

連立1次方程式における数値解の誤差評価†

布 広 永 示** 平 野 菅 保***

浮動小数点演算を用いて連立1次方程式を解いて得た数値解の誤差評価を考える。電子計算機を用いて数値計算を行う場合、連立1次方程式の係数、定数項は、数値を有限桁の数値に丸めたことによる誤差を含む。このため、計算途中で起こる丸め誤差が入るのを防ぐために高精度演算を用いて数値解を求めたとしても、数値解は、係数、定数項が含む誤差に起因する誤差を含む。ここで、係数、定数項が含む誤差はそれぞれすべて独立しているとは限らず、同じ原因で発生し、恒等的に等しい誤差を含んでいる場合がある。したがって、独立な誤差を用いて、係数、定数項が含む誤差を表すことができる。本論文では、数値解の誤差を評価する方法として、係数、定数項が含む誤差を独立な誤差に分け、数値解の各要素ごとに誤差を評価する新しい誤差評価方法を提案した。そして、幾つかの応用例に本論文で提案した数値解の誤差評価方法を適用することによって、この誤差評価方法の有効性を示した。

1. はじめに

連立1次方程式の数値解を求める場合の数値解の誤差評価に関しては、多くの研究がなされている。現在多く用いられている数値解の誤差評価は、係数行列の条件数を用いて行われている。この方法は、数値解の誤差を解ベクトルのノルムを用いて評価しているため、数値解の各要素ごとの誤差評価が必要な場合には適切ではない。数値解の誤差評価を解ベクトルの各要素ごとに行う方法としては、区間演算による方法などがある。この方法は、数値解の各要素ごとの誤差評価を行うことはできるが、数値解が実際に含んでいる誤差と比較して、誤差を過大評価する場合がある。その大きな原因として、数値を入力する時に、有限桁の数値に丸めたことにより係数あるいは定数項が含む誤差が計算途中でお互いに消失し合う場合があるのに、区間演算ではそれが考慮されていないなどがある。そこで、数値解の各要素ごとに誤差を評価することができただけでなく、誤差の消失も考慮した誤差評価方法を提案し、数値解の誤差評価を行った。

2. 係数、定数項が含む誤差と数値解の誤差

n 元連立1次方程式

$$\sum_{j=1}^n a_{ij} \cdot x_j = b_i \quad (i=1, \dots, n) \quad (1)$$

ただし、 a_{Tij} : 誤差を含まない真の係数

† The Error Evaluation on the Numerical Solution of Linear Equations by EIJI NUNOHIRO (Software Works, Hitachi Ltd.) and SUGAYASU HIRANO (Department of Mathematical Engineering, The College of Industrial Technology, Nihon University).

** (株)日立製作所ソフトウェア工場

*** 日本大学生産工学部数理工学科

b_{Ti} : 誤差を含まない真の定数項

x_j : 変数

を解くことを考える。

電子計算機を用いて数値計算を行う場合、数値は有限桁で扱われる。したがって、(1)ではなく、数値を有限桁の数値に丸めたことに起因する誤差を含んでいる係数、定数項を持つ n 元連立1次方程式

$$\sum_{j=1}^n a_{ij} \cdot x_j = b_i \quad (i=1, \dots, n) \quad (2)$$

を解くことになる。(2)が(1)を代表している n 元連立1次方程式であるとする、係数 a_{ij} 、定数項 b_i は、

$$a_{ij} = a_{Tij} + \Delta a_{Tij} \quad (i, j=1, \dots, n)$$

$$b_i = b_{Ti} + \Delta b_{Ti} \quad (i=1, \dots, n)$$

と表される。ここで、 Δa_{Tij} 、 Δb_{Ti} は、係数 a_{ij} 、定数項 b_i が含む誤差であり、

$$|\Delta a_{Tij}| \leq \Delta a_{\max ij} \quad (i, j=1, \dots, n) \quad (3)$$

$$|\Delta b_{Ti}| \leq \Delta b_{\max i} \quad (i=1, \dots, n) \quad (4)$$

ただし、 $\Delta a_{\max ij} \geq 0$ 、 $\Delta b_{\max i} \geq 0$

$\Delta a_{\max ij}$ 、 $\Delta b_{\max i}$: 定数

を満足している未知の値である。すなわち、

$$|\Delta a_{ij}| \leq \Delta a_{\max ij} \quad (i, j=1, \dots, n) \quad (5)$$

$$|\Delta b_i| \leq \Delta b_{\max i} \quad (i=1, \dots, n) \quad (6)$$

を満足する任意の誤差 Δa_{ij} 、 Δb_i を含む係数、定数項を持つ n 元連立1次方程式

$$\sum_{j=1}^n (a_{Tij} + \Delta a_{ij}) \cdot x_j = (b_{Ti} + \Delta b_i) \quad (i=1, \dots, n) \quad (7)$$

を満足する解を $\bar{x}_j$ とすると、解 $\bar{x}_j$ よりも(1)の解に近接している解を(2)より求めることは一般にはできない。したがって、解 $\bar{x}_j$ を(2)を満足する数値解

とする。

3. 係数、定数項が含む誤差に起因する数値解の誤差

本章では、(5)、(6)を満足する任意の誤差 Δa_{ij} , Δb_i を含む係数、定数項を持つ(7)の数値解の誤差を求める。

まず、(1)において、変数 x_j を

$$x_j = x_{Tj} \quad (8)$$

とする。ここで、 x_{Tj} は(1)の解である。そして、(8)を(1)に代入して、 n 元連立1次方程式

$$\sum_{j=1}^n a_{Tij} \cdot x_{Tj} = b_{Ti} \quad (i=1, \dots, n) \quad (9)$$

を作る。次に、(7)において、変数 x_j を

$$x_j = x_{Tj} + w_j \quad (10)$$

ただし、 w_j : 変数

とする。そして、(10)を(7)に代入して、 n 元連立1次方程式

$$\sum_{j=1}^n (a_{Tij} + \Delta a_{ij}) \cdot (x_{Tj} + w_j) = (b_{Ti} + \Delta b_i) \quad (i=1, \dots, n) \quad (11)$$

を作る。こうして、(9)と(11)より、 n 元連立1次方程式

$$\begin{aligned} \sum_{j=1}^n (a_{Tij} + \Delta a_{ij}) \cdot w_j \\ = - \sum_{j=1}^n \Delta a_{ij} \cdot x_{Tj} + \Delta b_i \quad (i=1, \dots, n) \end{aligned} \quad (12)$$

を得る。

4. 数値解の最大誤差

構造解析の計算、回路網の計算あるいは偏微分方程式を連立差分方程式に置き換えて解く計算など、実際の問題を連立1次方程式に置き換えて解く場合、連立1次方程式の係数あるいは定数項の各要素は規則性を持って係数あるいは定数項を構成している場合が多い。この時、係数、定数項の中に同じ誤差が線形性を持って存在するなら、それらの係数あるいは定数項が含む誤差は、互いに従属関係を持っている。したがって、係数、定数項が含む誤差を独立な誤差ごとに分けることができる。

ここで、独立な関係にある誤差が m 個あるとし、独立な誤差を、

$$\varepsilon_k \quad (k=1, \dots, m) \quad (13)$$

ただし、 $-1.0 \leq \varepsilon_k \leq 1.0$

と表す。そして、係数、定数項が含む誤差を独立な誤

差 ε_k ごとにまとめて、それぞれ独立に数値解に対する影響を調べ、数値解の最大誤差、すなわち(1)の解 x_{Tj} と(2)の数値解 $\bar{x}_j$ との差の最大値を求める。

まず、係数、定数項が含む誤差を独立な誤差ごとにまとめ、

$$\Delta a_{ij} = \sum_{k=1}^m \Delta a_{\max ijk} \cdot \varepsilon_k \quad (i, j=1, \dots, n) \quad (14)$$

$$\Delta b_i = \sum_{k=1}^m \Delta b_{\max ik} \cdot \varepsilon_k \quad (i=1, \dots, n) \quad (15)$$

