

二重指数関数型数値積分公式を用いた 行列符号関数計算法の丸め誤差解析

2023年12月8日

日本応用数理学会

「行列・固有値問題の解法とその応用」研究部会 第36回研究部会

電気通信大学大学院 情報理工学研究科

宮下朋也, 山本有作, 工藤周平

行列符号関数について

• 定義

- 複素数 z に対して、符号関数 $\text{sign}(z)$ は次のように定義される

$$\text{sign}(z) = \begin{cases} 1, & \text{Re } z > 0 \\ -1, & \text{Re } z < 0 \end{cases}$$

- これを行列に拡張したものが**行列符号関数**である
- 正方行列 A のJordan標準形を $A = XJX^{-1}$ とし、 $\text{diag}(J) = \text{diag}(\lambda_i)$ とすると、行列符号関数 $\text{sign}(A)$ は次式により定義される

$$\text{sign}(A) = X\text{sign}(J)X^{-1} = X\text{diag}(\text{sign}(\lambda_i))X^{-1}$$

• 応用例

- シルベスター方程式の求解
- 行列平方根計算

行列符号関数の数値計算方法

- 定義に基づく方法

- 対角化可能な行列の場合: 対角化の計算量大
- 対角化不可能な行列の場合: Jordan標準形の数値計算の不安定性

- Schur分解法

- Newton法

本研究で使用

- 積分表示に基づく方法



- [1]で提案
- 行列符号関数を半無限区間での積分として表示し，二重指数型数値積分公式(DE公式)で計算
- 数値計算の各標本点での計算が独立なので並列化に向く

本発表の目的

- **丸め誤差解析**
 - DE公式に基づく計算法の丸め誤差について上界を導出する
- **被積分関数の変形による丸め誤差の影響の削減**
 - 丸め誤差解析の結果よりわかった事実をもとに被積分関数を変形し，誤差を減らす
- **他の計算法との計算精度の比較**
 - Newton法，Schur分解法との比較を行う

目次

- DE公式に基づく行列符号関数の計算法
- 離散化誤差と打ち切り誤差の上界
- 数値実験による誤差評価
- 丸め誤差解析
- 数値誤差削減のための被積分関数の変形
- 他の計算法との比較

DE公式について

- 台形公式は積分区間が $(-\infty, +\infty)$ で被積分関数が急減少な解析関数の場合には非常に高精度な数値積分公式

→ 一般の積分についても、そうなるように変数変換をする

- 特に、変換後の被積分関数が二重指数関数的に減衰するように選んだものをDE公式という

変数変換

$$\begin{aligned} T_h &= \int_{-\infty}^{+\infty} f(t) dt \\ &= \int_{-\infty}^{+\infty} f(\phi(x)) \phi'(x) dx \end{aligned}$$

変数変換後に台形公式を適用

$$T_h^* = h \sum_{k=-N_-}^{N_+} f(\phi(kh)) \phi'(kh)$$

行列符号関数の積分表示形式

- 行列符号関数の積分表示形式

$$\text{sign}(A) = \frac{2}{\pi} A \int_0^{\infty} (t^2 I + A^2)^{-1} dt$$

- DE公式を適用すると次のようになる

$$\text{sign}(A) \simeq \frac{2}{\pi} A \sum_{k=-N_-}^{N_+} h(\phi^2(kh)I + A^2)^{-1} \phi'(kh)$$

- ただし,

$$\phi(t) \equiv \exp\left(\frac{\pi}{2} \sinh t\right)$$

DE公式を用いた行列符号関数計算の理論誤差上界

- 行列 $A \in \mathbb{R}^{n \times n}$ に対する理論誤差上界(離散化誤差 + 打ち切り誤差)
 - A が一般行列の場合[1]

$$\|E\|_F \leq \frac{n\sqrt{n}C(A, d)\|A\|_F(\sqrt{n}C'(A, d) + \|A^2\|_F)^{n-1}}{\pi R^{2n}} \exp\left(-\frac{2\pi dN}{\log(8dN)}\right)$$

- A が対角化可能 ($A = X\Lambda X^{-1}$) の場合 [2]

