cdot), fl_{\triangle}(\cdot)$ は括弧内の評価に必要なすべての演算に浮動小数点演算を用いるものとし，それぞれ最近点丸めのモード，下向き丸めのモード，上向き丸めのモードで計算するものとする．ただし，本論では浮動小数点演算でオーバーフロー・アンダーフローが発生しないことを仮定する．また，ベクトル $x, y \in \mathbb{R}^n$ に対して，絶対値と不等号の表記は

$$|x| = (|x_1|, |x_2|, \cdots, |x_n|)^T, \ x < y \Leftrightarrow x_i < y_i \text{ for } \forall i$$

を意味し，行列に対してもこれらを自然に拡張する．

行列の最大値ノルムに関して

$$\|A\|_\infty = \||A|e\|_\infty, \quad e = (1,\dots,1)^T \in \mathbb{R}^n$$

が成立する．また，$A, B \in \mathbb{R}^{n\times n}$ に対して不等式

$$\|AB\|_\infty \le \||A|(|B|e)\|_\infty \tag{1}$$

が成り立つ．式 (1) における上限は過大評価にはなるが，行列積のノルムの上限を行列積を直接計算せずに $\mathcal{O}(n^2)$ flops (floating-point operations) の計算量で得たいときに使用する．

実対称正定値行列 A に対して，

$$\|A^{-1}\|_2 = \frac{1}{\lambda_{\min}(A)} \tag{2}$$

が成り立つ．ここで，$\lambda_{\min}(A)$ は行列 A に対する最小固有値を表す．

三角行列 $T \in \mathbb{F}^{n\times n}$ の近似逆行列 $\hat{X}$ は，右辺を単位行列とする行列方程式 $XT = I$ の近似解とする．ただし，行列方程式の解法は前進代入・後退代入を行うものとする．

3. 高速な精度保証法

本章では，連立 1 次方程式の数値解に対する精度保証付き数値計算法の先行研究について紹介する．連立 1 次方程式に関して係数行列 $A \in \mathbb{F}^{n\times n}$ は実対称正定値行列とし，右辺ベクトルは $b \in \mathbb{F}^n$ とする．近似解 $\hat{x}$ に対する誤差上限は

$$\|\hat{x} - x\| \le \|A^{-1}\| \cdot \|b - A\hat{x}\| \tag{3}$$

と評価できる．この式を使用し，さらに浮動小数点演算を使用してこの上限を求めていく．式 (3) において，$\|b - A\hat{x}\|$ の上限に関しては $\mathcal{O}(n^2)$ flops の計算量で求められ，$\|A^{-1}\|$ の上限を求める計算に最も計算コストを要する．よって，以後は $\|A^{-1}\|$ の上限を高速に評価する方法に焦点を絞り，先行研究を紹介する．

3.1 Rump-Ogita の方法

本節では，シフト付きコレスキー分解を用いた精度保証付き数値計算法 [1] について紹介する．行列を $A \in \mathbb{F}^{n\times n}$ とし，A に対する数値計算を用いたコレスキー分解の計算結果を $\hat{R}$ とする．また，A と $\hat{R}$ に対して以下の関係式が知られている [3].

$$\hat{R}^T\hat{R} - A = \Delta A, \quad |\Delta A| \le (n+1)u|\hat{R}^T||\hat{R}| \tag{4}$$

また，式 (4) より，

$$\|\Delta A\|_2 \le \sum_{j=1}^{n} \gamma_{j+1} a_{jj} =: \rho, \quad \gamma_k = \frac{ku}{1-ku} \tag{5}$$

が成り立つ（詳しい導出は [4] を参照）．ここで，定数 $\bar{\rho}$ と行列 B を

$$(\rho \le)\ \bar{\rho} \in \mathbb{F}, \quad B = fl_\nabla(A - 2\bar{\rho}I)$$

とする（I は n 次単位行列とする）．ここで，B に対して浮動小数点演算を用いてコレスキー分解が成功したとき，このコレスキー分解により得られる上三角行列を $\hat{R}_B$，残差を $\Delta B = \hat{R}_B^T\hat{R}_B - B$ とする．このとき，

$$\lambda_{\min}(B) \ge \lambda_{\min}(\hat{R}_B^T\hat{R}_B) - \|\Delta B\|_2 \ge -\rho$$