ただし、 $\Delta a_{\max ijk} \geq 0$, $\Delta b_{\max ik} \geq 0$

$\Delta a_{\max ijk}$, $\Delta b_{\max ik}$: 定数

と表す。次に(14)の両辺に解 x_{Tj} を乗じて、

$$\Delta a_{ij} \cdot x_{Tj} = x_{Tj} \cdot \sum_{k=1}^m \Delta a_{\max ijk} \cdot \varepsilon_k \quad (i, j=1, \dots, n) \quad (16)$$

を作り、(15)、(16)を用いて、(12)の右辺を

$$\begin{aligned} - \sum_{j=1}^n \Delta a_{ij} \cdot x_{Tj} + \Delta b_i \quad (i=1, \dots, n) \\ = - \sum_{j=1}^n x_{Tj} \cdot \left(\sum_{k=1}^m \Delta a_{\max ijk} \cdot \varepsilon_k \right) \\ + \sum_{k=1}^m \Delta b_{\max ik} \cdot \varepsilon_k \quad (i=1, \dots, n) \\ = \sum_{k=1}^m \left(- \sum_{j=1}^n \Delta a_{\max ijk} \cdot x_{Tj} + \Delta b_{\max ik} \right) \cdot \varepsilon_k \quad (i=1, \dots, n) \\ = \sum_{k=1}^m f_{\max ik} \cdot \varepsilon_k \quad (i=1, \dots, n) \end{aligned} \quad (17)$$

と整理する。ここで、

$$f_{\max ik} = - \sum_{j=1}^n \Delta a_{\max ijk} \cdot x_{Tj} + \Delta b_{\max ik} \quad (i=1, \dots, n; k=1, \dots, m) \quad (18)$$

である。そして(12)の右辺を(17)で置き換えて、(12)を

$$\sum_{j=1}^n (a_{Tij} + \Delta a_{ij}) \cdot w_j = \sum_{k=1}^m f_{\max ik} \cdot \varepsilon_k \quad (i=1, \dots, n) \quad (19)$$

とする。こうして、係数、定数項が含む誤差で、互いに独立な関係にある誤差それぞれの数値解に対する影響を求める。すなわち、

$$w_j = \sum_{k=1}^m w_{jk} \cdot \varepsilon_k \quad (j=1, \dots, n) \quad (20)$$

ただし、 w_{jk} : 変数

とすると(19)は、

$$\sum_{j=1}^n (a_{Tij} + \Delta a_{ij}) \cdot w_{jk} \cdot \varepsilon_k = f_{\max ik} \cdot \varepsilon_k \quad (21)$$

$$(i=1, \dots, n; k=1, \dots, m)$$

となり、係数、定数項が含む誤差で互いに独立な関係にある誤差それぞれの数値解に対する影響は、

$$w_{jk} \cdot \varepsilon_k \quad (j=1, \dots, n; k=1, \dots, m) \quad (22)$$

となる。そして、数値解の最大誤差は、

$$w_{\max j} = \sum_{k=1}^m |w_{jk}| \quad (j=1, \dots, n) \quad (23)$$

である。

5. 解 x_{Tj} と数値解 $\bar{x}_j$

実際に $f_{\max ik}$ (18) の計算をする場合、数値解 $\bar{x}_j$ は、

$$\begin{aligned} & - \sum_{j=1}^n \Delta a_{ij} \cdot x_{Tj} + \Delta b_i \\ & \doteq - \sum_{j=1}^n \Delta a_{ij} \cdot \bar{x}_j + \Delta b_i \quad (i=1, \dots, n) \end{aligned} \quad (24)$$

を満足するとして、解 x_{Tj} の代わりに、(2) を満足している数値解 $\bar{x}_j$ を用いる。ここで、数値解 $\bar{x}_j$ が (24) を満足しているかどうかは、求めた数値解の最大誤差 $w_{\max j}$ を用いて、

$$\begin{aligned} & - \sum_{j=1}^n \Delta a_{ij} \cdot \bar{x}_j + \Delta b_i \\ & \doteq - \sum_{j=1}^n \Delta a_{ij} \cdot (\bar{x}_j + w_j) + \Delta b_i \quad (i=1, \dots, n) \end{aligned} \quad (25)$$

ここで、 $|w_j| \leq w_{\max j}$ が成り立つかどうかにより調べる。

すなわち、(25) を満足している場合、係数が含む誤差と数値解の誤差との積 $\Delta a_{ij} \cdot w_j$ を無視することができ、 $f_{\max ik}$ (18) の計算において、解 x_{Tj} の代わりに数値解 $\bar{x}_j$ を用いることができる。

6. 応 用 例

本章では、本論文で提案した数値解の誤差評価の有効性を四つの応用例を挙げることによって示す。

応用例1では、解の絶対値が桁違いに異なる場合の数値解の誤差評価を行う。応用例2では、式の中に誤差の消失が起きている場合の数値解の誤差評価を行う。応用例3では、係数が含む誤差の数値解に対する影響と定数項が含む誤差の数値解に対する影響との関係について述べる。応用例4では、連立1次方程式の係数行列が悪条件である場合の数値解の誤差評価を行う。

計算は、計算途中の丸め誤差を無視できるように高精度演算で行う。

6.1 解の絶対値が桁違いに異なる場合

一般に、連立1次方程式の解の絶対値が桁違いに異なる場合、他の解と比較して絶対値が桁違いに小さい解に対する数値解の精度は出にくい。しかし、問題によっては、絶対値が桁違いに小さい解に対する数値解も絶対値が大きい解に対する数値解と比べて相対誤差が同じ程度である場合がある。

この時、数値解の誤差評価を係数行列の条件数を用いて行ったのでは、解ベクトルのノルムで誤差を評価するため、絶対値が桁違いに小さい解に対する数値解の誤差を実際より大きく評価してしまう。

本論文で提案した数値解の誤差評価は、数値解の誤差を各要素ごとに評価しており、絶対値が桁違いに小さい数値解に対してもより正確な誤差評価を与えることができる。

計算例として、10元連立1次方程式

$$\begin{array}{c} a_{Ti,j} \\ \left(\begin{array}{cccccccccc} 3 \cdot r_T & -r_T & & & & & & & & \\ -r_T & 4 \cdot r_T & -r_T & & & & & & & \\ & -r_T & 4 \cdot r_T & -r_T & & & & & & \\ & & \ddots & \ddots & \ddots & & & & & \\ & & & -r_T & 4 \cdot r_T & r_T & & & & \\ & 0 & & & & -r_T & 4 \cdot r_T & & & \\ & & & & & & -r_T & 4 \cdot r_T & & \end{array} \right) \end{array} \quad \begin{array}{c} b_{Ti} \\ \left(\begin{array}{cc} 7.741203531815513D & 1 \\ 2.916036449271847D & 1 \\ 7.813496116821110D & 0 \\ 2.093619974565923D & 0 \\ 5.609837814426042D & -1 \\ 1.503151512045042D & -1 \\ 4.027682337540876D & -2 \\ 1.079214229713173D & -2 \\ 2.891745813118198D & -3 \\ 1.434116676563699D & -3 \end{array} \right) \end{array} \quad (26)$$

を挙げる。ここで、 $r_T = 3.1415926535 \dots$ であり、(26) の真の解は、

$k=3$

1. 560931360520551D -8
 4. 682794081561651D -8
 1. 254750891653368D -8
 3. 362094850518182D -9
 9. 008704855390526D -10
 2. 413870816380282D -10
 6. 467788101306020D -11
 1. 732443241421256D -11
 4. 619848643790017D -12
 1. 154962160947504D -12

(38)

 $k=10$

1. 539949547930006D -16
 4. 619848643790017D -16
 1. 693944502723006D -15
 6. 313793146513020D -15
 2. 356122808332907D -14
 8. 793111918680327D -14
 3. 281632486638840D -13
 1. 224721875468733D -12
 4. 570724253211046D -12
 1. 142681063302761D -12