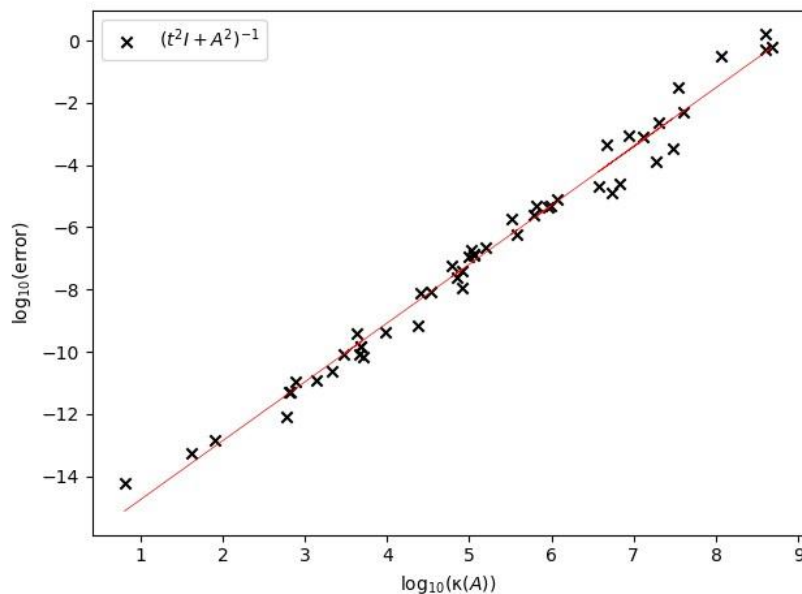
$$\|E\|_F \leq 4\kappa_2(X) \left\{ \frac{1}{\cos d} \left(\frac{\|A^2\|_F}{R^2} + \frac{\sqrt{n}}{3} \right) + \|A\|_F \right\} \exp\left(-\frac{2\pi dN_+}{\log(8dN_+)}\right)$$

[1] 中屋貴博, 田中健一郎, 二重指数関数型数値積分公式による行列符号関数公式による行列符号関数の数値計算, 日本応用数理学会論文誌, Vol. 31, No. 3(2021), 105-132.

[2] 宮下朋也, 山本有作, 二重指数関数型数値積分公式を用いた行列符号関数の計算の改良および応用, 日本応用数理学会2022年度年会予稿.

$\kappa(A)$ に対する誤差

- $\kappa_2(A) \in [10^1, 10^{10}]$ となるような $A \in \mathbb{R}^{100 \times 100}$ を50個作成
- それぞれに対して十分小さな h と十分大きな N (標本点)を用いてDE公式で $\text{sign}(A)$ を計算
- 解析解との差のフロベニウスノルムとして誤差を計算
- 条件数に対する誤差をプロット



$\kappa(A)$ に対するDE公式による誤差

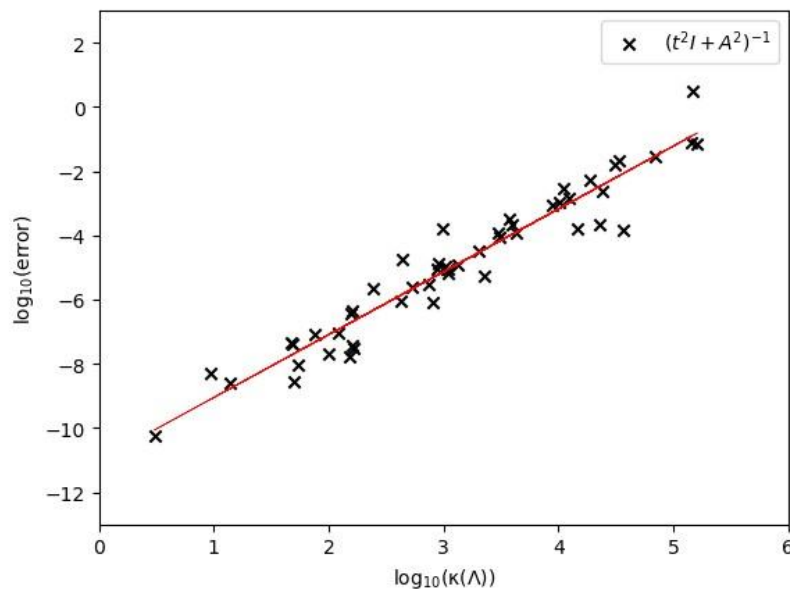
- 傾きは約1.9 $\rightarrow \text{error} \propto \kappa(A)^{1.9}$
- 条件数が 10^9 程度になるとほとんど意味のある解が得られない