であり，$\lambda_{\min}(A) - 2\bar{\rho} \ge \lambda_{\min}(B)$ であるから，

$$\lambda_{\min}(A) \ge \bar{\rho} \tag{6}$$

が成り立つ．よって，式 (6), (3), (2) より

$$\|\hat{x} - x\|_2 \le \frac{1}{\bar{\rho}}\|b - A\hat{x}\|_2$$

を得る．この方法で用いるシフト量 $\bar{\rho}$ の計算と対角シフトの計算量は $\mathcal{O}(n)$ flops である．よって Rump-Ogita 法は，コレスキー分解と同等の計算コストで連立 1 次方程式の数値解に対する精度保証が可能であり，非常に高速である．ただし，行列のサイズが大きいときや悪条件な場合は，原点方向に対角シフトをしたことによりコレスキー分解が破たんすることがある．

3.2 LU 分解を用いた精度保証法の適応

本節では，先行研究にある LU 分解を用いた精度保証法をコレスキー分解にそのまま適用させた連立 1 次方程式の数値解に対する精度保証付き数値計算法を紹介する．また，$\|A^{-1}\|$ の上限を得るための定理を紹介する．

定理 1
I を単位行列とする．$A \in \mathbb{R}^{n\times n}$ に対して

$$\|QA - I\| < 1 \tag{7}$$

を満たす $Q \in \mathbb{R}^{n\times n}$ が存在するとき A は正則であり，

$$\|A^{-1}\| \le \frac{\|Q\|}{1 - \|QA - I\|} \tag{8}$$

が成り立つ *[5]*.

式 (8) において，$\|QA - I\|$ の評価に最も計算コストを要するため，このノルムの上限の計算に焦点を絞って紹介する．

3.2.1 Oishi-Rump の方法 [6]

LU 分解を使用する精度保証法で最も高速な手法は Oishi-Rump の方法 [6] であり，この手法をコレスキー分解に適応させる．$A \in \mathbb{F}^{n\times n}$ とし，A に対する数値計算を用いたコレスキー分解の計算結果を $\hat{R}$ とする．また，$\hat{R}$ の近似逆行列を $\hat{X}$ とし，$\hat{R}^T$ の近似逆行列を $\hat{X}_t$ とする．ただし，$\hat{X}, \hat{X}_t$ は $XR = I, XR^T = I$ を前進代入・後退代入を用いてそれぞれ計算されたものとする．このとき，$\hat{R}$ と $\hat{X}, \hat{X}_t$ を得るのに必要な計算コストは合計で $n^3 + \mathcal{O}(n^2)$ flops で

ある．これらの行列を用い，また $Q = \hat{X}\hat{X}_t$ と置くことで，Oishi-Rump の方式をコレスキー分解に適用させた以下の定理が導出できる．

定理 2
$A \in \mathbb{F}^{n\times n}$ とする．A のコレスキー分解の計算結果を $\hat{R}$，その近似逆行列を $\hat{X}$ とする．また，$\hat{R}^T$ の近似逆行列を $\hat{X}_t$ とする．そのとき

$$\|QA - I\| \leq \|(2n+1)u|\hat{X}||\hat{X}_t||\hat{R}^T||\hat{R}| + nu|\hat{X}||\hat{R}|\| \tag{9}$$

が成り立つ．

定理 2 の評価式は $e = (1, 1, \ldots, 1) \in \mathbb{N}^n$ とすると

$$\begin{aligned}\|QA - I\|_\infty \leq &\|2(n+1)u|\hat{X}|(|\hat{X}^T|(|\hat{R}^T|(|\hat{R}|e))) \\ &+ nu|\hat{X}|(|\hat{R}|e)\|_\infty\end{aligned} \tag{10}$$