 $k=11$

3. 849873869825016D -17
 1. 154962160947504D -16
 4. 234861256807514D -16
 1. 578448286628255D -15
 5. 890307020832268D -15
 2. 198277979670082D -14
 8. 204081216597097D -14
 3. 061804688671831D -13
 1. 142681063302761D -12
 4. 264543784343862D -12

となる。

こうして、係数、定数項を含む互いに独立した誤差に起因する数値解の最大誤差は、

 $w_{\max j}$

3. 396564490079177D -7
 2. 970587200934544D -7
 1. 548246278225048D -7
 7. 166448055704136D -8
 2. 854645478455057D -8
 1. 095544303505128D -8
 3. 893544658213606D -9
 1. 365077037785090D -9
 4. 521754687664675D -10
 1. 378696605261696D -10

(39)

となる。ここで、数値解(30)の真の誤差と相対誤差は、

 $x_{Tj} - \bar{x}_j$

2. 002631447339809D -7
 1. 355000351921376D -7
 4. 777797535382433D -8
 1. 351584721920318D -8
 4. 536424069723211D -9
 1. 532718771346531D -9
 5. 042128173315950D -10
 1. 752409693051016D -10
 5. 142280264797126D -11
 1. 267662439862621D -11

(40)

 $(x_{Tj} - \bar{x}_j)/x_{Tj}$

2. 002631447339809D -8
 2. 528465078822144D -8
 2. 218204502197201D -8
 1. 756410097660650D -8
 1. 760085770551473D -8
 1. 849477122393291D -8
 1. 946261208318126D -8
 2. 208912396368162D -8
 2. 150273059430408D -8
 1. 780457249945601D -8

(41)

である。数値解の最大誤差(39)と数値解の真の誤差(40)とを比較すると、

$$|x_{Tj} - \bar{x}_j| \leq w_{\max j} \quad (j=1, \dots, 10) \quad (42)$$

となり、数値解の真の誤差は数値解の最大誤差より絶対値が小さくなっている。

数値解の誤差評価をノルムを用いて行った場合、それぞれの数値解は同じ位の大きさの誤差を含んでいると評価する。したがって、解の絶対値の大きさが異なる場合、絶対値の小さい解に対する数値解は、絶対値の大きい解に対する数値解に比べて相対誤差が大きいと評価されてしまう。しかし、(41)に示すように、解の絶対値の大きさが桁違いに異なる場合であっても、相対誤差がほぼ等しい。すなわち、数値解の精度はそれぞれほぼ同じ程度である場合がある。本論文で提案した数値解の誤差評価は、数値解の各要素ごとに行っているため、解の絶対値が桁違いに異なる場合においても、数値解の最大誤差(39)に示すように、数値解の誤差の範囲を十分正確に示すことができる。

6.2 誤差の消失が起きている場合

数値解の誤差を各要素ごとに評価する方法として区間演算がある。しかし、この方法では、数値解が実際に含んでいる誤差と比較して誤差を過大評価する場合

がある。その理由として、数値計算において誤差の消失が起こる場合があるのに、区間演算ではそれが考慮されていないなどがある。

誤差の消失には、式の中での誤差の消失、計算途中での誤差の消失などが挙げられる。しかし、誤差の消失が、どこでどのように発生しているのかを実際に調べることは難しい。ここで、式の中における誤差の消失とは、次のようなことである。例えば、係数 a_{11} が含む誤差 Δa_{11} と係数 a_{12} が含む誤差 Δa_{12} とが従属関係にあり値が等しく、解 x_{T1} と x_{T2} の絶対値が等しく符号が逆である場合、

$$\begin{aligned}\Delta a_{11} \cdot x_{T1} + \Delta a_{12} \cdot x_{T2} &= \Delta a_{11} \cdot (x_{T1} + x_{T2}) \\ &= \Delta a_{11} \cdot (x_{T1} - x_{T1}) \\ &= 0\end{aligned}$$

となる。この時、第 1 式

$$\sum_{j=1}^n a_{1j} \cdot x_{Tj} = b_1$$

において、誤差 Δa_{11} と Δa_{12} による誤差は消失する。

本論文で提案した数値解の誤差評価は、誤差の消失も考慮しており、これにより誤差の消失が起きている場合でもより正確な誤差評価を行うことができる。

計算例として、4 元連立 1 次方程式

$$\begin{array}{cc} & a_{Tij} \\ & \begin{array}{cc} j=1 & j=2 \\ \left[\begin{array}{cc} 1.414213562373095D & 0 \\ 0.000000000000000D & 0 \\ 1.732050807568878D & -1 \\ 2.236067977499790D & -2 \end{array} \right] & \begin{array}{cc} 0.000000000000000D & 0 \\ 1.414213562373095D & 0 \\ 1.732050807568878D & -1 \\ 2.236067977499790D & -2 \end{array} \\ & \begin{array}{cc} j=3 & j=4 \\ \left[\begin{array}{cc} 1.732050807568878D & -1 \\ 1.732050807568878D & -1 \\ 2.449489742783179D & -2 \\ 2.645751311064591D & -3 \end{array} \right] & \begin{array}{cc} 2.236067977499790D & -2 \\ 2.236067977499790D & -2 \\ 2.645751311064591D & -3 \\ 2.828427124746191D & -4 \end{array} \end{array} \end{array} \quad (43)$$

$$\begin{array}{cc} & b_{Ti} \\ & \left[\begin{array}{cc} 2.859466299820295D & 1 \\ -6.026299576496439D & 1 \\ -6.165739010948970D & -1 \\ -5.738929942506544D & -2 \end{array} \right] \end{array}$$

を挙げる。ここで、(43)の真の解は、

$$\begin{array}{cc} & x_{Tj} \\ & \begin{array}{cc} 3.141592653589793D & 1 \\ -3.141592653589793D & 1 \\ 3.141592653589793D & 2 \\ -3.141592653589793D & 3 \end{array} \end{array} \quad (44)$$

である。また、(43)の係数行列の条件数は、

$$\|A\|_2 \cdot \|A^{-1}\|_2 = 2.556832687706571D \quad 2$$

である。実際に(43)を解く場合、係数、定数項は、9桁目を四捨五入されて8桁に丸めて与えられたとする。したがって、4元連立1次方程式

$$\begin{array}{cc} & a_{ij} \\ & \begin{array}{cc} j=1 & j=2 \\ \left[\begin{array}{cc} 1.414213600000000D & 0 \\ 0.000000000000000D & 0 \\ 1.732050800000000D & -1 \\ 2.236068000000000D & -2 \end{array} \right] & \begin{array}{cc} 0.000000000000000D & 0 \\ 1.414213600000000D & 0 \\ 1.732050800000000D & -1 \\ 2.236068000000000D & -2 \end{array} \\ & \begin{array}{cc} j=3 & j=4 \\ \left[\begin{array}{cc} 1.732050800000000D & -1 \\ 1.732050800000000D & -1 \\ 2.449489700000000D & -2 \\ 2.645751300000000D & -3 \end{array} \right] & \begin{array}{cc} 2.236068000000000D & -2 \\ 2.236068000000000D & -2 \\ 2.645751300000000D & -3 \\ 2.828427100000000D & -4 \end{array} \end{array} \end{array} \quad (45)$$

$$\begin{array}{cc} & b_i \\ & \left[\begin{array}{cc} 2.859466300000000D & 1 \\ -6.026299600000000D & 1 \\ -6.165739000000000D & -1 \\ -5.738929900000000D & -2 \end{array} \right] \end{array}$$

を解くことになる。ここで、(45)の係数行列の条件数は、

$$\|A\|_2 \cdot \|A^{-1}\|_2 = 2.556832936384299D \quad 2$$

である。そして、(45)を解くと、数値解

$$\begin{array}{cc} & \bar{x}_j \\ & \begin{array}{cc} 3.141592351905940D & 1 \\ -3.141592804848316D & 1 \\ 3.141593674156939D & 2 \\ -3.141593263854156D & 3 \end{array} \end{array} \quad (46)$$