$\kappa(\Lambda), \kappa(X)$ に対する誤差

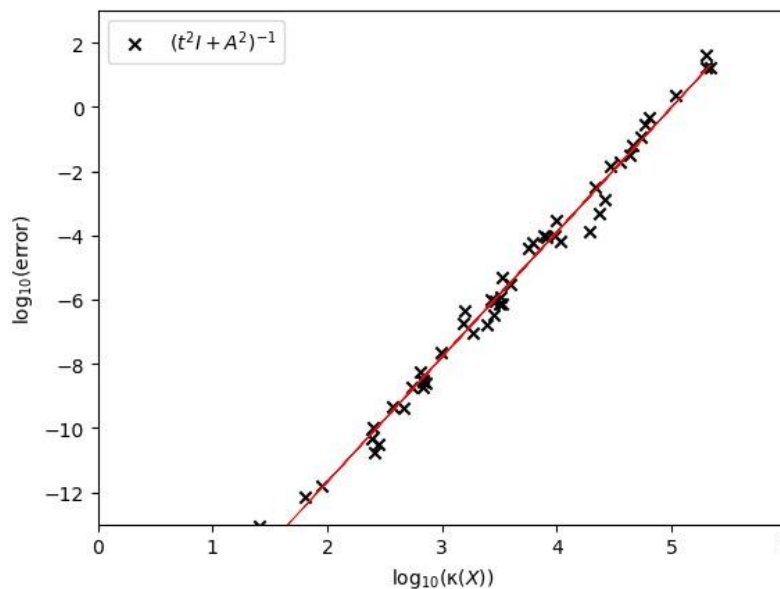
- $A = X\Lambda X^{-1} \in \mathbb{R}^{100 \times 100}$ の $\kappa_2(\Lambda), \kappa_2(X)$ が $[10^0, 10^6]$ の範囲内になるように作成し、それぞれにDE公式を適用し、収束誤差を求める (50回)
- Λ, X の条件数に対する誤差を調べる

対角化可能な場合の誤差上界

$$\|E\|_F \leq 4\kappa_2(X) \left\{ \frac{1}{\cos d} \left(\frac{\|A^2\|_F}{R^2} + \frac{\sqrt{n}}{3} \right) + \|A\|_F \right\} \exp \left(-\frac{2\pi d N_+}{\log(8dN_+)} \right)$$



$\kappa(\Lambda)$ に対するDE公式による誤差



$\kappa(X)$ に対するDE公式による誤差

右図より

$$\text{error} \propto \kappa(\Lambda)^{1.95}$$

$$\text{error} \propto \kappa(X)^{3.89}$$

DE公式での理論誤差上界より大きい

→ 丸め誤差の影響か

→ 丸め誤差解析を行う

丸め誤差解析における記号と表記

- 準備

- 丸め誤差の単位を \mathbf{u} とし, $\gamma_n = n\mathbf{u} / (1 - n\mathbf{u})$ とする
- 浮動小数点演算で計算した量を, $fl(\cdot)$ またはハット付きの記号(\hat{y})で表す
- $A = (a_{ij})$ に対し, $|a_{ij}|$ を要素とする行列を $|A|$ とする
- 行列に対し, $A \leq B$ は要素ごとに不等式が成り立つことを表す

- 注意点

- スカラーの乗算や加算の丸め誤差については考えない

丸め誤差解析

- 行列符号関数の積分表示形式(再掲)

$Y(t) \equiv (t^2 I + A^2)^{-1} A$ とすると

$$\begin{aligned} \text{sign}(A) &= \frac{2}{\pi} \int_0^\infty Y(t) dt \\ &\simeq \frac{2}{\pi} h \sum_{k=-N_-}^{N_+} Y(\phi(kh)) \phi'(kh) \end{aligned}$$

- 次の2点で生じる丸め誤差を考える
 - 各標本点での行列の計算における丸め誤差の合計
 - 和の計算における丸め誤差

1.

$$\frac{2}{\pi} h \sum_{k=N_-}^{N_+} \|Y(\phi(kh)) - \hat{Y}(\phi(kh))\|_F \phi'(kh)$$

2.

$$\frac{2}{\pi} h \sum_{k=N_-}^{N_+} \|Y(\phi(kh)) - \hat{Y}(\phi(kh))\|_F \phi'(kh)$$

各標本点での行列の計算における丸め誤差の合計

- 各項 $Y(\phi(kh))$ の計算で生じる誤差の合計を考える

$$\begin{aligned} & \frac{2}{\pi} h \sum_{k=N_-}^{N^+} \|Y(\phi(kh)) - \hat{Y}(\phi(kh))\|_F \phi'(kh) \\ & \simeq \frac{2}{\pi} \int_{-\infty}^{\infty} \|Y(\phi(x)) - \hat{Y}(\phi(x))\|_F \phi'(x) dx \\ & = \frac{2}{\pi} \int_0^{\infty} \|Y(t) - \hat{Y}(t)\|_F dt \end{aligned}$$