と計算できる．この式を評価する際に上向き丸めのモードを用いることで，浮動小数点のみを用いて上限が求まる．$\hat{R}, \hat{X}, \hat{X}_t$ を求めた後に必要な計算コストは $\mathcal{O}(n^2)$ flops である．よって，合計で $n^3 + \mathcal{O}(n^2)$ flops の計算量で精度保証が可能である．

3.2.2 Minamihata らによる方法

Minamihata らによって，M 行列や H 行列の性質を用いて Oishi-Rump の方法を改良した手法が提案されている [7]．この手法をコレスキー分解に適応させたものを以下に考える．Oishi-Rump の方法では，$Q = \hat{X}\hat{X}_t$ と定義したが，Minamihata ら手法では $Q = (\hat{R}^T\hat{R})^{-1}$ と定義する．

定理 3
$A \in \mathbb{F}^{n\times n}$ とする．A のコレスキー分解の計算結果を $\hat{R}$，$X\hat{R} = I$ の近似解を $\hat{X}$ とする．$I - nu|\hat{X}||\hat{R}|$ と $I - nu|\hat{X}_t||\hat{R}^T|$ が M 行列ならば $Q := (\hat{R}^T\hat{R})^{-1}$ が存在し

$$\begin{aligned}&|QA - I| \leq (n+1)uM_1|\hat{X}|M_2|\hat{X}_t||\hat{R}^T||\hat{R}| \\ &M_1 = (I - nu|\hat{X}||\hat{R}|)^{-1}, M_2 = (I - nu|\hat{X}_t||\hat{R}^T|)^{-1}\end{aligned} \tag{11}$$

が成り立つ．

定理 3 の評価式を最大値ノルムで評価する．まず，v_L, v_U を要素がすべて零より大きいベクトルとし，$u_L = M_2 v_L, u_U = M_1 v_U$ とする．このとき

$$\|QA - I\|_\infty \leq \alpha\|v_U\|_\infty$$

が成り立つ．ただし

$$\alpha = (n+1)u \max_{1\leq i\leq n} \frac{(|\hat{X}|v_L)_i}{(u_U)_i} \cdot \max_{1\leq i\leq n} \frac{(|\hat{X}_t||\hat{R}^T||\hat{R}|e)_i}{(u_L)_i}$$

である．

この手法の計算量は Oishi-Rump の方法と同等であり，$\|QA - I\|_\infty$ の上限が約 1/2 になる利点がある．

3.2.3 Ogita-Oishi の方法

次に，Ogita と Oishi らによる精度保証法 [8] をコレスキー分解に適応させた方法を紹介する．Oishi-Rump の方法と同じく $Q = \hat{X}\hat{X}^T$ を使用する．

定理 4
$A \in \mathbb{F}^{n\times n}$ とする．A のコレスキー分解の計算結果を $\hat{R}$，その近似逆行列を $\hat{X}$ とする．そのとき

$$\|QA - I\| \leq \||\hat{X}|(|\hat{X}^TA - \hat{R}| + nu|\hat{R}|)\| \tag{12}$$

が成り立つ．

式 (12) の $\hat{X}^TA - \hat{R} =: S$ に対して，$|S|$ の上限を

$$|S| \leq \max(|fl_\triangledown(\hat{X}^TA - R)|, |fl_\triangle(\hat{X}^TA - R)|) \tag{13}$$

と得る．ここで，$\hat{X}$ は三角行列であるため，$\hat{R}, \hat{X}$ を求めた後に, 式 (13) の評価に必要な計算量は $2n^3 + \mathcal{O}(n^2)$ flops である．合計の計算量は $\frac{8}{3}n^3 + \mathcal{O}(n^2)$ flops である．

