

チュートリアル
「精度保証付き数値計算の基礎」

数値線形代数における精度保証

荻田 武史
東京女子大学

2018/9/10

はじめに

連立一次方程式や固有値問題など線形代数に現れる諸問題について、数値計算によって効率良く解く手法（数値解法）を開発したり、それらの手法の性質を解析したりする学問を**数値線形代数**と呼ぶ。

ここでは、その中でも特に科学技術計算の基礎かつ重要である「連立一次方程式」及び「固有値問題」に対する精度保証付き数値計算法を中心に述べる。

準備

- 行列 $A = (a_{ij}), B = (b_{ij}) \in \mathbb{R}^{m \times n}$ に対し, $A \leq B$ はすべての (i, j) 要素に対して $a_{ij} \leq b_{ij}$ が成立していることを意味する。
- 行列の絶対値 $|A| = (|a_{ij}|) \in \mathbb{R}^{m \times n}$ は, 要素毎に絶対値をとった行列を表す。
- $A \geq O$ (あるいは $A > O$) は, A の要素が非負 (あるいは正) であることを意味する。
- ベクトルに対しても, 同様の記号を用いる。
- I は単位行列, e は要素がすべて1のベクトルを表す。

ベクトルノルムと行列ノルム

ベクトル $v = (v_1, v_2, \dots, v_n)^T \in \mathbb{C}^n$ の1ノルム, 2ノルム及び ∞ ノルムを, それぞれ

$$\|v\|_1 := \sum_{i=1}^n |v_i|, \quad \|v\|_2 := \sqrt{\sum_{i=1}^n |v_i|^2}, \quad \|v\|_\infty := \max_{1 \leq i \leq n} |v_i|$$

と定義する。特に, 2ノルムはユークリッドノルムと呼ばれる。

行列 $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ の固有値を $\lambda_i, 1 \leq i \leq n$ として

$$\lambda_{\max} := \max_{1 \leq i \leq n} |\lambda_i|, \quad \lambda_{\min} := \min_{1 \leq i \leq n} |\lambda_i|$$

と表す。また, λ_{\max} は A のスペクトル半径と呼ばれ, $\rho(A)$ と表すことがある。

上記のベクトルノルムから誘導される行列ノルム $\|A\| := \sup_{v \neq 0} \frac{\|Av\|}{\|v\|}$ は、それぞれ

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}| \quad (\text{最大列和ノルム})$$

$$\|A\|_2 = \sqrt{\rho(A^*A)} \quad (\text{スペクトルノルム})$$

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| \quad (\text{最大行和ノルム})$$

となる。ただし、 A^* は A の共役転置行列を意味する。このような行列ノルムは、 $\rho(A) \leq \|A\|$ を満たす。

さらに

$$\|A\|_F = \left(\sum_{i,j=1}^n |a_{ij}|^2 \right)^{1/2}$$

をフロベニウスノルムと呼ぶ。有用な行列ノルムの関係として

$$\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty} \quad (1)$$

$$\|A\|_2 \leq \|A\|_F \quad (2)$$

がある。また、 A の条件数は

$$\kappa(A) = \|A\| \cdot \|A^{-1}\|$$

のように表される。これは数値計算において問題の解きづらさの指標となる。特に A が実対称行列の場合、 $\kappa_2(A) = \|A\|_2 \cdot \|A^{-1}\|_2 = \lambda_{\max} / \lambda_{\min}$ ，すなわち、 A の絶対値最大・最小固有値の比となる。

以下では，本章で取り扱う特別な行列について説明する。

まず， $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ に対し，比較行列 $\mathcal{M}(A) = (m_{ij}) \in \mathbb{R}^{n \times n}$ を以下のように定義する。

$$m_{ij} = \begin{cases} |a_{ij}| & (i = j) \\ -|a_{ij}| & (i \neq j) \end{cases}$$

以下，すべて $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ とする。

定義 1. [狭義優対角行列] すべての $i \in \{1, 2, \dots, n\}$ に対して

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|$$

であるとき， A を狭義優対角行列と呼ぶ。

定義 2. [単調行列] A が正則かつ $A^{-1} \geq O$ であるとき、単調行列と呼ぶ。

定義 3. [M行列] A が単調行列で、 $a_{ii} > 0$ かつ $a_{ij} \leq 0$ ($i \neq j$) ならば、 M 行列と呼ぶ。

定義 4. [H行列] $M(A)$ が M 行列であるとき、 A を H 行列と呼ぶ。

補題 1. A が H 行列であることの必要十分条件は、 $M(A)v > \mathbf{0}$ を満たすベクトル $v > \mathbf{0}$ が存在することである。

補題 2. A が H 行列ならば、 $|A^{-1}| \leq M(A)^{-1}$ である。

定義 5. [正定値行列] 任意の $v \in \mathbb{R}^n$ ($v \neq \mathbf{0}$) に対し、 $v^T Av > 0$ であるとき A を正定値行列と呼ぶ。

実対称 (あるいはエルミート) 正定値行列の固有値はすべて正である。

区間行列積

数値線形代数における精度保証では行列積の区間演算が重要となるため、 $\mathbf{A} \in \mathbb{IR}^{m \times n}$ と $\mathbf{B} \in \mathbb{IR}^{n \times p}$ の積を区間包囲する、すなわち $\mathbf{A} \cdot \mathbf{B} \subset \mathbf{C}$ となる $\mathbf{C} \in \mathbb{IR}^{m \times p}$ を求める計算法について述べる。

区間演算を要素毎に用いる素朴な方式

$$C_{ij} = \sum_{k=1}^n A_{ik} B_{kj}, \quad 1 \leq i \leq m, 1 \leq j \leq p \quad (3)$$

では、プログラムの最適化の観点から計算機の性能を引き出すことが困難であるため、以下では、BLAS (Basic Linear Algebra Subprograms) などの既存の高速かつ信頼性の高い数値計算ライブラリを用いることが可能な点行列同士の積をベースとした手法を紹介する。

高速な区間行列積

以下の定理は、素朴な方式(3)よりも結果の区間幅が拡大されるが、高速な区間行列積の実現に大変有用である。

定理 1. $\mathbf{A} \in \mathbb{IR}^{m \times n}$, $\mathbf{B} \in \mathbb{IR}^{n \times p}$ に対して, $A_m := \text{mid}(\mathbf{A})$, $A_r := \text{rad}(\mathbf{A})$, $B_m := \text{mid}(\mathbf{B})$, $B_r := \text{rad}(\mathbf{B})$ とする。このとき

$$\mathbf{A} \cdot \mathbf{B} \subset \langle A_m B_m, T \rangle, \quad T := |A_m| B_r + A_r (|B_m| + B_r)$$

が成立する。

実際には、機械区間演算を考えるため、以下では $A_m, A_r \in \mathbb{IF}^{m \times n}$, $B_m, B_r \in \mathbb{IF}^{n \times p}$ である場合を考える。もし \mathbf{A} が下端・上端型で保存されている場合は、 \mathbf{A} を包含する中心・半径型の区間に変換してから計算を行う。

定理1において

$$\mathbf{A} \cdot \mathbf{B} \subset \langle A_m B_m, T \rangle = [A_m B_m - T, A_m B_m + T] \quad (4)$$

であるから

$$\bar{T} := \mathbf{fl}_{\Delta}(|A_m|B_r + A_r(|B_m| + B_r)) \quad (5)$$

とすれば, $T \leq \bar{T}$ であるため

$$\mathbf{A} \cdot \mathbf{B} \subset [\mathbf{fl}_{\nabla}(A_m B_m - \bar{T}), \mathbf{fl}_{\Delta}(A_m B_m + \bar{T})] \quad (6)$$