- $Y(t)$ の計算における丸め誤差上界を求める必要がある

$$\hat{Y}(t) = fl(\underline{(t^2 I + A^2)^{-1} A})$$

A^2 の計算における丸め誤差
連立1次方程式の求解

$B = t^2 I + A^2$ の計算における丸め誤差

- $t^2 I$ と A^2 の加算は対角要素に1個のスカラーを加えるだけなので無視する
- 丸め誤差を以下のように表記する

$$\tilde{B} = fl(t^2 I + A^2) = B + \Delta B'$$

- このとき、行列乗算の丸め誤差は以下ようになる[3]

$$|\Delta B'| \leq \gamma_n |A|^2$$

- よって、次の不等式が成り立つ

$$\|\Delta B'\|_F \leq \gamma_n \|A\|_F^2$$

LU分解の後退誤差

- 次の連立1次方程式を部分軸選択付きLU分解を用いて解くことを考える

$$\tilde{B}\mathbf{y} = \mathbf{c}$$

- \tilde{B} を浮動小数点演算によりLU分解して \hat{L}, \hat{U} が得られたとし, それらを用いて浮動小数点による前進消去, 後退代入を行って近似解 $\hat{\mathbf{y}}$ が得られたとする
- このとき, 次の式が成り立つ[3]

$$(\tilde{B} + \Delta B'')\mathbf{y} = \mathbf{c}, \quad |\Delta B''| \leq \gamma_{3n} |\hat{L}| |\hat{U}|$$

- これより,

$$\|\Delta B''\|_F \leq \gamma_{3n} \|\hat{L}\|_F \|\hat{U}\|_F$$

LU分解の後退誤差

- 部分軸選択付きのLU分解では L と U は次の式を満たす

$$\|\hat{L}\|_F \leq n$$

$$\|\hat{U}\|_F \leq n \hat{\rho}_n \|\tilde{B}\|_2$$

ただし, $\hat{\rho}_n = \max_{i,j,k} |\tilde{b}_{i,j}^k| / \max_{i,j} |\tilde{b}_{i,j}|$ とする(成長因子)

- これらを $\|\Delta B''\|_F \leq \gamma_{3n} \|\hat{L}\|_F \|\hat{U}\|_F$ に代入し, $\|\Delta B'\|_F \leq \gamma_n \|A\|_F^2$ を用いると次の評価が得られる

$$\begin{aligned} \|\Delta B''\|_F &\leq n^2 \gamma_{3n} \hat{\rho}_n \|\tilde{B}\|_2 \\ &\leq n^2 \gamma_{3n} \hat{\rho}_n (\|B\|_2 + \gamma_n \|A\|_F^2) \\ &\simeq n^2 \gamma_{3n} \hat{\rho}_n \|t^2 I + A^2\|_2 \end{aligned}$$

連立1次方程式の求解における丸め誤差

摂動による解への影響の一般論

- 解きたい連立1次方程式の解を \mathbf{y} , 係数行列に摂動 ΔB が加わった方程式の解を $\hat{\mathbf{y}}$ とすると次のようになる

$$B\mathbf{y} = \mathbf{c},$$

$$(B + \Delta B)\hat{\mathbf{y}} = \mathbf{c}$$

- $\|B^{-1}\| \|\Delta B\| \leq 1$ が成り立つとすると, 次の不等式が成り立つ[3]

$$\frac{\|\hat{\mathbf{y}} - \mathbf{y}\|}{\|\mathbf{y}\|} \leq \frac{\|B^{-1}\| \|\Delta B\|}{1 - \|B^{-1}\| \|\Delta B\|}$$

- ベクトルのノルムは何でもよく, 行列のノルムはベクトルのノルムに従属なノルムとする
- 特に, $\|B^{-1}\| \|\Delta B\| < 1/2$ なら, 次の式が成り立つ

$$\|\hat{\mathbf{y}} - \mathbf{y}\| \leq 2\|B^{-1}\| \|\Delta B\| \|\mathbf{y}\|$$