3.2.4 Ozaki-Ogita-Oishi の方法

次に，式 (12) の評価を用いたもう一つの方法を紹介する [9]．式 (12) の $|\hat{X}^TA - \hat{R}|$ の上限を得るために，行列積に対する事前誤差評価を用いる．ここで，$A, B \in \mathbb{F}^{n\times n}$ の行列積に対して

$$|fl(AB) - AB| \leq nu|A||B| \tag{14}$$

が成り立つ．よって，$|S| = |\hat{X}^TA - \hat{R}|$ の上限は

$$|S| \leq |fl(\hat{X}^TA - \hat{R})| + (n+1)u(|\hat{X}^T||A| + |\hat{R}|) \tag{15}$$

と得る．式 (15) を使用して式 (12) の上限を求める場合，$fl(\hat{X}^TA - \hat{R})$ に必要な計算量は $n^3 + \mathcal{O}(n^2)$ flops であり，精度保証全体に必要な計算量は $\frac{5}{3}n^3 + \mathcal{O}(n^2)$ flops である．

3.3 先行研究のまとめ

以下に，先行研究とその計算量を示す．

表 1 先行研究における，計算手法と計算量の比較 (計算量順)

year	method	計算量 (flops)	比
2007	Rump-Ogita	$\frac{1}{3}n^3 + \mathcal{O}(n^2)$	1
2002	Oishi-Rump	$n^3 + \mathcal{O}(n^2)$	3
2011	Ozaki-Ogita-Oishi	$\frac{5}{3}n^3 + \mathcal{O}(n^2)$	5
2005	Ogita-Oishi	$\frac{8}{3}n^3 + \mathcal{O}(n^2)$	8

ここで，表の比は各手法とコレスキー分解の計算量（$\frac{1}{3}n^3 + \mathcal{O}(n^2)$ flops）の比を表す．これらは，連立 1 次方程式の数値解に対する高速な精度保証法として知られている．表より，Rump-Ogita の方法が精度保証に失敗した場合，少なくともコレスキー分解の 3 倍の計算量が必要であり，Ogita-Oishi の方法を用いる場合，コレスキー分解の計算量の 8 倍の計算量が必要である．

4. 提案手法

本章では，コレスキー分解を用いた精度保証法を提案する．まず，提案手法の導出に必要な補題を紹介する．

補題 5
$\hat{R} \in \mathbb{F}^{n\times n}$ を三角行列とし，$\hat{X}$ を $\hat{R}$ の近似逆行列とする．このとき

$$I - \hat{X}\hat{R} = \Delta R, \quad |\Delta R| \leq nu|\hat{X}||\hat{R}|$$

が成り立つ *[3]*.

補題 6
$F \in \mathbb{R}^{n\times n}$ とする．$\|F\| < 1$ であるとき

$$\|(I-F)^{-1}\| \leq \frac{1}{1-\|F\|} \tag{16}$$

が成り立つ．また，

$$(I-F)^{-1} = I + F(I-F)^{-1} \tag{17}$$

も成立する *[10]*.

上述の補題を用いて，$\|QA - I\|$ の上限に関する定理を提案する．

定理 7
$A \in \mathbb{F}^{n\times n}$ とする．A に対する数値計算を用いたコレスキー分解が成功し，得られた上三角行列を $\hat{R}$ とする．行列方程式 $X\hat{R} = I$ を前進代入して得られた近似逆行列を $\hat{X}$ とする．ここで，$\Delta A := A - \hat{R}^T\hat{R}, \Delta R := I - \hat{X}\hat{R}$ としたとき，$\|\Delta R\| < 1$ ならば

$$\|QA - I\| \leq \frac{\|\hat{X}\hat{X}^T \cdot \Delta A\| + \|S\|}{1 - \|\Delta R\|} \tag{18}$$

$$\|T\| \leq \frac{\|\hat{X}\hat{X}^T \cdot (\Delta R)^T\| \cdot \|\Delta A\|}{1 - \|(\Delta R)^T\|} \tag{19}$$