となる。この方法では, 行列積が

$$\mathbf{fl}_{\Delta}(|A_m|B_r), \quad \mathbf{fl}_{\Delta}(A_r(|B_m| + B_r)), \quad \mathbf{fl}_{\nabla}(A_m B_m), \quad \mathbf{fl}_{\Delta}(A_m B_m)$$

と4回必要であることがわかる。

連立一次方程式

A を n 次行列， b を n 次元ベクトルとする。 A, b の要素は，実数または複素数とする。このとき，連立一次方程式

$$Ax = b \tag{7}$$

の解の存在範囲を特定することを考える。ここでは，古典的な区間演算を用いた方法から現代的な高速精度保証法までを取り扱う。

係数行列 A が正則 ($\det(A) \neq 0$) であれば A^{-1} が存在し，(7) の解の一意性が保証され，その厳密解は $x = A^{-1}b$ となる。したがって， A の正則性の保証が重要である。

逆に，精度保証付き数値計算では， A が特異 ($\det(A) = 0$) であることを保証するのは一般に困難である。

なぜなら、 A が特異であったとしても、微小な摂動 ΔA によって、 $A + \Delta A$ は正則になるからである。もし A が特異に近い場合、 A は**悪条件**と呼ばれる。そのような場合は、数値計算における丸め誤差によって、 A の正則性の保証が失敗することがある¹。

数値計算によって(7)を解いたときに得られるのは近似解 \hat{x} である。

近似解 \hat{x} の誤差を評価するときは、**ノルム評価**

$$\|x^* - \hat{x}\| \leq \varepsilon \quad (8)$$

や成分毎評価

$$|x_i^* - \hat{x}_i| \leq d_i, \quad i = 1, 2, \dots, n \quad (9)$$

が用いられる。

¹この場合でも、 A が正則である可能性があることに注意しなければならない。 A が悪条件であっても、演算の精度を高めることによって、正則性の保証が可能となる場合がある。

ガウスの消去法とLU分解

ガウスの消去法は，基本変形の繰り返しによって行列 $A \in \mathbb{R}^{n \times n}$ を三角行列に変換する方法である。

具体的には， $A^{(1)} := A$, $b^{(1)} := b$ として， $k = 1, 2, \dots, n - 1$ に対して

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik}a_{kj}^{(k)}, \quad i = k + 1, \dots, n, \quad j = k + 1, \dots, n$$
$$b_i^{(k+1)} = b_i^{(k)} - m_{ik}b_k^{(k)}, \quad i = k + 1, \dots, n$$

を実行する。ただし， $m_{ik} = a_{ik}^{(k)} / a_{kk}^{(k)}$ である。

これによって，途中で $a_{kk}^{(k)}$ が0になって破綻しない限り，最終的に A が上三角行列 $U := A^{(n)}$ に変換される。

このとき、 $c := b^{(n)}$ とすると、三角方程式 $Ux = c$ は後退代入

$$x_i = \left(c_i - \sum_{j=i+1}^n u_{ij}x_j \right) / u_{ii}, \quad i = n, n-1, \dots, 1$$

によって容易に解くことが可能である。

ただし、 $a_{kk}^{(k)}$ が 0 でなかったとしても、絶対値が相対的に小さくなる場合は数値計算誤差の意味で不利になる。これを避けるため、通常、行の交換や列の交換による軸交換を行う²。

²実用上は、行の交換のみで十分である。これを部分軸交換 (partial pivoting) と呼ぶ。詳細は、例えば文献 [4] を参照されたい。

ガウスの消去法を用いて、行列 A を $A = LU$ のように下三角行列 $L = (l_{ij})$ と上三角行列 $U = (u_{ij})$ の積に分解することができる (LU分解)。

A の LU 分解は無数に存在するが、 L か U の対角成分を 1 に固定することによって一意に定まる。 L の対角成分について $l_{kk} = 1$ と固定する LU 分解 (これは Doolittle 法と呼ばれる) は、 $k = 1, 2, \dots, n - 1$ に対して

$$u_{kj} = a_{kj} - \sum_{i=1}^{k-1} l_{ki} u_{ij}, \quad j = k, \dots, n \quad (10)$$

$$l_{ik} = \left(a_{ik} - \sum_{j=1}^{k-1} l_{ij} u_{jk} \right) / u_{kk}, \quad i = k + 1, \dots, m \quad (11)$$

となる³。Doolittle 法は、前述の軸交換なしのガウスの消去法と数学的に等しく、計算コストは $2n^3/3$ flops である。

³ U の対角成分について $u_{kk} = 1$ と固定する LU 分解は Crout 法と呼ばれる。

以下は，LU分解に対する後退誤差についての定理である。

定理 2. $A \in \mathbb{F}^{m \times n}$ の浮動小数点演算によるLU分解が破綻せずに終了したとする。このとき，得られたLU分解因子 $\hat{L} = (\hat{l}_{ij}) \in \mathbb{F}^{m \times n}$ ， $\hat{U} = (\hat{u}_{ij}) \in \mathbb{F}^{n \times n}$ は，計算順序に関わらず

$$\hat{L}\hat{U} = A + \Delta A, \quad |\Delta A| \leq nu|\hat{L}||\hat{U}| \quad (12)$$

を満たす。ただし，アンダーフローが発生する場合を除く。

u: 単位丸め誤差 (IEEE 754 binary64 では $u = 2^{-53} \approx 10^{-16}$)

また，三角方程式については，以下の後退誤差評価が成り立つ。

定理 3. 下三角行列 $L \in \mathbb{F}^{n \times n}$, $b \in \mathbb{F}^n$ に対し，浮動小数点演算を用いて以下の前進代入によって $Lx = b$ を解いたときの近似解を $\hat{x} \in \mathbb{F}^n$ とする。

$$x_i = \left(b_i - \sum_{j=1}^{i-1} l_{ij} x_j \right) / l_{ii}, \quad i = 1, 2, \dots, n$$

このとき，計算順序に関わらず

$$(L + \Delta L)\hat{x} = b, \quad |\Delta L| \leq nu|L| \quad (13)$$

が成り立つ。ただし，アンダーフローが発生する場合を除く。

u : 単位丸め誤差 (IEEE 754 binary64 では $u = 2^{-53} \approx 10^{-16}$)

コレスキー分解

A が実対称正定値行列のとき、LU分解ではなくコレスキー分解 $A = R^T R$ を用いると計算コストがLU分解の半分 ($n^3/3$ flops) で済む。ここで、 R は上三角行列である。

コレスキー分解は、 $j = 1, 2, \dots, n$ に対して

$$r_{ij} = (a_{ij} - \sum_{k=1}^{i-1} r_{ki}r_{kj})/r_{ii}, \quad i = 1, 2, \dots, j-1$$

$$r_{jj} = (a_{jj} - \sum_{k=1}^{j-1} r_{kj}^2)^{1/2}$$

を実行する。ただし、 A が実対称正定値行列であっても、浮動小数点演算における丸め誤差の累積によって、 r_{jj} の計算に負の平方根が現れ、途中で計算が破綻する場合がある。

以下は，コレスキー分解に対する後退誤差についての定理である。

定理 4. $A = A^T \in \mathbb{F}^{n \times n}$ に対して浮動小数点演算によるコレスキー分解を適用し，破綻せずに終了したとする。このとき，得られたコレスキー分解因子 $\hat{R} = (\hat{r}_{ij}) \in \mathbb{F}^{n \times n}$ は，計算順序に関わらず

$$\hat{R}^T \hat{R} = A + \Delta A, \quad |\Delta A| \leq (n + 1)u |\hat{R}^T| |\hat{R}| \quad (14)$$

を満たす。ただし，アンダーフローが発生する場合を除く。

u : 単位丸め誤差 (IEEE 754 binary64 では $u = 2^{-53} \approx 10^{-16}$)