を得る。

それでは、係数 a_{ij} 、定数項 b_i が含む互いに独立した誤差ごとに、数値解に対する影響を調べる。

係数 a_{ij} 、定数項 b_i が含む誤差は、(43)の係数 a_{Tji} 、定数項 b_{Ti} を入力する時に、係数、定数項の9桁目を四捨五入して8桁に丸めたことによるもので、係数 a_{ij} が含む誤差は、

$$\begin{aligned} -5 \cdot 10^{-8} &\leq \Delta a_{11} \equiv \Delta a_{22} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_1 \leq 5 \cdot 10^{-8} \\ -5 \cdot 10^{-9} &\leq \Delta a_{13} \equiv \Delta a_{23} \equiv \Delta a_{31} \equiv \Delta a_{32} \equiv 5 \cdot 10^{-9} \cdot \varepsilon_2 \\ &\leq 5 \cdot 10^{-9} \\ -5 \cdot 10^{-10} &\leq \Delta a_{14} \equiv \Delta a_{24} \equiv \Delta a_{41} \equiv \Delta a_{42} \equiv 5 \cdot 10^{-10} \cdot \varepsilon_3 \\ &\leq 5 \cdot 10^{-10} \\ -5 \cdot 10^{-10} &\leq \Delta a_{33} \equiv 5 \cdot 10^{-10} \cdot \varepsilon_4 \leq 5 \cdot 10^{-10} \quad (47) \\ -5 \cdot 10^{-11} &\leq \Delta a_{34} \equiv \Delta a_{43} \equiv 5 \cdot 10^{-11} \cdot \varepsilon_5 \leq 5 \cdot 10^{-11} \end{aligned}$$

$-5 \cdot 10^{-12} \leq \Delta a_{44} \equiv 5 \cdot 10^{-12} \cdot \varepsilon_6 \leq 5 \cdot 10^{-12}$
 の範囲にあり、定数項 b_i が含む誤差は、

$$\begin{aligned} -5 \cdot 10^{-7} &\leq \Delta b_1 \equiv 5 \cdot 10^{-7} \cdot \varepsilon_7 \leq 5 \cdot 10^{-7} \\ -5 \cdot 10^{-7} &\leq \Delta b_2 \equiv 5 \cdot 10^{-7} \cdot \varepsilon_8 \leq 5 \cdot 10^{-7} \\ -5 \cdot 10^{-8} &\leq \Delta b_3 \equiv 5 \cdot 10^{-8} \cdot \varepsilon_9 \leq 5 \cdot 10^{-8} \\ -5 \cdot 10^{-10} &\leq \Delta b_4 \equiv 5 \cdot 10^{-10} \cdot \varepsilon_{10} \leq 5 \cdot 10^{-10} \end{aligned} \quad (48)$$

Δ の範囲にある。次に、係数、定数項が含む誤差(47), (48)と数値解(46)とを用いて、係数、定数項が含む互いに独立した誤差 ε_k ごとに $f_{\max ik}$ を計算すると、係数 a_{ij} が含む互いに独立した誤差 ε_k ($k=1, \dots, 6$) に対する $f_{\max ik}$ は、

$k=1$		$f_{\max ik}$ $k=2$		$k=3$	
-1.570796175952970D	-6	-1.570796837078469D	-6	1.570796631927078D	-6
1.570796402424158D	-6	-1.570796837078469D	-6	1.570796631927078D	-6
0.000000000000000D	0	2.264711879900748D	-14	0.000000000000000D	0
0.000000000000000D	0	0.000000000000000D	0	2.264711878577260D	-15
(49)					
$k=4$		$k=5$		$k=6$	
0.000000000000000D	0	0.000000000000000D	0	0.000000000000000D	0
0.000000000000000D	0	0.000000000000000D	0	0.000000000000000D	0
-1.570796837078470D	-7	1.570796631927078D	-7	0.000000000000000D	0
0.000000000000000D	0	-1.570796837078469D	-8	1.570796631927077D	-8

となり、定数項 b_i が含む互いに独立した誤差 ε_k ($k=7, \dots, 10$) に対する $f_{\max ik}$ は、

$f_{\max ik}$ $k=7$		$k=8$		$k=9$	
5.000000000000000D	-7	0.000000000000000D	0	0.000000000000000D	0
0.000000000000000D	0	5.000000000000000D	-7	0.000000000000000D	0
0.000000000000000D	0	0.000000000000000D	0	5.000000000000000D	-9
0.000000000000000D	0	0.000000000000000D	0	0.000000000000000D	0
(50)					
$k=10$					
0.000000000000000D	0				
0.000000000000000D	0				
0.000000000000000D	0				
5.000000000000000D	-10				

となる。

そして、(45)の係数 a_{ij} を係数とし、 $f_{\max ik}$ をそれぞれ定数項として、10個の4元連立1方程式

$$\sum_{j=1}^4 a_{ij} \cdot w_{jk} = f_{\max ik} \quad (i=1, \dots, 4; k=1, \dots, 10) \quad (51)$$

を作り、係数、定数項が含む互いに独立した誤差 ε_k それぞれに起因する数値解の誤差 w_{jk} を求めると、係数 a_{ij} が含む互いに独立した誤差 ε_k ($k=1, \dots, 6$) により数値解に入る誤差 w_{jk} は、

w_{jk} $k=1$		$k=2$		$k=3$	
-1.110720664359088D	-6	-1.947167352464479D	-7	1.947166619362522D	-7
1.110720692432509D	-6	-1.947167352464479D	-7	1.947166619362522D	-7
-3.976980908948050D	-12	5.516844549610524D	-5	-5.516843048753866D	-5
3.498185136507258D	-11	-4.852661366995054D	-4	4.852660159055451D	-4
(52)					
$k=4$		$k=5$		$k=6$	
2.758421101770325D	-6	-5.184750642142943D	-6	2.426329583744855D	-6
2.758421101770325D	-6	-5.184750642142943D	-6	2.426329583744855D	-6
-1.627226478740849D	-4	2.713212587127153D	-4	-1.085986179074104D	-4
1.085986320907599D	-3	-1.773732974592422D	-3	6.877467056962550D	-4

となり、定数項 b_i が含む互いに独立した誤差 ε_k ($k=7, \dots, 10$) により数値解に入る誤差 w_{jk} は、

$$\begin{array}{rcccl}
 & & w_{jk} & & \\
 & k=7 & k=8 & k=9 & \\
 2.077667481475196D & -7 & -1.457866330390281D & -7 & -8.780324217168399D & -8 \\
 -1.457866330390281D & -7 & 2.077667481475196D & -7 & -8.780324217168399D & -8 \\
 -8.780324217168400D & -6 & -8.780324217168400D & -6 & 5.179621069798345D & -6 \\
 7.723245436196903D & -5 & 7.723245436196903D & -5 & -3.456800699087949D & -5 \\
 & k=10 & & & & \\
 7.723245436196900D & -8 & & & & \\
 7.723245436196900D & -8 & & & & \\
 -3.456800699087950D & -6 & & & & \\
 2.189165330882193D & -5 & & & &
 \end{array} \quad (53)$$

となる。

こうして、係数、定数項が含む互いに独立した誤差に起因する数値解の最大誤差は、

$$\begin{array}{rcl}
 w_{\max j} & & \\
 1.238824446692011D & -5 & \\
 1.238824449499353D & -5 & \\
 6.791764746580585D & -4 & \\
 4.728922757806818D & -3 &
 \end{array} \quad (54)$$

となる。ここで、数値解(46)の真の誤差と相対誤差は、

$$\begin{array}{rcl}
 x_{Tj} - \bar{x}_j & & \\
 3.016838530101040D & -6 & \\
 1.512585232177344D & -6 & \\
 -1.020567146383655D & -4 & \\
 6.102643619669834D & -4 & \\
 (x_{Tj} - \bar{x}_j) / x_{Tj} & & \\
 9.602895291513362D & -8 & \\
 -4.814708330976530D & -8 & \\
 -3.248566122082973D & -7 & \\
 -1.942531795997341D & -7 &
 \end{array} \quad (55)$$