が成り立つ．

証明 $Q = (\hat{R}^T\hat{R})^{-1}$ と置く．仮定より，$\hat{R}, \hat{X}, I + \Delta R$ はそれぞれ正則である．よって，$\hat{X}\hat{R} = I - \Delta R$ より，

$$\hat{R}^{-1} = (I - \Delta R)^{-1}\hat{X} \tag{20}$$

が成り立つ．また，式 (20) の両辺を転置することにより

$$(\hat{R}^{-1})^T = \hat{X}^T(I - (\Delta R)^T)^{-1} \tag{21}$$

を得る．ここで，$Q = (\hat{R}^T\hat{R})^{-1}$ であるから，

$$QA - I = Q(\hat{R}^T\hat{R} + \Delta A) - I = Q \cdot \Delta A$$

が成り立つ．よって，式 (20) と式 (21) より

$$\begin{aligned} QA - I &= (\hat{R}^T\hat{R})^{-1}\Delta A = \hat{R}^{-1}(\hat{R}^{-1})^T\Delta A \\ &= (I - \Delta R)^{-1}\hat{X}\hat{X}^T(I - (\Delta R)^T)^{-1} \cdot \Delta A \end{aligned} \tag{22}$$

が成り立つ．ここで，補題 6，式 (17) より，

$$(I - (\Delta R)^T)^{-1} = I + (\Delta R)^T(I - (\Delta R)^T)^{-1}$$

であるから，

$$\begin{aligned} &QA - I \\ =&(I - \Delta R)^{-1}\hat{X}\hat{X}^T(I + (\Delta R)^T(I - (\Delta R)^T)^{-1}) \cdot \Delta A \\ =&(I - \Delta R)^{-1}(\hat{X}\hat{X}^T \cdot \Delta A + T) \end{aligned}$$

が成り立つ．ただし，$T = \hat{X}\hat{X}^T \cdot (\Delta R)^T(I - (\Delta R^T))^{-1} \cdot \Delta A$ である．よって補題 6，式 (16) より

$$\|QA - I\| \leq \frac{\|\hat{X}\hat{X}^T \cdot \Delta A\| + \|S\|}{1 - \|\Delta R\|}$$

$$\|T\| \leq \frac{\|\hat{X}\hat{X}^T \cdot (\Delta R)^T\| \cdot \|\Delta A\|}{1 - \|(\Delta R)^T\|}$$

が成り立つ． □

また，Minamihata らの方法のように，H 行列の性質を用いた計算法も適用可能であるが本稿では省略する．これより定理 7 に基づく 4 通りの計算方法を提案する．定理 7 において，$\|S\|$ は $\|\hat{X}\hat{X}^T \cdot \Delta A\|$ と比較して非常に小さいことが期待される．よって $\|\hat{X}\hat{X}^T \cdot \Delta A\|$ の評価が非常に重要である．以下．

$$\|\hat{X}\hat{X}^T \cdot \Delta A\|_\infty \leq \||\hat{X}\hat{X}^T|(|\Delta A|e)\|_\infty \tag{23}$$

により上限を得る方法について説明する．

はじめに，$|\Delta A|e$ について説明を行う．コレスキー分解の残差に対して，式 (4) が成り立つから

$$|\Delta A|e \leq (n+1)u|\hat{R}^T|(|\hat{R}|e) \tag{24}$$

が成り立つ．この式を上向き丸めのモードで計算することにより数値計算で上限を得る．この手法は $\mathcal{O}(n^2)$ flops の計算量で $|\Delta A|e$ の上限を評価することができる．また，丸めのモードの変更を利用した行列積の包含を行う場合，

$$|\Delta A|e \leq \max(|fl_\triangledown(\hat{R}^T\hat{R} - A)|, |fl_\triangle(\hat{R}^T\hat{R} - A)|)e \tag{25}$$