反復改良法

反復改良法は、近似解の精度を向上させる手法の一つである。

$Ax = b$ の近似解 \hat{x} に対する残差を $r := b - A\hat{x}$ とすると、厳密解に対する \hat{x} の誤差は、 $y^* := A^{-1}r$ で与えられる。

反復改良法は以下のような手順で実行できる。まず、 $\hat{x}^{(0)} := \hat{x}$ として、近似解の精度が十分に良くなるまで $k = 0, 1, 2, \dots$ として以下を反復する。

1. 残差 $r := b - A\hat{x}^{(k)}$ を高精度に計算する（近似値を \hat{r} とする）。
2. $Ay = \hat{r}$ を解く（ A のLU分解等の結果を用いて近似解 \hat{y} を求める）。
3. 近似解を $\hat{x}^{(k+1)} \leftarrow \hat{x}^{(k)} + \hat{y}$ と更新する。

ここで、残差 r の計算については、通常の数値精度で計算し、最終的に通常の数値精度で保持すれば良い。たとえば、アルゴリズムDot2を用いると、これを実現できる。

区間ガウスの消去法

区間ガウスの消去法は，通常の高ウスの消去法におけるすべての演算を区間演算に置き換える手法である。これによって，連立一次方程式 $Ax = b$ の真の解の存在範囲を得ることが理論上可能である。

しかしながら，区間ガウスの消去法は，特別な場合を除いて，係数行列 A の次数 n がある程度以上の大きさになると， A の条件数（問題の解きづらさ）とは無関係に適用できなくなるという致命的な欠点があることが知られている。

これは，区間演算による区間幅の増大に起因する問題であり，係数行列が区間行列の場合に限らず，点行列の場合でも，丸め誤差によって途中の計算結果が区間になるため，浮動小数点演算を用いる限り避けることができない。

たとえば, A を n 次の乱数行列, $b = A \cdot e$ として, 区間ガウスの消去法によって得られる解の包含の最大区間半径を図1に示す。

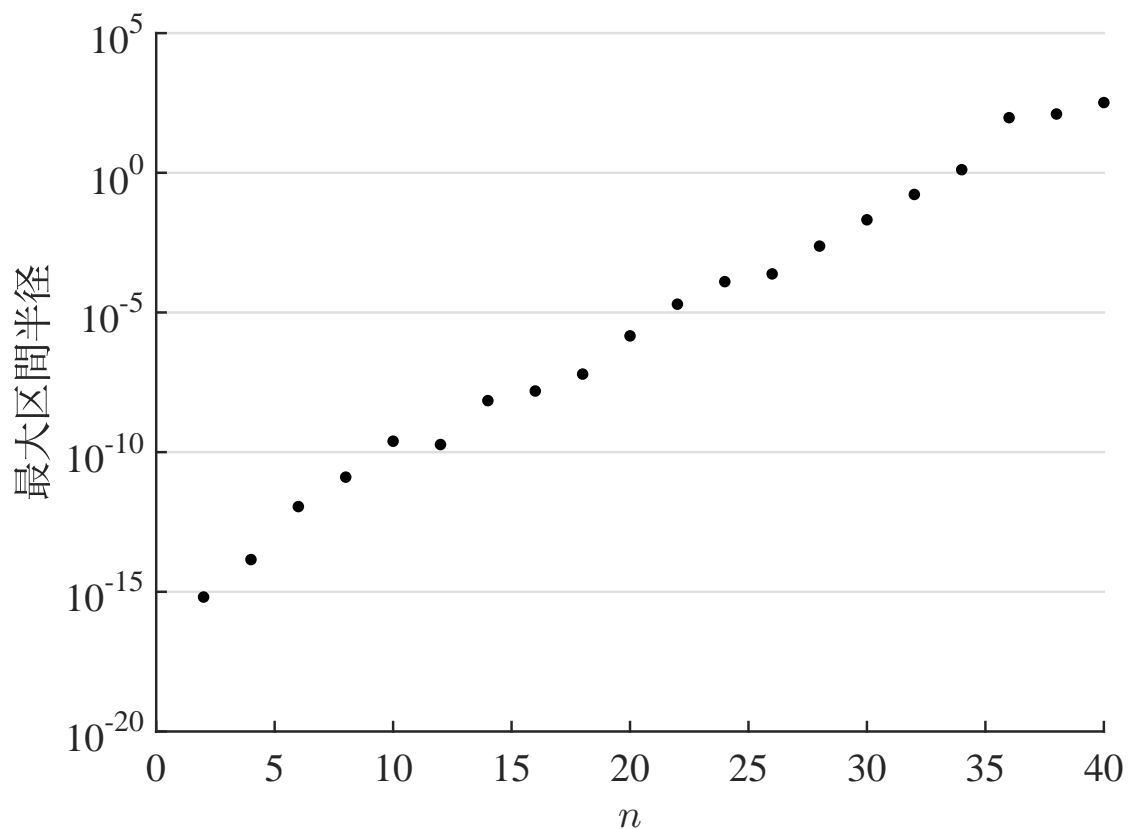


図 1: 区間ガウスの消去法による区間増大の例 (n を 2 から 40 まで 2 刻みで動かす, 各 n について 20 サンプル実行した時の平均)。

密行列に対する精度保証法

係数行列 A が密行列（要素のほとんどが非ゼロ）の場合に有効な精度保証法について説明する。

ここで紹介する方法は、いずれも行列計算単位で区間演算を実行することができるため、BLASやLAPACKなどの高速で信頼性の高い数値計算ライブラリを利用可能である。

また、区間ガウスの消去法と違い、次数 n が数千以上のような比較的大規模な行列にも適用可能である。

ノルム評価

連立一次方程式 $Ax = b$ の近似解 \hat{x} に対して、 A^{-1} の近似 $R \in \mathbb{R}^{n \times n}$ が $\|I - RA\|_\infty < 1$ を満たすとき

$$\|A^{-1}b - \hat{x}\|_\infty \leq \frac{\|R(b - A\hat{x})\|_\infty}{1 - \|I - RA\|_\infty} \quad (15)$$

が成り立つ。ここで、 A の正則性を示す上で、 $\|I - RA\|_\infty$ の評価が重要となる。

A^{-1} の近似 R の計算については、LU分解の結果を利用することができる。

以下のアルゴリズムを用いると、 $\|I - RA\|_\infty \leq \alpha$ を満たす α を求めることができる。

アルゴリズム 1. $A \in \mathbb{F}^{n \times n}$ の近似逆行列 $R \in \mathbb{F}^{n \times n}$ を用いた $\|I - RA\|_\infty$ の上限 α の計算

$$\begin{aligned} \underline{C} &= \text{fl}_{\nabla}(R \cdot A - I); & \% RA - I \text{ の下限} \\ \overline{C} &= \text{fl}_{\Delta}(R \cdot A - I); & \% RA - I \text{ の上限} \\ \overline{G} &= \max(|\underline{C}|, |\overline{C}|); & \% |I - RA| \text{ の上限} \\ \alpha &= \max_{1 \leq i \leq n} \text{fl}_{\Delta} \left(\sum_{j=1}^n \overline{G}_{ij} \right); & \% \|I - RA\|_\infty \text{ の上限} \end{aligned}$$

この方式を用いると、近似逆行列 R の計算も含めて、 $\|I - RA\|_\infty$ の評価に $6n^3$ flops の計算コストが必要となる。

式(15)の分子 $\|R(b - A\hat{x})\|_\infty$ の評価については

$$\|R(b - A\hat{x})\|_\infty \leq \|R\|_\infty \cdot \|b - A\hat{x}\|_\infty$$

を用いれば、アルゴリズム1と同様に上限を簡単に計算できるが、これは過大評価になりやすい。

区間演算等を用いて $b - A\hat{x} \in \mathbf{r}$ を満たす区間ベクトル \mathbf{r} を求め、 $\|\text{mag}(R \cdot \mathbf{r})\|_\infty$ の計算をしたほうが良い結果を得られる。