である。数値解の最大誤差(54)と数値解の真の誤差(55)とを比較すると、

$$|x_{Tj} - \bar{x}_j| \leq w_{\max j} \quad (j=1, \dots, 4) \quad (57)$$

となり、数値解の真の誤差は数値解の最大誤差より絶対値が小さくなっている。

係数が含まれている互いに独立した誤差の数値解に対する影響 w_{jk} ($j=1, 2; k=2, 3$) と w_{jk} ($j=1, 2; k=1, 4, 5, 6$) とを比べると、 w_{jk} ($j=1, 2; k=2, 3$) の絶対値が w_{jk} ($j=1, 2; k=1, 4, 5, 6$) の絶対値よりも桁違いに小さくなっている。これは、解 x_{T1} と x_{T2} の絶対値が等しく、符号が逆であり、かつ、係数を含む誤差 Δa_{31} と Δa_{32} , Δa_{41} と Δa_{42} がそれぞれ恒等的に等しいことにより、

$$\begin{aligned}
 \Delta a_{31} \cdot x_{T1} + \Delta a_{32} \cdot x_{T2} &= 0 \\
 \Delta a_{41} \cdot x_{T1} + \Delta a_{42} \cdot x_{T2} &= 0
 \end{aligned}$$

となる。つまり Δa_{31} と Δa_{32} , Δa_{41} と Δa_{42} による誤差の消失が起きているからである。この場合、係数を含む誤差 Δa_{31} と Δa_{32} , Δa_{41} と Δa_{42} の数値解に対する影響は小さくなる。このように、本論文で提案した数値解の誤差評価は、誤差の消失が起きている場合でも、誤差の消失に対する評価をより正確に行うことができる。

6.3 解が絶対値最小固有値に対する固有ベクトルである場合

解の絶対値が同じ位の大きさであっても、解それぞれの要素の関係によって、係数を含む誤差と定数項が含む誤差それぞれの数値解に対する影響が異なる場合がある。特に、解が絶対値最小固有値に対応する固有ベクトルである場合には、定数項が含む誤差は数値解に対してあまり影響せず、係数を含む誤差が数値解に対して大きく影響する。

そこで、解が絶対値最小固有値に対応する固有ベクトルである場合を例として挙げ、本論文で提案した数値解の誤差評価を用いることによって、係数を含む誤差と定数項が含む誤差がそれぞれどのような関係を持って数値解に影響しているかを調べる。

計算例として、4元連立1次方程式

$$\begin{array}{rcccl}
 & & a_{Tij} & & \\
 & j=1 & j=2 & & \\
 \left[\begin{array}{l} 3.098822653589793D \\ 2.236067977499790D \\ 1.732050807568877D \\ 1.414213562373095D \end{array} \right] & \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array} & \begin{array}{l} 2.236067977499790D \\ 3.098822653589793D \\ 2.718281828459045D \\ 2.828427124746190D \end{array} & \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array} &
 \end{array}$$

$$\begin{array}{cc}
 j=3 & j=4 \\
 \begin{array}{l} 1.732050807568877D \\ 2.718281828459045D \\ 2.407829742783178D \\ 2.645751311064592D \end{array} & \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array} \begin{array}{l} 1.414213562373095D \\ 2.828427124746190D \\ 2.645751311064592D \\ 2.407829742783178D \end{array} \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array}
 \end{array}$$

$$\begin{array}{c} b_{Ti} \\ \begin{bmatrix} -8.980965059357526D & -5 \\ 4.807087679936097D & -4 \\ -4.075426740590138D & -4 \\ -6.410009573720731D & -5 \end{bmatrix} \end{array} \quad (58)$$

を挙げる。ここで、(58)の真の解は、(58)の係数行列の絶対値最小固有値に対応する固有ベクトル

$$\begin{array}{c} x_{Ti} \\ \begin{array}{l} 1.403710153801282D \\ -7.513399442148851D \\ 6.369825357452139D \\ 1.001874015229526D \end{array} \end{array} \quad (59)$$

である。また、(58)の係数行列の条件数は、

$$\|A\|_2 \cdot \|A^{-1}\|_2 = 1.226440285686587D \quad 2$$

である。実際に(58)を解く場合、係数、定数項は、9桁目を四捨五入して8桁に丸めて与えられたとする。したがって、4元連立1次方程式

$$\begin{array}{cc}
 a_{ij} & \\
 j=1 & j=2 \\
 \begin{array}{l} 3.098822700000000D \\ 2.236068000000000D \\ 1.732050800000000D \\ 1.414213600000000D \end{array} & \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array} \begin{array}{l} 2.236068000000000D \\ 3.098822700000000D \\ 2.718281800000000D \\ 2.828427100000000D \end{array} \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array}
 \end{array}$$

$$\begin{array}{cc}
 j=3 & j=4 \\
 \begin{array}{l} 1.732050800000000D \\ 2.718281800000000D \\ 2.407829700000000D \\ 2.645751300000000D \end{array} & \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array} \begin{array}{l} 1.414213600000000D \\ 2.828427100000000D \\ 2.645751300000000D \\ 2.407829700000000D \end{array} \begin{array}{l} 0 \\ 0 \\ 0 \\ 0 \end{array}
 \end{array}$$

$$\begin{array}{c} b_i \\ \begin{bmatrix} -8.980965100000000D & -5 \\ 4.807087700000000D & -4 \\ -4.075426700000000D & -4 \\ -6.410009600000000D & -5 \end{bmatrix} \end{array} \quad (60)$$

を解くことになる。ここで、(60)の係数行列の条件数は、

$$\|A\|_2 \cdot \|A^{-1}\|_2 = 1.226472718095646D \quad 2$$

である。そして、(60)を解くと、数値解

$$\begin{array}{c} \bar{x}_j \\ \begin{array}{l} 1.408784600929528D \\ -7.513797102682162D \\ 6.370161752465768D \\ 1.001927726145258D \end{array} \end{array} \quad (61)$$

を得る。

それでは、係数 a_{ij} 、定数項 b_i が含む互いに独立した誤差ごとに数値解に対する影響を調べる。

係数 a_{ij} 、定数項 b_i が含む誤差は、(58)の係数 a_{Tij} 、定数項 b_{Ti} を入力する時に、係数、定数項の9桁目を四捨五入して8桁に丸めたことによるもので、係数 a_{ij} が含む誤差は、

$$\begin{array}{l} -5 \cdot 10^{-8} \leq \Delta a_{11} \equiv \Delta a_{22} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_1 \leq 5 \cdot 10^{-8} \\ -5 \cdot 10^{-8} \leq \Delta a_{1k} \equiv \Delta a_{k1} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_k \leq 5 \cdot 10^{-8} \\ (k=2, 3, 4) \\ -5 \cdot 10^{-8} \leq \Delta a_{23} \equiv \Delta a_{32} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_5 \leq 5 \cdot 10^{-8} \\ -5 \cdot 10^{-8} \leq \Delta a_{24} \equiv \Delta a_{42} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_6 \leq 5 \cdot 10^{-8} \\ -5 \cdot 10^{-8} \leq \Delta a_{33} \equiv \Delta a_{44} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_8 \leq 5 \cdot 10^{-8} \\ -5 \cdot 10^{-8} \leq \Delta a_{34} \equiv \Delta a_{43} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_8 \leq 5 \cdot 10^{-8} \end{array} \quad (62)$$

の範囲にあり、定数項 b_i が含む誤差は、

$$\begin{array}{l} -5 \cdot 10^{-13} \leq \Delta b_1 \equiv 5 \cdot 10^{-13} \cdot \varepsilon_9 \leq 5 \cdot 10^{-13} \\ -5 \cdot 10^{-12} \leq \Delta b_2 \equiv 5 \cdot 10^{-12} \cdot \varepsilon_{10} \leq 5 \cdot 10^{-12} \\ -5 \cdot 10^{-12} \leq \Delta b_3 \equiv 5 \cdot 10^{-12} \cdot \varepsilon_{11} \leq 5 \cdot 10^{-12} \\ -5 \cdot 10^{-13} \leq \Delta b_4 \equiv 5 \cdot 10^{-13} \cdot \varepsilon_{12} \leq 5 \cdot 10^{-13} \end{array} \quad (63)$$