と評価できる．このときの計算量は $\frac{1}{2}n^3 + \mathcal{O}(n^2)$ flops である．

次に，$v \geq |\Delta A|e$ とし $|\hat{X}\hat{X}^T|v$ の上限を考える．高速に上限を得るために

$$|\hat{X}\hat{X}^T|v \leq |\hat{X}|(|\hat{X}^T|v) \tag{26}$$

と計算を行えば，計算量は $\mathcal{O}(n^2)$ flops である．また，行列積の事前誤差評価である式 (14) を用いて

$$|\hat{X}\hat{X}^T|v \leq |fl(\hat{X}\hat{X}^T)|v + nu|\hat{X}|(|\hat{X}^T|v) \qquad (27)$$

と評価できる．$fl(\hat{X}\hat{X}^T)$ の評価に必要な計算量は $\frac{1}{4}n^3 + \mathcal{O}(n^2)$ flops である．

以下に，計算手法と計算量の比較結果を表す (表 2)．ただし，表の計算量はコレスキー分解と近似逆行列の計算量を含む．

表 2 提案手法における，計算手法と計算量の比較

	計算手法 (ΔA)	計算手法 ($\hat{X}\hat{X}^T$)	計算量 (flops)
T1	式 (24)	式 (26)	$\frac{2}{3}n^3 + \mathcal{O}(n^2)$
T2	式 (24)	式 (27)	$\frac{11}{12}n^3 + \mathcal{O}(n^2)$
T3	式 (25)	式 (26)	$\frac{7}{6}n^3 + \mathcal{O}(n^2)$
T4	式 (25)	式 (27)	$\frac{17}{12}n^3 + \mathcal{O}(n^2)$

また，式 (23) において，$\hat{X}\hat{X}^T$ と ΔA を絶対値で分離せずに，区間演算にて評価をすると，より上限を過大評価しない手法を開発することが可能である．

5. 数値実験

本章では，数値実験結果について紹介する．まず，本論で用いるテスト行列について説明を行う．行列サイズ n=1024 とし．行列 A を MATLAB 上で

$$A = \texttt{gallery}(\text{`}\texttt{randsvd}\text{'}, n, -\text{cnd}, \text{mode}, n, n, 1);$$

と生成する．ここで cnd は A の条件数であり，mode により特異値の分布を設定する．また mode は以下の 5 つが用意されている．

1： 1 つの大きな特異値
2： 1 つの小さな特異値
3： 幾何学的に分布する特異値
4： 算術的に分布する特異値
5： 対数的に一様分布する乱数の特異値

このテスト行列を用いて，条件数の変化による $\|QA-I\|_\infty$ の上限の評価を比較する．はじめに，提案手法の 4 つを比較する．表 3, 4, 5 にそれぞれ $\text{mode}=1, \text{mode}=2, \text{mode}=3$ の場合の計算結果を示す．mode = 4 の場合は mode = 2，mode = 5 の場合は mode = 3 と類似した傾向であるため省略する．

表 3 $n = 1024, \text{mode} = 1$. $\|QA-I\|_\infty$ の上限の比較

cnd	T1	T2	T3	T4
10^8	$1.1*10^{-4}$	$7.1*10^{-5}$	$1.3*10^{-5}$	$7.5*10^{-6}$
10^{10}	$2.4*10^{-2}$	$6.4*10^{-3}$	$1.7*10^{-3}$	$8.0*10^{-4}$
10^{12}	–	$5.9*10^{-1}$	$1.5*10^{-1}$	$7.9*10^{-2}$

実験結果より，提案手法は計算量と精度がトレードオフの関係にある．ただし，表 4 より，mode = 2 の場合は T1 と T2，T3 と T4 がほぼ同等の精度となっている．これは，$\|XX^T\|_\infty \approx \||X||X^T|\|_\infty$ となっているためである．また，

表 4 $n = 1024, \text{mode} = 2$. $\|QA-I\|_\infty$ の上限の比較

cnd	T1	T2	T3	T4
10^8	$1.1*10^{-4}$	$1.1*10^{-4}$	$1.5*10^{-6}$	$1.5*10^{-6}$
10^{10}	$1.5*10^{-2}$	$1.5*10^{-2}$	$2.0*10^{-4}$	$2.0*10^{-4}$
10^{12}	–	–	$1.8*10^{-2}$	$1.8*10^{-2}$