$Y(t) \equiv (t^2 I + A^2)^{-1} A$ の計算における丸め誤差

- A , $\hat{Y}(t)$ の第 j 列をそれぞれ \mathbf{a}_j , $\hat{\mathbf{y}}_j$ とすると, $\hat{\mathbf{y}}_j$ は次の方程式を満たす

$$(B + \Delta B' + \Delta B_j'') \hat{\mathbf{y}}_j = \mathbf{a}_j$$

$$\text{ただし, } \|\Delta B'\|_F \leq \gamma_n \|A\|_F^2$$

$t^2 I + A^2$ の計算における誤差

$$\|\Delta B_j''\|_F \leq n^2 \gamma_{3n} \hat{\rho}_n \|t^2 I + A^2\|_2$$

LU分解における誤差

- したがって, $\|B^{-1}\|_2 \|\Delta B' + \Delta B_j''\|_2 < 1/2$ ならば, 次が成り立つ

$$\|\hat{\mathbf{y}}_j - \mathbf{y}_j\| \leq 2 \|B^{-1}\|_2 \|\Delta B' + \Delta B_j''\|_2 \|\mathbf{y}_j\|$$

$$\leq 2 \|(t^2 I + A^2)^{-1}\|_2 (\gamma_n \|A\|_F^2 + n^2 \gamma_{3n} \hat{\rho}_n \|t^2 I + A^2\|_2) \|\mathbf{y}_j\|$$

- よって,

$$\|\hat{Y}(t) - Y(t)\|_F \lesssim 2 \left\| (t^2 I + A^2)^{-1} \right\|_2 \left(\gamma_n \|A\|_F^2 + n^2 \gamma_{3n} \hat{\rho}_n \|t^2 I + A^2\|_2 \right) \|Y(t)\|_F$$

$$= 2 \|(t^2 I + A^2)^{-1}\|_2 (\gamma_n \|A\|_F^2 + n^2 \gamma_{3n} \hat{\rho}_n \|t^2 I + A^2\|_2) \|(t^2 I + A^2)^{-1} A\|_F$$

$Y(t) \equiv (t^2 I + A^2)^{-1} A$ の計算における丸め誤差

- $A = X\Lambda X^{-1}$ と対角化できるとすると以下のようなになる

$$\|A\|_F^2 = \|X\Lambda X^{-1}\|_F^2 \leq (\|X\|_2 \| \Lambda \|_F \|X^{-1}\|_2)^2 = (\kappa_2(X))^2 \|\Lambda\|_F^2$$

同様に $\|t^2 I + A^2\|_2 \leq \kappa_2(X) \|t^2 I + \Lambda^2\|_2$

$$\|(t^2 I + A^2)^{-1}\|_2 \leq \kappa_2(X) \|(t^2 I + \Lambda^2)^{-1}\|_2$$

$$\|(t^2 I + A^2)^{-1} A\|_F \leq \kappa_2(X) \|(t^2 I + \Lambda^2)^{-1} \Lambda\|_F$$

- よって、 $Y(t)$ の計算における丸め誤差の上界が次のようになる

$$\begin{aligned} \|\hat{Y}(t) - Y(t)\|_F &\lesssim 2 \left\| (t^2 I + A^2)^{-1} \right\|_2 \left(\gamma_n \|A\|_F^2 + n^2 \gamma_{3n} \hat{\rho}_n \|t^2 I + A^2\|_2 \right) \left\| (t^2 I + A^2)^{-1} A \right\|_F \\ &\leq 2 \left\{ \gamma_n (\kappa_2(X))^4 \|\Lambda\|_F^2 + n^2 \gamma_{3n} \hat{\rho}_n (\kappa_2(X))^3 \|t^2 I + \Lambda^2\|_2 \right\} \|(t^2 I + \Lambda^2)^{-1}\|_2 \|(t^2 I + \Lambda^2)^{-1} \Lambda\|_F \end{aligned}$$

各標本点での行列の計算における丸め誤差の合計

- $Y(t)$ の計算における丸め誤差の上界が得られたので、先に述べた各標本点での行列の計算における丸め誤差の合計を評価できる

$$\begin{aligned}
 & \frac{2}{\pi} h \sum_{k=N_-}^{N^+} \|Y(\phi(kh)) - \hat{Y}(\phi(kh))\|_F \phi'(kh) \\
 & \simeq \frac{2}{\pi} \int_{-\infty}^{\infty} \|Y(\phi(x)) - \hat{Y}(\phi(x))\|_F \phi'(x) dx \\
 & = \frac{2}{\pi} \int_0^{\infty} \|Y(t) - \hat{Y}(t)\|_F dt \\
 & \leq \frac{4}{\pi} \int_0^{\infty} \left\{ \gamma_n(\kappa_2(X))^4 \|\Lambda\|_F^2 + n^2 \gamma_{3n} \hat{\rho}_n(\kappa_2(X))^3 \|t^2 I + \Lambda^2\|_2 \right\} \\
 & \quad \times \|(t^2 I + \Lambda^2)^{-1}\|_2 \|(t^2 I + \Lambda^2)^{-1} \Lambda\|_F dt
 \end{aligned}$$