$\text{mag}(\mathbf{a})$ は、区間 \mathbf{a} における絶対値の上限

$$\text{mag}(\mathbf{a}) := \sup_{a \in \mathbf{a}} |a| \geq 0$$

を表す。

いずれの場合も、 $O(n^2)$ flops の計算コストで済む。

成分毎評価

式(15)によるノルム評価は使い勝手が良いが、解ベクトルの成分間で絶対値の大きさにばらつきがある場合に、絶対値の小さい成分に対して誤差評価が過大になってしまう場合がある。

そのようなときは、以下の成分毎評価法 [15] が有用である。

定理 5. $A, R \in \mathbb{R}^{n \times n}$, $b, \hat{x} \in \mathbb{R}^n$, $G := I - RA$ とする。このとき、 $\|G\|_\infty < 1$ ならば、 A は正則であり

$$|A^{-1}b - \hat{x}| \leq |R(b - A\hat{x})| + \frac{\|R(b - A\hat{x})\|_\infty}{1 - \|G\|_\infty} |G|e \quad (16)$$

が成り立つ。

H行列を用いる方法

行列 A が H 行列のときに、効率の良い精度保証法がある。これは、疎行列の場合にも適用可能である。

定理 6. $A \in \mathbb{R}^{n \times n}$ に対し、 $v \in \mathbb{R}^n$ が $v > \mathbf{0}$ かつ $u := \mathcal{M}(A)v > \mathbf{0}$ を満たしているとする。また、 $\mathcal{M}(A) = D + E$ を、 $\mathcal{M}(A)$ の対角部分 D と非対角部分 E の分離とする。このとき、 $r \in \mathbb{R}^n$ に対して

$$|A^{-1}r| \leq D^{-1}|r| + \alpha v, \quad \alpha := \max_{1 \leq i \leq n} \frac{(-ED^{-1}|r|)_i}{u_i} \quad (17)$$

が成り立つ。

定理6において、 $v > \mathbf{0}$ かつ $\mathcal{M}(A)v > \mathbf{0}$ を満たすような $v \in \mathbb{R}^n$ を見つける必要がある。

v の候補として、非負行列 $S := -D^{-1}E$ のペロンベクトルが考えられる。具体的には、 $v^{(0)} := D^{-1}e$ として、べき乗法 $v^{(k+1)} := Sv^{(k)}$ による反復を数回繰り返せば良い。

S のペロンベクトルを $q > \mathbf{0}$ とすると、ペロン根 $\rho(S)$ に対して $Sq = \rho(S)q$ であるため⁴、もし $\rho(S) < 1$ であれば

$$\mathcal{M}(A)q = D(I - D^{-1}E)q = (1 - \rho(S))Dq > \mathbf{0}$$

となる。

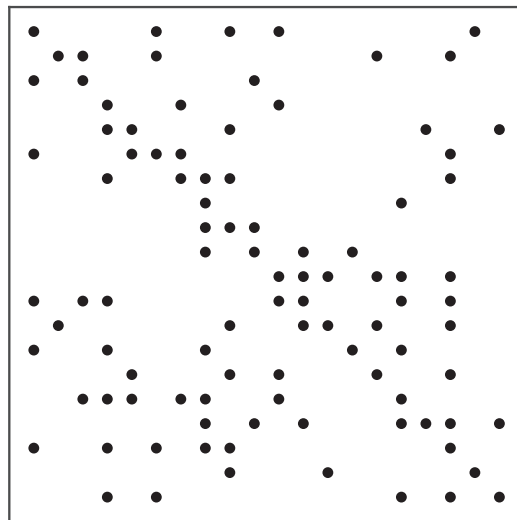
したがって、 S のペロンベクトルの近似を $v > \mathbf{0}$ とすれば、 v が $\mathcal{M}(A)v > \mathbf{0}$ を満たすことを期待できる。

⁴ペロン=フロベニウスの定理については、たとえば文献[2]を参照されたい。

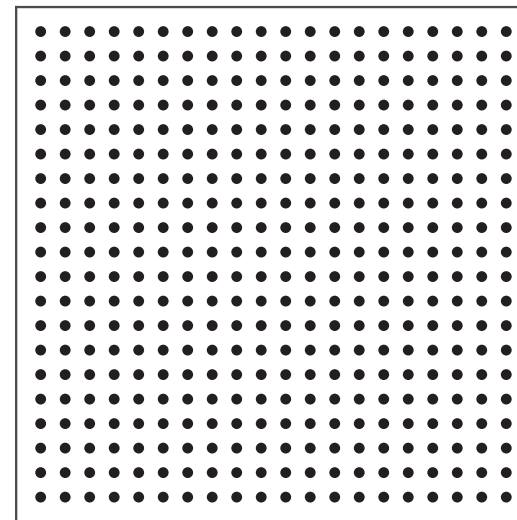
疎行列に対する精度保証法

大規模疎行列に対する精度保証は，特殊な場合を除いて非常に難しいことが知られており，区間解析における難問の1つとなっている。

たとえば， A が疎行列であっても，その逆行列は一般に密行列となる。



nz = 97



nz = 400

図 2: 疎性の崩れ ($n = 20$ の乱数行列 A と A^{-1} の非ゼロ要素パターン)

現状では，係数行列 A が以下の場合に，疎行列に対して有効な方法が知られている。

- 単調行列系（狭義優対角行列， M 行列， H 行列を含む）
- 実対称（あるいはエルミート）正定値行列

係数行列 A あるいはその比較行列 $M(A)$ が単調行列の場合については，前節の H 行列に対する精度保証法が有効である。

ここでは実対称正定値行列の場合に限定して，疎行列の場合でも有効な方法を説明する。

A が正則であるとき、 A の最小特異値 $\sigma_{\min}(A) = \sqrt{\lambda_{\min}(A^T A)}$ の下限 $\sigma > 0$ を計算することができれば

$$\|A^{-1}\|_2 = \frac{1}{\sigma_{\min}(A)} \leq \frac{1}{\sigma}$$

であるので

$$\|A^{-1}b - \hat{x}\|_{\infty} \leq \|A^{-1}b - \hat{x}\|_2 \leq \frac{\|b - A\hat{x}\|_2}{\sigma} \quad (18)$$

が成り立つ。

実対称正定値行列 A に対し，アンダーフローが生じる場合を除いて

$$\widehat{R}^T \widehat{R} = A + \Delta A, \quad \|\Delta A\|_2 \leq \sum_{i=1}^n \gamma_{i+1} a_{ii}, \quad \gamma_k = \frac{ku}{1 - ku} \quad (19)$$

が成り立つことが知られている⁵。これを用いると以下の定理を得る。

定理 7. 実対称行列 $A = (a_{ij}) \in \mathbb{F}^{n \times n}$, $a_{ii} \geq 0$ の浮動小数点演算によるコレスキー分解が破綻せずに完了したとする。このとき

$$\lambda_{\min}(A) \geq - \sum_{i=1}^n \gamma_{i+1} a_{ii} \quad (20)$$

が成り立つ。ただし，アンダーフローが発生する場合を除く。

⁵疎行列の場合， $(i+1)u$ の i は， \widehat{R} の非ゼロ要素パターンに関連した値に置き換えることができる。

$C := \text{fl}_{\nabla}(A - 2\sigma I)$, $\sigma := \sum_{i=1}^n \gamma_{i+1} a_{ii}$ とする。 C の浮動小数点演算によるコレスキー分解が破綻せずに完了した場合は、定理7より、アンダーフローが発生した場合を除いて

$$\begin{aligned} \lambda_{\min}(A) - 2\sigma &= \lambda_{\min}(A - 2\sigma I) \geq \lambda_{\min}(C) \geq -\sum_{i=1}^n \gamma_{i+1} c_{ii} \\ &\geq -\sum_{i=1}^n \gamma_{i+1} (a_{ii} - 2\sigma) \geq -\sum_{i=1}^n \gamma_{i+1} a_{ii} = -\sigma \end{aligned}$$