表 5 $n = 1024, \text{mode} = 3$. $\|QA-I\|_\infty$ の上限の比較

cnd	T1	T2	T3	T4
10^8	$1.9*10^{-2}$	$1.8*10^{-3}$	$1.7*10^{-4}$	$2.3*10^{-5}$
10^{10}	–	$1.4*10^{-1}$	$1.2*10^{-2}$	$1.9*10^{-3}$
10^{12}	–	–	–	$1.6*10^{-1}$

特異値の分布が mode = 4 の場合でも同様の結果が得られる．

次に，Rump-Ogita 法の精度保証可能な条件数の上限を紹介する (表 6)．

表 6 $n = 1024, \text{mode} = 1, 2, 3$.
Rump-Ogita 法の計算可能な条件数の上限

mode	1	2	3
cnd	$7.9*10^{12}$	$1.0*10^{10}$	$2.5*10^{11}$

mode = 1 のとき，Rump-Ogita 法は提案手法 T2 や T3 と同様の条件数まで精度保証が可能である．また，mode = 2 のとき，Rump-Ogita 法は提案手法 T1 より適用範囲が狭いことがわかる．最後に，mode = 3 のとき，Rump-Ogita 法は提案手法 T2 と同等であり，T3 より適用範囲が狭い．つまり，提案手法 T4 は総じて Rump-Ogita 法より悪条件問題に対して精度保証が可能であり，mode = 2 のような場合では，提案手法 T1 も有効な場合がある．

次に，Ogita-Oishi 法，Ozaki-Ogita-Oishi 法との比較を行う．まず，提案手法 T4 と Ogita-Oishi の方法の比較を行う．図 1, 2 は，mode = 2, 4 の場合の，条件数における $\|QA-I\|_\infty$ の上限値を表す．図 1, 2 より，精度の差は非常に小さく mode = 1, 3, 5 の場合はさらに小さい．また，T4，Ogita-Oishi の方法の計算量は，それぞれ $\frac{17}{12}n^3, \frac{8}{3}n^3$ flops 程度であり，提案手法はほぼ同等な上限を得ながらも計算量を大幅に低減できている．

次に，T2，T3，Ozaki 等の方法の比較を行う．図 3, 4 は，mode = 1, 4 の場合の，条件数における $\|QA-I\|_\infty$ の上限値を表す．これらの手法は，各 mode において T3 の精度が良い．また，T2 と Ozaki 等の方法は，mode により精度の優劣は異なる．ただし，T2 は計算量が一番小さく $\frac{11}{12}n^3$ flops 程度である．それに対して，Ozaki 等の方法は $\frac{5}{3}n^3$ flops 程度であるから，計算量は T2 の約 2 倍である．

6. まとめ

本論では，連立一次方程式の数値解に対するコレスキー分解を用いた精度保証法を提案した．提案手法は，Rump-Ogita の方法が精度保証に失敗する問題に対して有効な精