- 後は積分などを評価していく

DE公式による行列符号関数計算の丸め誤差上界

1. 各標本での行列の計算における丸め誤差の合計

$$\begin{aligned} & \frac{2}{\pi} h \sum_{k=N_-}^{N_+} \|Y(\phi(kh)) - \hat{Y}(\phi(kh))\|_F \phi'(kh) \\ & \leq \left(\gamma_n(\kappa_2(X))^4 + 3n^2 \gamma_{3n} \hat{\rho}_n(\kappa_2(X))^3 \right) \left(\frac{4\sqrt{2}}{\pi} + \|\Lambda\|_F^3 \left\| \Lambda^{-1} |\operatorname{Re}(\Lambda)|^{-\frac{1}{2}} \right\|_F^2 \right) \end{aligned}$$

2. 和の計算における丸め誤差

$$\|\hat{S}_n - S_n\|_F \leq \gamma_{N-1} \kappa_2(X) \left\{ n - \frac{2}{\pi} \sum_{j=1}^n \log \left(\frac{|\operatorname{Re}(\lambda_j)|}{|\lambda_j|} \right) \right\}$$

- 1での誤差による影響が大きいことがわかる
- 誤差上界は、 $\kappa_2(X)$ について4次のオーダー、 $\kappa_2(\Lambda)$ について3次のオーダーとなる

誤差解析まとめ

- 丸め誤差解析により次の誤差上界が得られる

$$\begin{aligned} & \frac{2}{\pi} h \sum_{k=N_-}^{N_+} \|Y(\phi(kh)) - \hat{Y}(\phi(kh))\|_F \phi'(kh) \\ & \leq \left(\gamma_n(\kappa_2(X))^4 + 3n^2 \gamma_{3n} \hat{\rho}_n (\kappa_2(X))^3 \right) \left(\frac{4\sqrt{2}}{\pi} + \|\Lambda\|_F^3 \left\| \Lambda^{-1} |\operatorname{Re}(\Lambda)|^{-\frac{1}{2}} \right\|_F^2 \right) \end{aligned}$$

- 逆行列の中に A^2 がある影響で丸め誤差上界が大きくなっている
 - A^2 の誤差を考慮しないと誤差上界は $\kappa_2(X)$ については3次のオーダーとなる
- そこで、逆行列の中に A^2 が含まれないように被積分関数を変形することで丸め誤差を減らすことを考えていく

2つの被積分変換方法

- 被積分関数の逆行列の中の A^2 がなくなるように変形

1. A を逆行列の中に入れる

$$A(t^2 I + A^2)^{-1} = (t^2 A^{-1} + A)^{-1}$$

$$\text{sign}(A) \approx \frac{2}{\pi} h \sum_{k=-N_-}^{N_+} (\phi(kh)^2 A^{-1} + A)^{-1} \phi'(kh)$$