となり、 $\lambda_{\min}(A) \geq \sigma > 0$ を得る。すなわち、 A が正定値であることが証明され、さらに $\|A^{-1}\|_2 = 1/\lambda_{\min}(A) \leq 1/\sigma$ となる。

もし、コレスキー分解が途中で破綻した場合は、 A が正定値かどうかの判定はできない⁶。

⁶浮動小数点演算を用いた場合は、 A が正定値であったとしても、丸め誤差の累積が原因でコレスキー分解が破綻する場合があるからである。

式(18)による方法は誤差ノルム $\|A^{-1}b - \hat{x}\|_\infty$ を過大評価しやすいので、それを抑制するために、下記のような誤差評価法が有効である。

定理 8. $A \in \mathbb{R}^{n \times n}$, $b, \hat{x}, \hat{y} \in \mathbb{R}^n$ とする。 $\|A^{-1}\|_2 \leq \tau$ ならば

$$\|A^{-1}b - \hat{x}\| \leq \|\hat{y}\| + \tau \|b - A(\hat{x} + \hat{y})\|_2 e \quad (21)$$

が成り立つ。ただし、 $e = (1, 1, \dots, 1)^T \in \mathbb{R}^n$ である。

定理8において、通常、連立一次方程式 $Ay = r$ ($r := b - A\hat{x}$) の近似解を \hat{y} とする。反復改良法にあるように、残差ベクトル r を高精度に計算した場合に、 \hat{y} が $Ax = b$ の近似解 \hat{x} の修正項として有効となる。そのためには高精度な内積計算が必要である。

区間連立一次方程式

係数行列や右辺ベクトルが区間である場合，すなわち， $\mathbf{A} \in \mathbb{IR}^{n \times n}$ ， $\mathbf{b} \in \mathbb{IR}^n$ について，その解集合は

$$\Sigma(\mathbf{A}, \mathbf{b}) := \{x \in \mathbb{R}^n \mid Ax = b, A \in \mathbf{A}, b \in \mathbf{b}\}$$

と定義される。すべての $A \in \mathbf{A}$ が正則であれば

$$\Sigma(\mathbf{A}, \mathbf{b}) = \{A^{-1}b \in \mathbb{R}^n \mid A \in \mathbf{A}, b \in \mathbf{b}\}$$

である。

この $\Sigma(\mathbf{A}, \mathbf{b})$ を求めるのは一般に困難であるので，代わりにこれを包含する区間ベクトル $\mathbf{z} \supset \Sigma(\mathbf{A}, \mathbf{b})$ を求める。

たとえば,

$$\mathbf{A} = \begin{bmatrix} [1, 2] & [-1, 0] \\ [-1, 2] & [2, 3] \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} [-1, 1] \\ [-2, 3] \end{bmatrix}$$

のとき, 解集合 $\Sigma(\mathbf{A}, \mathbf{b})$ とその包含は図3のようになる。

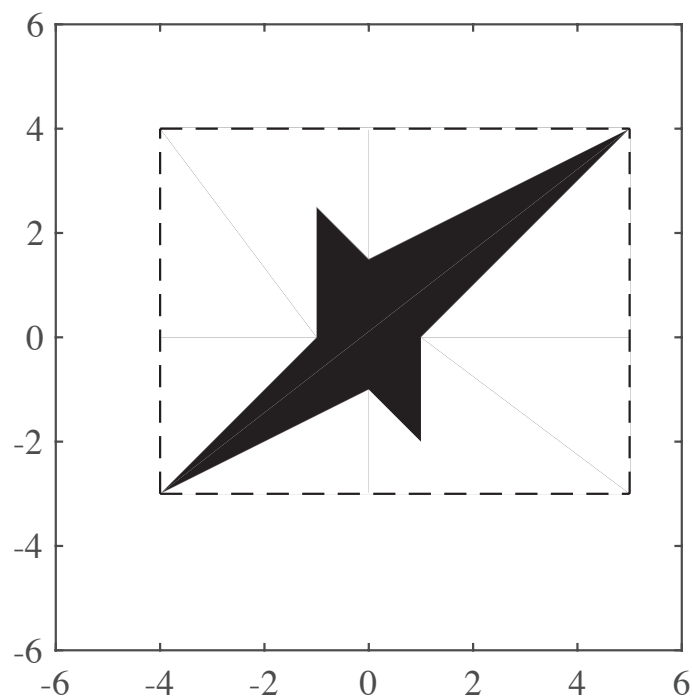


図 3: 区間連立一次方程式の解集合 (星形) とその包含 (破線)

$\mathbf{A} \in \mathbb{R}^{n \times n}$ の区間幅が狭い場合については、点行列のときとほぼ同様に考えることができる。

たとえば、 $A_m := \text{mid}(\mathbf{A})$, $b_m := \text{mid}(\mathbf{b})$ として、連立一次方程式 $A_m x = b_m$ の近似解 \hat{x} に対して、 A_m^{-1} の近似 $R \in \mathbb{R}^{n \times n}$ が $\|\text{mag}(I - R \cdot \mathbf{A})\|_\infty < 1$ を満たすとき、 \mathbf{A} に含まれるすべての行列は正則で

$$\Sigma(\mathbf{A}, \mathbf{b}) \subset \hat{x} + [-y, y], \quad y := \frac{\|\text{mag}(R(\mathbf{b} - \mathbf{A} \cdot \hat{x}))\|_\infty}{1 - \|\text{mag}(I - R \cdot \mathbf{A})\|_\infty} e \quad (22)$$

が成り立つ。ただし、 $e = (1, 1, \dots, 1)^T \in \mathbb{R}^n$ である。

$\text{mag}(\mathbf{a})$ は、区間 \mathbf{a} における絶対値の上限

$$\text{mag}(\mathbf{a}) := \sup_{a \in \mathbf{a}} |a| \geq 0$$

を表す。

行列固有値問題

固有値問題は，連立一次方程式と同様に科学技術計算の基礎であり，その解（固有値及び固有ベクトル）の精度保証付き数値計算は重要である。

行列固有値問題は，標準固有値問題

$$Ax = \lambda x, \quad A \in \mathbb{C}^{n \times n} \quad (23)$$

あるいは一般化固有値問題

$$Ax = \lambda Bx, \quad A, B \in \mathbb{C}^{n \times n} \quad (24)$$

を満たす $\lambda \in \mathbb{C}$, $x \in \mathbb{C}^n$ ($x \neq \mathbf{0}$) を求める問題であり， (λ, x) の組を固有対と呼ぶ。

実際には，重複を含めて n 個の固有値 λ_i , $i = 1, 2, \dots, n$ が存在する。このとき， λ_i に対応する固有ベクトルを $x^{(i)}$, $i = 1, 2, \dots, n$ とする。