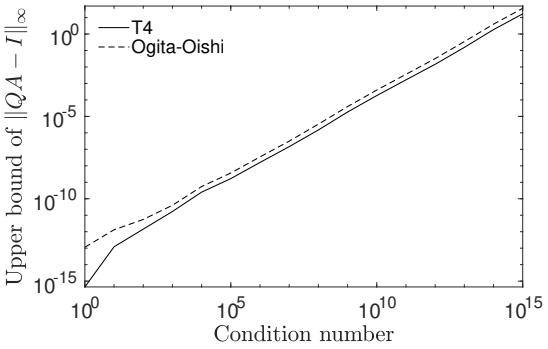

図 1 $n = 1024, \mathrm{mode} = 2.$
T4 と Ogita-Oishi の方法の比較

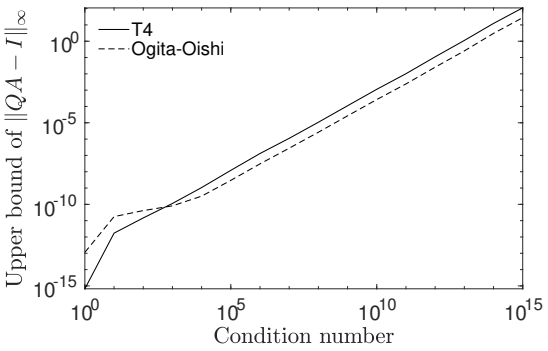

図 2 $n = 1024, \mathrm{mode} = 4.$
T4 と Ogita-Oishi の方法の比較

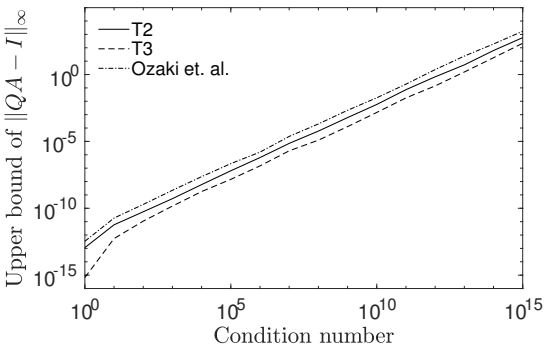

図 3 $n = 1024, \mathrm{mode} = 1.$
T2, T3, Ozaki 等の方法の比較

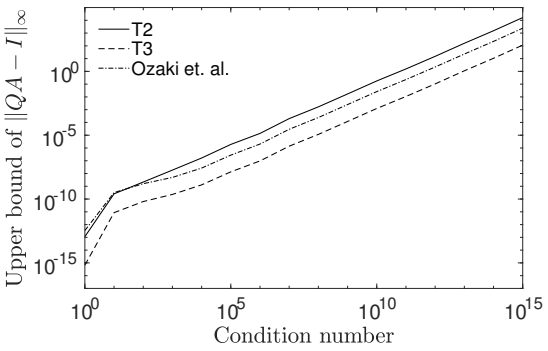

図 4 $n = 1024, \mathrm{mode} = 4.$
T2, T3, Ozaki 等の方法の比較

度保証法である．また，この手法は先行研究の近似逆行列を用いる方法と同等の精度で計算量を大幅に低減することができる．また，提案手法は T1 → T2 or T3 → T4 のように段階的に検証が可能であり，計算の効率化に対して優れている．

参考文献

[1] S.M. Rump and T. Ogita, *Super-fast validated solution of linear system*, Journal of Computation and Applied Mathematics, **199** (2007), 199–206.
[2] *IEEE Standard for Floating-Point Arithmetic*, 754–2008 (2008).
[3] S.M. Rump and C.-P. Jeannerod, *Improved backward error bounds for LU and Cholesky factorizations*, SIAM. J. Matrix Anal. & Appl., **35** (2014), 684 – 698.
[4] S.M. Rump, *Verification of positive definiteness*, BIT Numer. Math., **46** (2006), 433 – 452.
[5] N.J.Higham, *Accuracy and Stability of Numerical Algorithms, 2nd edition*, SIAM Publications, Philadelphia (2002).
[6] S. Oishi and S.M. Rump, *Fast verification of solution of matrix equations*, Numer. Math., **90** (2002), 755 – 773.
[7] A.Mimamihata, Y.Morikura, T.Ogita and S.Oishi, *A Modified Verification Method for Linear Systems by Using LU Decomposition*, The 14th Asia Simulation Conference & The 33rd JSST Annual Conference: International Conference on Simulation Technology (2014).
[8] T. Ogita and S. Oishi, *Fast verified solutions of linear systems*, IPSJ Trans, **46** (2005), 10 – 18 (in Japanese).
[9] K. Ozaki, T. Ogita and S. Oishi, *An algorithm for automatically selecting a suitable verification method for linear systems*, Numerical Algorithms, **56** (2011), 363 – 382.
[10] G. H. Golub and C. F. Van Loan, *Matrix computations, 4th edition*, The Johns Hopkins University Press (2013).