2. 被積分関数を部分分数分解

$$A(t^2 I + A^2)^{-1} = \frac{1}{2} \{(A - itI)^{-1} + (A + itI)^{-1}\}$$

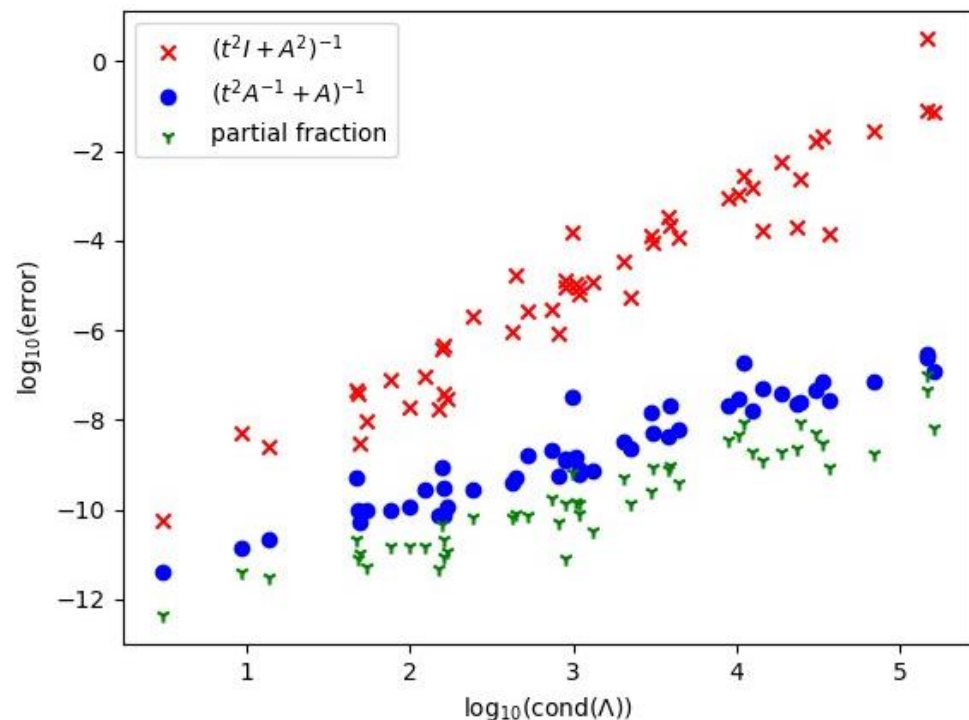
$$\text{sign}(A) \approx \frac{h}{\pi} \sum_{k=-N_-}^{N_+} \left\{ (A - i\phi(kh))^{-1} + (A + i\phi(kh))^{-1} \right\} \phi'(kh)$$

数値実験条件

- $\kappa_2(\Lambda)$ と $\kappa_2(X)$ の誤差への影響を調べる
- 条件
 - 行列サイズ: 100×100
 - 誤差の求め方: 解析解との差のフロベニウスノルム
 - DE公式適用法: 刻み幅を変えて収束するまで計算し, 最も誤差の小さい解を選択
 - テスト行列数: 50個
 - 行列タイプ: 対称行列
 - $\kappa_2(\Lambda)$ を調べるときは, $\kappa_2(X)$ などは固定の値になるようにする($\kappa_2(X)$ のときも同様)

$\kappa_2(\Lambda)$ に対する誤差

- 以下の3つの被積分関数で実験
 - オリジナル: $(t^2I + A^2)^{-1}A$
 - A を逆行列の中に入れる: $(t^2A^{-1} + A)^{-1}$
 - 部分分数分解: $(A - itI)^{-1} + (A + itI)^{-1}$