以後、 A が実対称行列である場合は

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

と仮定する。また、特に $\lambda_i(C)$ のように表記した場合は、行列 C の固有値を意味することとする。

固有値問題の数値解法

固有値問題に対して様々な数値解法が提案されているが、密行列の場合はハウスホルダー変換を用いる方法が現在の主流である。

これによって、実対称（エルミート）行列の場合は三重対角行列，非対称行列の場合はヘッセンベルク行列に変形することができる。

変形後，対称系の場合はQR法，逆反復法を用いた二分法，分割統治法，MRRR法等によって，非対称系の場合はQR法によって効率的に固有値・固有ベクトルを求める。

いずれにしても，数値計算によって(23)あるいは(24)を解いたときに得られるのは近似固有値 $\hat{\lambda}_i$ 及び近似固有ベクトル $\hat{x}^{(i)}$ である。

$\hat{\lambda}_i$ の誤差を評価するときは

$$|\lambda_i - \hat{\lambda}_i| \leq \varepsilon \quad (25)$$

を満たす ε を求める。

また，近似固有ベクトル $\hat{x}^{(i)}$ の誤差を評価するときは，たとえば

$$\|x^{(i)} - \hat{x}^{(i)}\|_2 \leq \alpha \quad (26)$$

を満たす α や

$$|\sin \angle(x^{(i)}, \hat{x}^{(i)})| \leq \beta \quad (27)$$

を満たす β を求める。ただし，2つの n 次元実ベクトル u, v に対して， $\angle(u, v)$ は u と v のなす角を意味する。

定義 6. [固有値の重複度] $A \in \mathbb{C}^{n \times n}$ のすべての異なる固有値を $\lambda_i \in \mathbb{C}$, $1 \leq i \leq r$ とする。

1. A の固有多項式は

$$\phi(\lambda) = (-1)^n \prod_{i=1}^r (\lambda - \lambda_i)^{m_i}$$

と表せる。このとき、 m_i を固有値 λ_i に対する代数的重複度と呼ぶ。

2. 固有値 λ_i に対する固有空間

$$V_{\lambda_i} = \{x \in \mathbb{C}^n : Ax = \lambda_i x\}$$

の次元 $\dim V_{\lambda_i}$ を固有値 λ_i に対する幾何的重複度と呼ぶ。

固有値の存在範囲を特定する上で有用となるいくつかの定理を紹介する。

定理 9. [ゲルシュゴリンの包含定理] $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ について、 A のすべての固有値は

$$\Lambda = \bigcup_{1 \leq i \leq n} U_i$$

に含まれる。ただし

$$U_i = \left\{ z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \right\}$$

である。特に、単連結領域 C が m 個の U_i からなる場合、重複度も含めて m 個の固有値が C 内に存在する。

定理 9 に現れる円板領域 U_i をゲルシュゴリン円板と呼ぶ。 A が強い優対角性を持つ場合、 A の対角成分が A の固有値の良い近似となる。

例: 3次行列

$$A = \begin{bmatrix} 3 & 0 & -2 \\ 1 & -2 & 0 \\ 1 & 1 & 6 \end{bmatrix}$$

の真の固有値は, $\lambda_1 = -1.8965 \dots$, $\lambda_2 = 2.7411 \dots$, $\lambda_3 = 6.1554 \dots$ である。ゲルシュゴリンの定理を用いると, 図4のように, λ_1 は U_2 に, λ_2 と λ_3 は $U_1 \cup U_3$ に, それぞれ包含されていることがわかる。

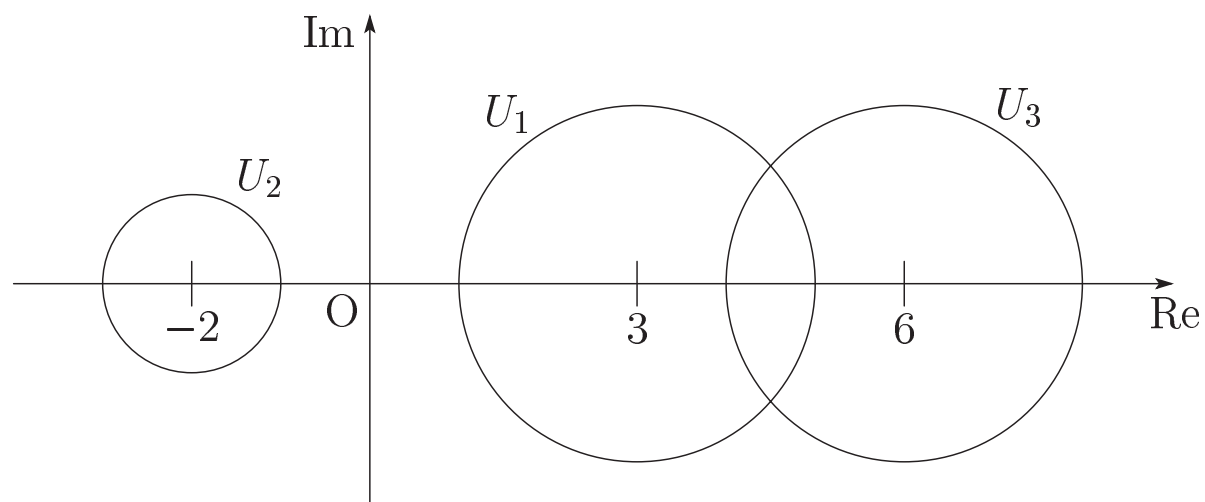


図 4: ゲルシュゴリン円板による固有値の包含

以下の定理を用いると，実対称行列 A に対して， μ の近傍に A の固有値が存在することが証明できる⁷。

定理 10. A を n 次実対称行列とする。任意の $\mu \in \mathbb{R}$ 及び任意の n 次元実ベクトル $y \neq \mathbf{0}$ に対して

$$\min_{1 \leq i \leq n} |\lambda_i - \mu| \leq \frac{\|Ay - \mu y\|_2}{\|y\|_2}$$

が成り立つ。

⁷ただし，それが何番目に小さい固有値なのかはわからない。

以下の定理は，行列に摂動が加わった場合に有用である。

定理 11. [摂動定理] A, B を n 次実対称行列とする。このとき

$$|\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|_2, \quad i = 1, 2, \dots, n$$

が成り立つ。

密行列に対する精度保証法

密行列の場合は、すべての固有値に対する効率的な精度保証法が知られている。

A を n 次実対称行列, $\hat{\lambda}_i, i = 1, 2, \dots, n$ を A の固有値の近似とする。ただし

$$\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \dots \leq \hat{\lambda}_n$$

とする。

また, $\hat{\lambda}_i$ に対応する近似固有ベクトル $\hat{x}^{(i)}, i = 1, 2, \dots, n$ を並べた行列を

$$\hat{X} := \left[\hat{x}^{(1)}, \hat{x}^{(2)}, \dots, \hat{x}^{(n)} \right] \in \mathbb{R}^{n \times n}$$

とする。

対称行列の場合

定理 12. A を n 次実対称行列, $\hat{D} = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_n)$ とする。このとき, 任意の正則行列 $\hat{X} \in \mathbb{R}^{n \times n}$ に対して

$$|\lambda_i - \hat{\lambda}_i| \leq \frac{\|A\hat{X} - \hat{X}\hat{D}\|_2}{\sqrt{\lambda_{\min}(\hat{X}^T \hat{X})}}, \quad i = 1, 2, \dots, n$$

が成り立つ。

系 1. 定理 12において, $\|I - \hat{X}^T \hat{X}\|_2 < 1$ であれば

$$|\lambda_i - \hat{\lambda}_i| \leq \frac{\|A\hat{X} - \hat{X}\hat{D}\|_2}{(1 - \|I - \hat{X}^T \hat{X}\|_2)^{1/2}}, \quad i = 1, 2, \dots, n$$