の範囲にある。次に、係数、定数項が含む誤差(62)、(63)と数値解(61)とを用いて、係数、定数項が含む互いに独立した誤差 ε_k ごとに $f_{\max ik}$ を計算すると、係数 a_{ij} が含む互いに独立した誤差 ε_k ($k=1, \dots, 8$) に対する $f_{\max ik}$ は、

$$\begin{array}{cc}
 f_{\max ik} & \\
 k=1 & k=2 \\
 \begin{array}{l} -7.018923004647637D \\ 3.756898551341081D \\ 0.000000000000000D \\ 0.000000000000000D \end{array} & \begin{array}{l} -9 \\ -8 \\ 0 \\ 0 \end{array} \begin{array}{l} 3.756898551341081D \\ -7.018923004647637D \\ 0.000000000000000D \\ 0.000000000000000D \end{array} \begin{array}{l} -8 \\ -9 \\ 0 \\ 0 \end{array} \\
 k=3 & k=4 \\
 \begin{array}{l} -3.185080876232884D \\ 0.000000000000000D \\ -7.018923004647637D \\ 0.000000000000000D \end{array} & \begin{array}{l} -8 \\ 0 \\ -9 \\ 0 \end{array} \begin{array}{l} -5.009638630726287D \\ 0.000000000000000D \\ 0.000000000000000D \\ -7.018923004647637D \end{array} \begin{array}{l} -9 \\ 0 \\ 0 \\ -9 \end{array} \\
 k=5 & k=6 \\
 \begin{array}{l} 0.000000000000000D \\ -3.185080876232884D \\ 3.756898551341081D \\ 0.000000000000000D \end{array} & \begin{array}{l} 0 \\ -8 \\ -8 \\ 0 \end{array} \begin{array}{l} 0.000000000000000D \\ -5.009638630726287D \\ 0.000000000000000D \\ 3.756898551341081D \end{array} \begin{array}{l} 0 \\ -9 \\ 0 \\ -8 \end{array} \\
 k=7 & k=8 \\
 \begin{array}{l} 0.000000000000000D \\ 0.000000000000000D \\ -3.185080876232884D \\ -5.009638630726287D \end{array} & \begin{array}{l} 0 \\ 0 \\ -8 \\ -9 \end{array} \begin{array}{l} 0.000000000000000D \\ 0.000000000000000D \\ -5.009638630726287D \\ -3.185080876232884D \end{array} \begin{array}{l} 0 \\ 0 \\ -9 \\ -8 \end{array} \end{array} \quad (64)$$

となり、定数項 b_i が含む互いに独立した誤差 ε_k ($k=9, \dots, 12$) に対する $f_{\max ik}$ は、

$$f_{\max ik} \quad (65)$$

$k=9$		$k=10$		$k=11$	
5.000000000000000D	-13	0.000000000000000D	0	0.000000000000000D	0
0.000000000000000D	0	5.000000000000000D	-12	0.000000000000000D	0
0.000000000000000D	0	0.000000000000000D	0	5.000000000000000D	-12
0.000000000000000D	0	0.000000000000000D	0	0.000000000000000D	0
$k=12$					
0.000000000000000D	0				
0.000000000000000D	0				
0.000000000000000D	0				
5.000000000000000D	-13				

となる。

そして、(60)の係数 a_{ij} を係数とし、 $f_{\max ik}$ をそれぞれ定数項として、12 個の4元連立1次方程式

$$\sum_{j=1}^4 a_{ij} \cdot w_{jk} = f_{\max ik} \quad (i=1, \dots, 4; k=1, \dots, 12) \quad (66)$$

を作り、係数、定数項が含む互いに独立した誤差 ε_k それぞれに起因する数値解の誤差 w_{jk} を求めると、係数 a_{ij} が含む互いに独立した誤差 ε_k ($k=1, \dots, 8$) により数値解に入る誤差 w_{jk} は、

$$w_{jk} \quad (67)$$

$k=1$		$k=2$		$k=3$	
6.412072213738757D	-6	-2.300640065541626D	-6	1.948398349210947D	-6
-3.431883665833530D	-5	1.239395470434443D	-5	-1.050218744561639D	-5
2.906464425010269D	-5	-1.049108924684617D	-5	8.903438971839271D	-6
4.610985406198316D	-6	-1.679931051306133D	-6	1.409125779358512D	-6
$k=4$		$k=5$		$k=6$	
3.107994270162306D	-7	-1.049920445105616D	-5	-1.674627245753724D	-6
-1.658411535660343D	-6	5.619845107952411D	-5	8.874366170860612D	-6
1.389401745487055D	-6	-4.766944733833003D	-5	-7.438941521188016D	-6
2.359506390782011D	-7	-7.468801477869435D	-6	-1.251358832100687D	-6
$k=7$		$k=8$			
4.558969908005028D	-6	1.418261192290553D	-6		
-2.440397910876985D	-5	-7.519828098427866D	-6		
2.072028203893456D	-5	6.313225191391216D	-6		
3.219412001471084D	-6	1.050110291590368D	-6		

となり、定数項 b_i が含む互いに独立した誤差 ε_k ($k=9, \dots, 12$) により数値解に入る誤差 w_{jk} は、

$$w_{jk} \quad (68)$$

$k=9$		$k=10$		$k=11$	
-1.520627861449414D	-11	8.249635612280998D	-10	-6.979231744587608D	-10
8.249635612280989D	-11	-4.413316604319554D	-9	3.737778657230482D	-9
-6.979231744587605D	-11	3.737778657230484D	-9	-3.175384212642720D	-9
-1.128688157412480D	-11	5.925821312297877D	-10	-4.916232631562779D	-10
$k=12$					
-1.128688157412485D	-11				
5.925821312297895D	-11				
-4.916232631562795D	-11				
-8.752357241168542D	-12				

となる。

こうして、係数、定数項が含む互いに独立した誤差に起因する数値解の最大誤差は、

$$\begin{aligned} w_{\max j} \\ 2.912452223250891D \quad -5 \\ 1.558783076513697D \quad -4 \\ 1.319975024216327D \quad -4 \\ 2.092677972360594D \quad -5 \end{aligned} \quad (69)$$

となる。ここで、数値解(61)の真の誤差と相対誤差は、

$$\begin{aligned} x_{Tj} - \bar{x}_j \\ -7.444712824572485D \quad -6 \\ 3.976605333111394D \quad -5 \\ -3.363950136292537D \quad -5 \\ -5.371091573169728D \quad -6 \\ (x_{Tj} - \bar{x}_j) / x_{Tj} \\ -5.303596903115662D \quad -5 \\ -5.292684574712394D \quad -5 \\ -5.281071218627698D \quad -5 \\ -5.361044893393337D \quad -5 \end{aligned} \quad (70)$$

である。数値解の最大誤差(69)と数値解の真の誤差(70)とを比較すると、

$$|x_{Tj} - \bar{x}_j| \leq w_{\max j} \quad (j=1, \dots, 4) \quad (72)$$

となり、数値解の真の誤差は数値解の最大誤差より絶対値が小さくなっている。

係数 a_{ij} が含む誤差の数値解に対する影響(67)と、定数項 b_i が含む誤差の数値解に対する影響(68)とを比較すると、(67)の方が(68)に比べて絶対値が桁違いに大きい。すなわち、解が絶対値最小固有値に対応する固有ベクトルである場合には、定数項が含む誤差は数値解の誤差としてはほとんど影響せず、係数が含む誤差が数値解の誤差として大きく影響する。