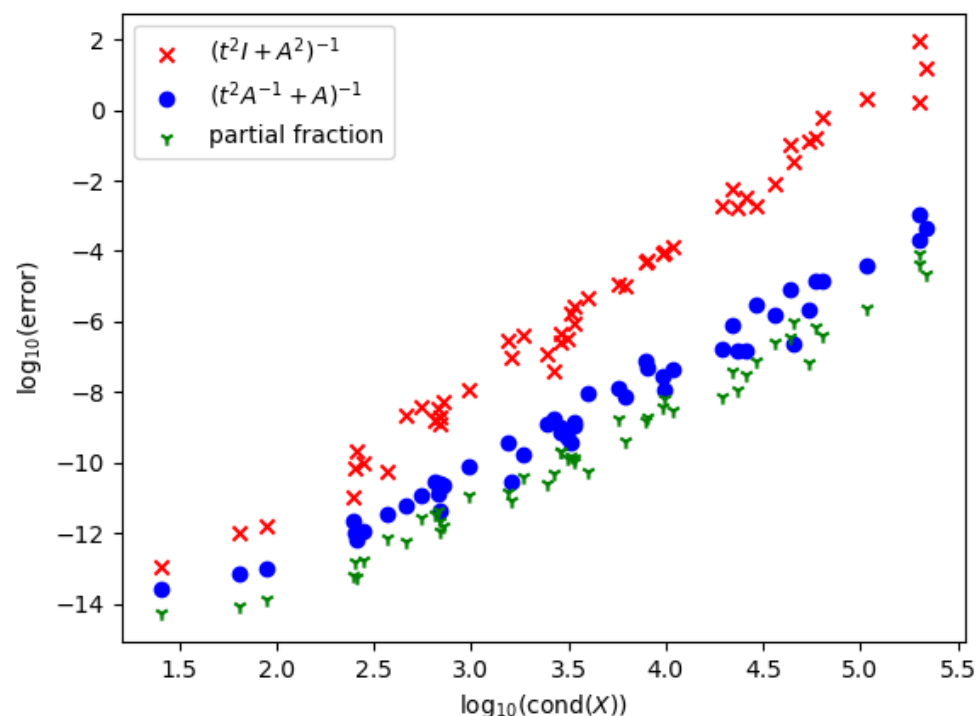


被積分関数	傾き
オリジナル	1.95701569
A を逆行列の中に入れる	1.00755635
部分分数分解	0.97182402

理論誤差上界では $O(\kappa_2(\Lambda)^3)$ であるため、それ以下になっていることが確認できる

$\kappa_2(X)$ に対する誤差

- 以下の3つの被積分関数で実験
 - オリジナル: $(t^2I + A^2)^{-1}A$
 - A を逆行列の中に入れる: $(t^2A^{-1} + A)^{-1}$
 - 部分分数分解: $(A - itI)^{-1} + (A + itI)^{-1}$

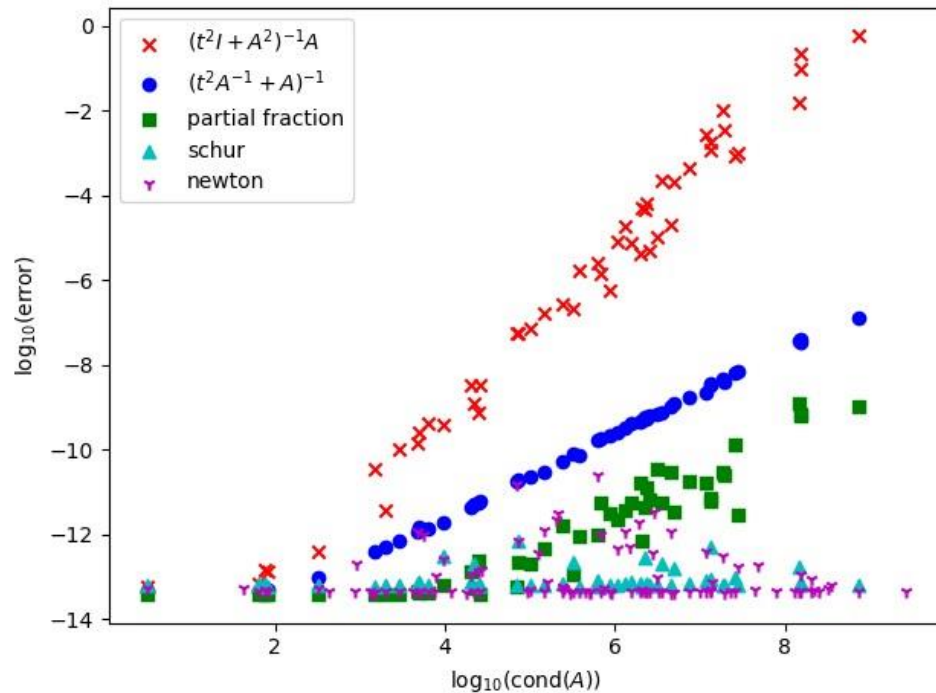


被積分関数	傾き
オリジナル	3.85292305
A を逆行列の中に入れる	2.80341657
部分分数分解	2.71465548

理論誤差上界では $O(\kappa_2(X)^4)$ であるため、それ以下になっていることが確認できる

$\kappa_2(A)$ に対する誤差

- 以下の3つの被積分関数で実験(Schur分解法, Newton法とも比較)
 - オリジナル: $(t^2I + A^2)^{-1}A$
 - A を逆行列の中に入れる: $(t^2A^{-1} + A)^{-1}$
 - 部分分数分解: $(A - itI)^{-1} + (A + itI)^{-1}$



- 理論的な検証はできないが、被積分変形後の方が $\kappa_2(A)$ による影響が減少していることが分かる
- Schur分解法, Newton法は $\kappa_2(A)$ に対しても安定

まとめ

- 丸め誤差解析により誤差上界は $\kappa_2(\Lambda)$ について4次のオーダー， $\kappa_2(X)$ について3次のオーダーになることがわかった
- 数値実験での誤差は， $\kappa_2(\Lambda)^{1.96}$ ， $\kappa_2(X)^{3.85}$ に比例していて，丸め誤差上界以下に抑えられていることがわかる
- 被積分関数の逆行列の中に A^2 が含まれないように変形することで，丸め誤差の影響を減らし，誤差を減らすことができた
- 他の計算法と比較すると，まだ誤差が大きいため，さらなる改善が必要