が成り立つ。

重複固有値（代数的重複度が2以上）が存在しても，上記の方法は適用可能である。

また, $R := A\hat{X} - \hat{X}\hat{D}$ や $G := I - \hat{X}^T \hat{X}$ の2ノルムを精度保証付きで求めるのは, 1ノルムや ∞ ノルムあるいはフロベニウスノルムのそれと比べて計算量が多くなるので, これを避けるために式(1)や式(2)のようなノルムの関係を利用して良い。

また，固有ベクトルの精度保証については，以下の定理が有用である。

定理 13. A を n 次実対称行列とする。このとき， $(\hat{\lambda}_i, \hat{x}^{(i)}) \in \mathbb{R} \times \mathbb{R}^n$ ， $\|\hat{x}^{(i)}\|_2 = 1$ ， $i = 1, 2, \dots, n$ に対して

$$|\hat{\lambda}_i - \hat{\lambda}_j| \geq d > 0 \quad (i \neq j) \quad (28)$$

$$\|A\hat{x}^{(i)} - \hat{\lambda}_i\hat{x}^{(i)}\|_2 \leq \varepsilon$$

が成り立っているとす。このとき， $d > 2\varepsilon$ ならば

$$|\lambda_i - \hat{\lambda}_i| \leq \varepsilon$$

$$\|x^{(i)} - \hat{x}^{(i)}\|_2 \leq \frac{\varepsilon}{d - \varepsilon}$$

が成り立つ。ただし， $x^{(i)}$ は A の固有ベクトル ($\|x^{(i)}\|_2 = 1$) を表す。

また，定理10を用いると，任意の $\mu \in \mathbb{R}$, $y \in \mathbb{R}^n$ ($\|y\|_2 = 1$) に対して

$$|\lambda_k - \mu| \leq \|Ay - \mu y\|_2$$

を満たすような A の固有値 λ_k が存在することがわかる。

このとき， $\text{gap}(\mu)$ を

$$\text{gap}(\mu) := \min_{\lambda_i \neq \lambda_k} |\lambda_i - \mu| \quad (29)$$

のように定める。これは， λ_k と隣り合う固有値との距離に対応する。

定理 14. n 次実対称行列 A の固有値 λ_k に対応する固有ベクトルを $x^{(k)}$ とする。このとき、任意の $\mu \in \mathbb{R}$, $y \in \mathbb{R}^n$ に対して

$$|\sin \angle(x^{(k)}, y)| \leq \frac{\|Ay - \mu y\|_2}{\text{gap}(\mu)}$$
$$|\lambda_k - \mu| \leq \frac{\|Ay - \mu y\|_2^2}{\text{gap}(\mu)}$$

が成り立つ。

重複固有値やそれに近い固有値が存在する場合

$$|\hat{\lambda}_i - \hat{\lambda}_j| \geq d > 0 \quad (i \neq j)$$

の条件を満たさなくなるため、上記の定理 13 及び定理 14 は適用できない。

非対称行列の場合

今、 $A \in \mathbb{C}^{n \times n}$ は対角化可能 (A の相異なる固有値 $\lambda_i \in \mathbb{C}$ について、それぞれ代数的重複度と幾何的重複度が等しい) であるとする。

このとき、 λ_i に対応する固有ベクトル $x^{(i)} \in \mathbb{C}^n$, $i = 1, 2, \dots, n$ を並べた行列

$$X := \begin{bmatrix} x^{(1)} & x^{(2)} & \cdots & x^{(n)} \end{bmatrix} \in \mathbb{C}^{n \times n}$$

によって、 A は

$$X^{-1}AX = D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

のように対角化される。

数値計算によって得られるのは X の近似 \hat{X} であるが、 $G := \hat{X}^{-1}A\hat{X}$ は強い優対角性を持つことが期待でき、そのような場合、 G の対角成分が A の固有値の良い近似となる。

そこで、 G に対してゲルシュゴリンの定理（定理9）を適用することを考える。

ただし、数値計算では行列積や逆行列を厳密に計算することは困難なので、代わりに区間演算を用いて G の包含を求めることにする。

アルゴリズム 2. $A \in \mathbb{C}^{n \times n}$ のすべての固有値に対する精度保証アルゴリズム

1. A のすべての近似固有ベクトル $\hat{X} \in \mathbb{C}^{n \times n}$ を求める。
2. 行列積 $A \cdot \hat{X}$ の包含 $C \in \mathbb{IC}^{n \times n}$ を求める。
3. 区間連立一次方程式 $\hat{X}G = C$ の解集合の包含 $G \supset \hat{X}^{-1} \cdot C$ を求める。
4. $G \in \mathbb{IC}^{n \times n}$ に対してゲルシュゴリンの定理を適用する。

このアルゴリズムが成功裏に停止し、ゲルシュゴリン円板に共通部分があれば、 $i = 1, 2, \dots, n$ について

$$|\lambda_i - c_i| \leq r_i$$

を満たす A の固有値 λ_i が各円板内に一つずつ存在することがわかる。

非線形方程式を利用した精度保証法

固有値問題は，非線形方程式の解を求める問題に帰着できる。

たとえば， $A \in \mathbb{C}^{n \times n}$ ， $0 \neq \alpha \in \mathbb{C}$ ， $1 \leq k \leq n$ について， $f : \mathbb{C}^{n+1} \rightarrow \mathbb{C}^{n+1}$ を

$$f(z) = f \left(\begin{bmatrix} x \\ \lambda \end{bmatrix} \right) = \begin{bmatrix} Ax - \lambda x \\ x_k - \alpha \end{bmatrix}$$

のようにして $f(z) = \mathbf{0}$ の解を求めると， (λ, x) は $x_k = \alpha$ に正規化された行列 A の固有対となる ($\|x\|_2 = 1$ のように正規化しても良い)。

したがって，非線形方程式の解の精度保証ができれば，固有値問題に対する精度保証も可能であることがわかる。

以下は，非線形方程式に対する精度保証法であるクラフチック法を一般化固有値問題(24)に適用した場合に得られる定理[10]である。

定理 15. $A, B \in \mathbb{C}^{n \times n}$, $R \in \mathbb{C}^{(n+1) \times (n+1)}$, $\hat{x} \in \mathbb{C}^n$, $\hat{\lambda}, \alpha \in \mathbb{C}$ ($\alpha \neq 0$) とする。 $\mathbf{y} \in \mathbb{IC}^n$, $\mathbf{d} \in \mathbb{IC}$ に対し, $\mathbf{w} := [\mathbf{y}^T, \mathbf{d}]^T \in \mathbb{IC}^{n+1}$ として

$$g(\mathbf{w}) := z + (I_{n+1} - R \cdot S(\mathbf{y}))\mathbf{w}$$

と定義する。ただし, ある固定された k ($1 \leq k \leq n$) について

$$S(\mathbf{y}) := \begin{bmatrix} A - \hat{\lambda}B & -B(\hat{x} + \mathbf{y}) \\ e_k^T & 0 \end{bmatrix} \in \mathbb{IC}^{(n+1) \times (n+1)}$$

$$z := -R \begin{bmatrix} A\hat{x} - \hat{\lambda}\hat{x} \\ \hat{x}_k - \alpha \end{bmatrix} \in \mathbb{C}^{n+1}$$

である⁸。また, $\text{int}(\mathbf{w})$ を \mathbf{w} の内部とする。このとき, $g(\mathbf{w}) \subset \text{int}(\mathbf{w})$ ならば, 一般化固有値問題 $Ax = \lambda Bx$ に対して以下が成り立つ。

⁸実際には, R は $S(\mathbf{y})$ の近似逆行列, $(\hat{\lambda}, \hat{x})$ は $Ax = \lambda Bx$ の近似固有対とする。

1. $\mathbf{v} := \hat{\mathbf{x}} + \mathbf{y}$ とすると, \mathbf{v} 内に $x_k = \alpha$ に正規化された唯一の固有ベクトル \mathbf{x} が存在する。
2. $\mathbf{c} := \hat{\lambda} + \mathbf{d}$ とすると, \mathbf{c} 内に唯一の固有値 λ が存在する。