また、解が絶対値最大固有値に対応する固有ベクトルである場合には、係数が含む誤差と定数項が含む誤差とは同じ程度の大きさで数値解に影響する。同じ桁数で計算した場合、特に、係数と解の積の総和

$$\sum_{j=1}^n a_{Tij} \cdot x_{Tj} \quad (i=1, \dots, n) \quad (73)$$

が桁上りを起こし、定数項が係数と解の積より絶対値が桁違いに大きくなると、数値解の誤差は定数項が含む誤差で決まる。

6.4 係数行列が悪条件である場合

一般に、係数行列が悪条件であるかそうではないかを議論する場合、与えられた係数行列の条件数などを用いて行われている。そして、条件数が大きな場合、

係数行列は悪条件であり数値解の精度は出にくいとされている。しかし、与えられた係数行列の条件数がそれほど大きくない場合でも、係数行列の各要素が大きな誤差を含んでいる場合、数値解の精度は悪い。この場合、その係数行列は悪条件であるべきだと考える。

そこで、本論文で提案した数値解の誤差評価を、係数行列が悪条件である連立1次方程式の数値解の誤差評価に適用することにより、数値解の誤差は、計算途中において起こる誤差によるものであるとは限らず、係数、定数項が与えられた時点において、係数、定数項に含まれる誤差で決まってしまうことを示す。

計算例として、4元連立1次方程式

$$\begin{aligned} & a_{Tij} \\ & \begin{matrix} j=1 & j=2 \\ \left[\begin{array}{cc} 3.099463496548540D & 0 \\ 2.236067977499790D & 0 \\ 1.732050807568877D & 0 \\ 1.414213562373095D & 0 \end{array} \right] \end{matrix} \\ & \begin{matrix} j=3 & j=4 \\ \left[\begin{array}{cc} 2.236067977499790D & 0 \\ 1.732050807568877D & 0 \\ 2.718281828459045D & 0 \\ 2.828427124746190D & 0 \end{array} \right] \end{matrix} \\ & b_{Ti} \\ & \left[\begin{array}{c} 3.951718059394517D \\ 2.333583630614791D \\ 1.566060347982934D \\ 2.006803101763138D \end{array} \right] \end{aligned} \quad (74)$$

を挙げる。ここで、(74)の真の解は、

$$\begin{aligned} x_{Tj} \\ 1.414213562373095D \quad 0 \\ -1.732050807568877D \quad 0 \\ 3.141592653589793D \quad 0 \\ -1.414213562373095D \quad 0 \end{aligned} \quad (75)$$

である。また、(74)の係数行列の条件数は、

$$\|A\|_2 \cdot \|A^{-1}\|_2 = 3.978062260794834D \quad 3$$

である。実際に(74)を解く場合、係数、定数項は、9桁目を四捨五入されて8桁に丸めて与えられたとする。したがって、4元連立1次方程式

$$\begin{aligned} & a_{ij} \\ & \begin{matrix} j=1 & j=2 \\ \left[\begin{array}{cc} 3.099463500000000D & 0 \\ 2.236068000000000D & 0 \\ 1.732050800000000D & 0 \\ 1.414213600000000D & 0 \end{array} \right] \end{matrix} \\ & \begin{matrix} j=3 & j=4 \\ \left[\begin{array}{cc} 2.236068000000000D & 0 \\ 1.732050800000000D & 0 \\ 2.718281800000000D & 0 \\ 2.828427100000000D & 0 \end{array} \right] \end{matrix} \end{aligned}$$

$$\begin{array}{cc}
 j=3 & j=4 \\
 \begin{array}{l} 1.732050800000000D \ 0 \\ 2.718281800000000D \ 0 \\ 2.408469500000000D \ 0 \\ 2.645751300000000D \ 0 \end{array} & \begin{array}{l} 1.414213600000000D \ 0 \\ 2.828427100000000D \ 0 \\ 2.645751300000000D \ 0 \\ 2.408469500000000D \ 0 \end{array}
 \end{array}$$

$$\begin{array}{cc}
 b_i & (76) \\
 \begin{array}{l} 3.951718100000000D \ 0 \\ 2.333583600000000D \ 0 \\ 1.566060300000000D \ 0 \\ 2.006803100000000D \ 0 \end{array}
 \end{array}$$

を解くことになる。ここで、(76)の係数行列の条件数は、

$$\|A\|_2 \cdot \|A^{-1}\|_2 = 3.952390850592313D \ 3$$

である。そして、(76)を解くと、数値解

$$\begin{array}{cc}
 \bar{x}_j & (77) \\
 \begin{array}{l} 1.422095628203050D \ 0 \\ -1.774239144948132D \ 0 \\ 3.177359416399883D \ 0 \\ -1.408587737964049D \ 0 \end{array}
 \end{array}$$

を得る。

それでは、係数 a_{ij} 、定数項 b_i が含む互いに独立した誤差ごとに、数値解に対する影響を調べる。

係数 a_{ij} 、定数項 b_i が含む誤差は、(74)の係数 a_{Tij} 、定数項 b_{Ti} を入力する時に、係数、定数項の9桁目を四捨五入して8桁に丸めたことによるもので、係数 a_{ij} が含む誤差は、

$$\begin{array}{l}
 -5 \cdot 10^{-8} \leq \Delta a_{11} \equiv \Delta a_{22} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_1 \leq 5 \cdot 10^{-8} \\
 -5 \cdot 10^{-8} \leq \Delta a_{1k} \equiv \Delta a_{k1} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_k \leq 5 \cdot 10^{-8} \\
 (k=2, 3, 4) \\
 -5 \cdot 10^{-8} \leq \Delta a_{23} \equiv \Delta a_{32} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_5 \leq 5 \cdot 10^{-8} \\
 -5 \cdot 10^{-8} \leq \Delta a_{24} \equiv \Delta a_{42} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_6 \leq 5 \cdot 10^{-8} \\
 -5 \cdot 10^{-8} \leq \Delta a_{33} \equiv \Delta a_{44} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_7 \leq 5 \cdot 10^{-8} \\
 -5 \cdot 10^{-8} \leq \Delta a_{34} \equiv \Delta a_{43} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_8 \leq 5 \cdot 10^{-8}
 \end{array}
 \quad (78)$$

の範囲にあり、定数項 b_i が含む誤差は、

$$-5 \cdot 10^{-8} \leq \Delta b_{k-8} \equiv 5 \cdot 10^{-8} \cdot \varepsilon_k \leq 5 \cdot 10^{-8} \quad (79)$$

$$(k=9, \dots, 12)$$

の範囲にある。次に、係数、定数項が含む誤差(78)、(79)と数値解(77)とを用いて、係数、定数項が含む互いに独立した誤差 ε_k ごとに $f_{\max ik}$ を計算すると、係数 a_{ij} が含む互いに独立した誤差 ε_k ($k=1, \dots, 8$) に対する $f_{\max ik}$ は、

$$\begin{array}{cc}
 f_{\max ik} & \\
 \begin{array}{cc}
 k=1 & k=2 \\
 \begin{array}{l} -7.110478141015246D \ -8 \\ 8.871195724740660D \ -8 \\ 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \end{array} & \begin{array}{l} 8.871195724740660D \ -8 \\ -7.110478141015246D \ -8 \\ 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \end{array}
 \end{array} \\
 \begin{array}{cc}
 k=3 & k=4 \\
 \begin{array}{l} -1.588679708199941D \ -7 \\ 0.000000000000000D \ 0 \\ -7.110478141015246D \ -8 \\ 0.000000000000000D \ 0 \end{array} & \begin{array}{l} 7.042938689820243D \ -8 \\ 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \\ -7.110478141015246D \ -8 \end{array}
 \end{array} \\
 \begin{array}{cc}
 k=5 & k=6 \\
 \begin{array}{l} 0.000000000000000D \ 0 \\ -1.588679708199941D \ -7 \\ 8.871195724740660D \ -8 \\ 0.000000000000000D \ 0 \end{array} & \begin{array}{l} 0.000000000000000D \ 0 \\ 7.042938689820243D \ -8 \\ 0.000000000000000D \ 0 \\ 8.871195724740660D \ -8 \end{array}
 \end{array} \\
 \begin{array}{cc}
 k=7 & k=8 \\
 \begin{array}{l} 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \\ -1.588679708199941D \ -7 \\ 7.042938689820243D \ -8 \end{array} & \begin{array}{l} 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \\ 7.042938689820243D \ -8 \\ -1.588679708199941D \ -7 \end{array}
 \end{array}
 \end{array}
 \quad (80)$$