上記の定理で示すことができるのは, 数値計算によって得られた近似固有対 $(\hat{\lambda}, \hat{\mathbf{x}})$ の近傍に真の固有対 (λ, \mathbf{x}) が唯一存在する, ということである。

逆に, λ が重複固有値の場合やそれに近い場合は, この精度保証法は適用できない。そのような場合でも適用可能な精度保証法については, 文献[8]を参照されたい。

また, 幾何的重複度が2以上の場合は, 特別な場合を除き, 精度保証付き数値計算では固有ベクトルを狭い区間で包含することができないことが文献[13]で証明されている。これは, 実対称行列の場合でも同様である。

大規模疎行列の場合

行列 A が大規模疎行列の場合，すべての固有値及び固有ベクトルを求めるのは計算量及び計算容量の両面で非現実的である。

そこで，たとえば A が実対称行列であれば「小さい方から何番目の固有値がどの範囲にあるか」など，いくつかの固有値に対する精度保証法を考えるのが現実的である。

そのような精度保証法の開発は，実は非常に困難なことが知られているが，たとえば文献 [14] では，実対称行列に対して LDL^T 分解（実対称行列 A を， $A = LDL^T$ のように正則な下三角行列 L と対角行列 D の積に分解すること）を用いた効率的な方法が提案されている。

この原理を簡単に説明しよう。

定理 16. [シルベスターの慣性則] A を実対称行列, S を正則な実行列とする。 A とその合同変換 $S^T A S$ について, その正・負・零の固有値の個数は等しい (ただし, 重複固有値も重複度分だけ個数に含める)。

実対称行列 A に対して, その対角成分を $\alpha \in \mathbb{R}$ だけ引いた行列 $A - \alpha I$ を考える (この操作をシフトと呼ぶ)。

このとき, $A - \alpha I$ の固有値は, A の固有値から α だけ負の方向にずらしたものになるため, $A - \alpha I$ の固有値の符号を調べれば, 原理的には A の固有値について「 α より大きいもの」, 「 α より小さいもの」, 「 α と等しいもの」の個数がわかることになる。

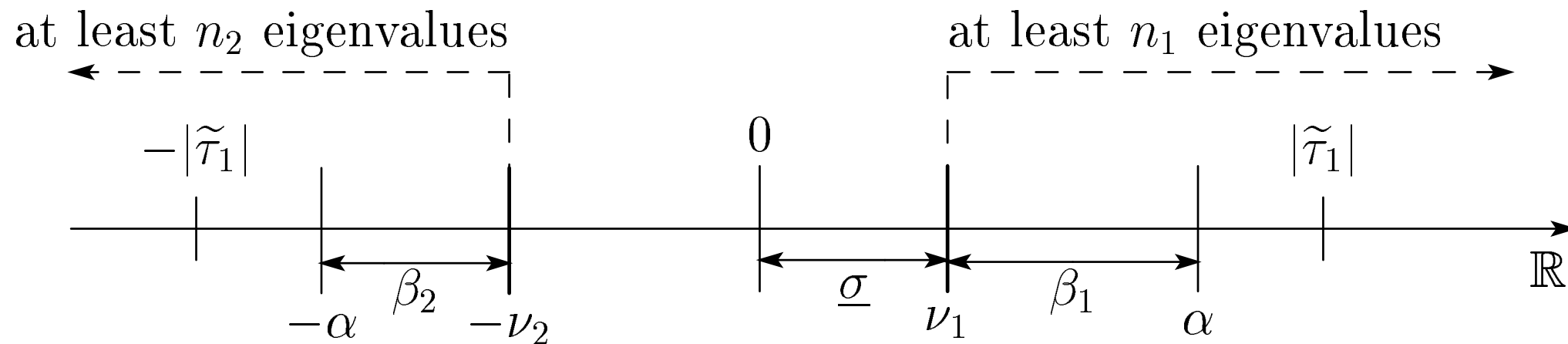
次に、 $A - \alpha I$ のLDL^T分解が成功したとすると

$$A - \alpha I = LDL^T$$

となるが、 $L^{-1}(A - \alpha I)L^{-T}$ は $A - \alpha I$ の合同変換であるため、シルベスタの慣性則から $A - \alpha I$ と D の固有値について、その正・負・零の個数は等しい。

ここで、 D は対角行列であるから、 D の固有値及びその符号がすぐに分かるため、 $A - \alpha I$ の固有値についても、その符号がわかることになる。

実際の数値計算では、 LDL^T 分解において丸め誤差が混入するため、 $A - \alpha I$ の固有値の正負の個数を厳密にカウントするためには、その丸め誤差も考慮しなければならない（固有値が零であることの判定はできない）。



したがって、 A の固有値から十分に離れたところに α をとる必要がある。

また，定理10を用いると，何番目の固有値かはわからないが，近似固有値の近くに真の固有値が存在することは容易に示すことができる。

A を n 次実対称行列とする。任意の $\mu \in \mathbb{R}$ 及び任意の n 次元実ベクトル $y \neq \mathbf{0}$ に対して

$$\min_{1 \leq i \leq n} |\lambda_i - \mu| \leq \frac{\|Ay - \mu y\|_2}{\|y\|_2}$$

が成り立つ。

References

- [1] 大石 進一：精度保証付き数値計算, コロナ社, 1999.
- [2] 山本 哲朗：数値解析入門[増訂版], サイエンス社, 2003.
- [3] J. R. Bunch, L. Kaufman: Some stable methods for calculating inertia and solving symmetric linear systems, *Math. Comp.*, **31** (1977), 163–179.
- [4] G. H. Golub, C. F. Van Loan: *Matrix Computations*, 4th ed., The Johns Hopkins University Press, Baltimore, 2013.
- [5] T. Ogita: Accurate matrix factorization: Inverse LU and inverse QR factorizations, *SIAM J. Matrix Anal. Appl.*, **31** (2010), 2477–2497.

- [6] S. Oishi, S. M. Rump: Fast verification of solutions of matrix equations, *Numer. Math.*, **90** (2002), 755–773.
- [7] B. N. Parlett: *The Symmetric Eigenvalue Problem*, *Classics in Applied Mathematics*, **20**, SIAM Publications, Philadelphia, 1997.
- [8] S. M. Rump: Computational error bounds for multiple or nearly multiple eigenvalues, *Linear Alg. Appl.*, **324** (2001), 209–226.
- [9] S. M. Rump: Inversion of extremely ill-conditioned matrices in floating-point, *Japan J. Indust. Appl. Math.*, **26** (2009), 249–277.
- [10] S. M. Rump: Verification methods for dense and sparse systems of equations, *Topics in Validated Computations – Studies in Computational Mathematics* (J. Herzberger ed.), Elsevier, Amsterdam, 1994, pp.63–136.

- [11] S. M. Rump, C.-P. Jeannerod: Improved backward error bounds for LU and Cholesky factorizations, *SIAM J. Matrix Anal. Appl.*, **35** (2014), 684–698.
- [12] S. M. Rump, T. Ogita: Super-fast validated solution of linear systems, *J. Comp. Appl. Math.*, **199** (2007), 199–206.
- [13] S. M. Rump, J.-P. M. Zemke: On eigenvector bounds, *BIT*, **43** (2003), 823–837.
- [14] N. Yamamoto: A simple method for error bounds of eigenvalues of symmetric matrices, *Linear Alg. Appl.*, **324** (2001), 227–234.
- [15] T. Yamamoto: Error bounds for approximate solutions of systems of equations, *Japan J. Appl. Math.*, **1** (1984), 157–171.