となり、定数項 b_i が含む互いに独立した誤差 ε_k ($k=9, \dots, 12$) に対する $f_{\max ik}$ は、

$$\begin{array}{cc}
 f_{\max ik} & \\
 \begin{array}{cc}
 k=9 & k=10 \\
 \begin{array}{l} 5.000000000000000D \ -8 \\ 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \end{array} & \begin{array}{l} 0.000000000000000D \ 0 \\ 5.000000000000000D \ -8 \\ 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \end{array}
 \end{array} \\
 \begin{array}{cc}
 k=11 & k=12 \\
 \begin{array}{l} 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \\ 5.000000000000000D \ -8 \\ 0.000000000000000D \ 0 \end{array} & \begin{array}{l} 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \\ 0.000000000000000D \ 0 \\ 5.000000000000000D \ -8 \end{array}
 \end{array}
 \end{array}
 \quad (81)$$

となる。

そして、(76)の係数 a_{ij} を係数とし、 $f_{\max ik}$ をそれぞれ定数項として、12個の4元連立1次方程式

$$\sum_{j=1}^n a_{ij} \cdot w_{jk} = f_{\max ik} \quad (82)$$

$$(i=1, \dots, 4; k=1, \dots, 12)$$

を作り、係数、定数項が含む互いに独立した誤差 ε_k それぞれに起因する数値解の誤差 w_{jk} を求めると、係数 a_{ij} が含む互いに独立した誤差 ε_k ($k=1, \dots, 8$) により数値解に入る誤差 w_{jk} は、

$$\begin{array}{rcl}
 & w_{jk} & x_{Tj} - \bar{x}_j \\
 k=1 & & \\
 -1.746029407966540D & -2 & 1.500932697516221D & -2 & -7.882065829954099D & -3 \\
 9.345660128412195D & -2 & -8.033768361132205D & -2 & 4.218833737925486D & -2 \\
 -7.923207813101126D & -2 & 6.810992410892566D & -2 & -3.576676281008950D & -2 \\
 -1.246194925101540D & -2 & 1.071259855616875D & -2 & -5.625824409046797D & -3
 \end{array} \quad (86)$$

$$\begin{array}{rcl}
 & & (x_{Tj} - \bar{x}_j)/x_{Tj} \\
 k=3 & & k=4 \\
 -1.540045632808241D & -2 & 6.294524017809728D & -4 & -5.573462198119314D & -3 \\
 8.243099614311981D & -2 & -3.368841850542017D & -3 & -2.435744794257555D & -2 \\
 -6.988449653143430D & -2 & 2.855952957352898D & -3 & -1.138491419924223D & -2 \\
 -1.099184815831281D & -2 & 4.493023823217153D & -4 & 3.978058589401791D & -3
 \end{array} \quad (87)$$

である。数値解の最大誤差(85)と数値解の真の誤差

$$|x_{Tj} - \bar{x}_j| \leq w_{\max j} \quad (88)$$

となり、数値解の真の誤差は数値解の最大誤差より絶対値が小さくなっている。

本章では、係数と定数項とが同じ位である連立1次方程式を例として挙げた。しかし、同じ係数行列であっても、3章で述べたように、係数と解の積の総和(73)が桁上りを起こして、定数項が係数と解の積より絶対値が桁違いに大きくなる場合がある。この時、定数項を含む誤差の位も係数を含む誤差と解の積の位より大きくなる。このため同じ桁数で計算する場合、数値解の精度は、定数項が桁上りをした桁数だけ悪くなり保証できない。

となり、定数項 b_i が含む互いに独立した誤差 ε_i ($k=9, \dots, 12$) により数値解に入る誤差 w_{jk} は、

$$\begin{array}{rcl}
 & w_{jk} & \\
 k=9 & & k=10 \\
 1.599127438753624D & -3 & -8.559264394157130D & -3 \\
 -8.559264394157127D & -3 & 4.581372755752909D & -2 \\
 7.256511797012374D & -3 & -3.884066284113616D & -2 \\
 1.141314879042368D & -3 & -6.109035741877412D & -3 \\
 k=11 & & k=12 \\
 7.256511797012387D & -3 & 1.141314879042362D & -3 \\
 -3.884066284113620D & -2 & -6.109035741877378D & -3 \\
 3.292883369745794D & -2 & 5.179314551172041D & -3 \\
 5.179314551172073D & -3 & 8.145301472566623D & -4
 \end{array} \quad (84)$$

となる。

こうして、係数、定数項を含む誤差に起因する数値解の最大誤差は、

$$\begin{array}{rcl}
 & w_{\max j} & \\
 1.452018741407582D & -1 & \\
 7.771953931779581D & -1 & \\
 6.589011133321252D & -1 & \\
 1.036364507087713D & -1 &
 \end{array} \quad (85)$$

となる。ここで、数値解(77)の真の誤差と相対誤差は

7. おわりに

本論文では、連立1次方程式の係数、定数項を含む誤差を独立な誤差に分け、それぞれ独立に数値解に対する影響を調べることによって、係数、定数項を含む誤差に起因する数値解の最大誤差を求める誤差評価方法を提案した。

本論文で提案した誤差評価方法は、数値解の各要素ごとに誤差を評価できるだけでなく、これまでの誤差評価方法では考慮されていなかった誤差の過大評価の原因である誤差の消失に対する問題も考慮している。

したがって、できるだけ正確な誤差評価を必要とする場合、あるいは、係数、定数項を含む誤差がどのような関係を持って数値解に影響しているかを調べる必要がある場合などにおいては、本論文で提案した数値解の誤差評価方法は十分有効である。

また、本論文で挙げた応用例は、係数の各要素が同じ値、またはその定数倍で与えられている場合を取り挙げたが、連立1次方程式の係数あるいは定数項が、いくつかの変数あるいは定数を含んだ関数で与えられている場合であっても、その関数を含んでいる誤差が

線形性を持っているなら、本論文で挙げた応用例と同様に誤差評価を行うことができる。

今後の課題は、本論文で提案した誤差評価方法を、より広い分野で用いられているいろいろな数値計算で得られた数値解の誤差評価に適用することによって、数値計算におけるいろいろな問題を解決していくことである。

参 考 文 献

- 1) 宇野利雄：計算機のための数値計算，朝倉書店，東京（1963）。
- 2) 一松 信，戸川隼人（編）：数値計算の誤差（bit），pp. 33-43，共立出版，東京（1975）。
- 3) 戸川隼人：マトリクスの数値計算，オーム社，東京（1971）。
- 4) 鳥居達生，牧之内三郎：数値解析，オーム社，東京（1981）。
- 5) Wilkinson, J. H. (一松 信，四糸忠雄共訳)：基本的演算における丸め誤差解析，培風館，東京（1974）。
- 6) 平野菅保ほか：コンピュータによる構造工学講座Ⅱ-1-A，計算技術および数値計算法，培風館，東京（1971）。
- 7) 平野菅保，布広永示：解の誤差と絶対値の小さい枢軸に導入する誤差，応用数学シンポジウム（1982）。

- 8) 平野菅保，布広永示：連立1次方程式の解が含む誤差，情報処理学会数値解析研究会（1984）。
- 9) 布広永示：係数，定数の誤差と数値解の誤差，数値解析シンポジウム（1985）。

（昭和60年9月18日受付）

（昭和61年8月27日採録）



布広 永示（正会員）

昭和32年7月30日生。昭和55年日本大学生産工学部数理工学科卒業。昭和57年同大学院数理工学専攻博士前期課程修了。昭和60年同大学院数理工学専攻博士後期課程中退。同年（株）日立製作所入社。現在、同社ソフトウェア工場勤務。数値解析の分野において、とくに丸め誤差解析に興味を持っている。



平野 菅保（正会員）

昭和5年8月25日生。昭和28年東京都立大学理学部物理学科卒業。理学博士。主に代数方程式、連立一次方程式の解法の丸め誤差解析に従事。現在、日本大学生産工学部数理工学科教授。日本物理学会、日本OR学会、電気学会各会員。