多重および近接固有値をもつ楕円型固有値問題に対する精度保証付き数値計算

指導教官　　中尾充宏 教授
提出者　　吉田哲基
提出日　　平成13年1月31日

目次

1 Introduction **2**

2 行列の固有値に対する Rump の方法 **3**
2.1 記号の定義 3
2.2 行列の多重固有値及び、近接固有値に対する精度保証 3
2.3 一般化固有値問題に対する精度保証の適用 5
2.4 疎行列に対する精度保証の適用 6
2.5 欠陥のある固有値に対する考察 7

3 楕円型固有値問題の単純固有値 **9**
3.1 問題と記号の定義 9
3.2 不動点定式化 10
3.3 検証条件とアルゴリズム 12

4 楕円型固有値問題の多重固有値 **16**
4.1 多重固有値に対する不動点定式化 16
4.2 検証条件とアルゴリズム 18
4.3 数値例 26

5 結果の考察と今後の課題 **27**

1 Introduction

本論文では、以下のような自己共役楕円型固有値問題：

$$\left\{ \begin{array}{rcll} -\Delta u + qu & = & \lambda u & \text{in } \Omega \\ u & = & 0 & \text{on } \partial\Omega \end{array} \right.$$

の多重固有値及び、近接固有値に対する精度保証付き数値計算を考える。

ここで、Ω は $\mathbf{R}^2$ の有界凸領域、$q \in L^\infty(\Omega)$ とする。

本固有値問題に関しては [2], [3] でその固有値の精度保証付き計算法が提案され、実際の適用が得られているが、そこで与えられている方法は単純固有値にしか適用できなかった。本論文は、行列の多重及び、近接固有値を精度保証付きで計算するという Rump の考え方を、この楕円型固有値問題にも拡張することを目標にしたものである。

本論文においてはまず第 2 節において、 Rump の方法 [5] に基づいた行列の固有値の精度保証の方法を紹介する。これは、従来の方法ではシステムが singular になってしまうために適用できなかった、行列の多重固有値に対する精度保証を実現したものである。 J.J.Dongarra, C.B.Moler, J.H.Wilkinson による論文 [1] では、二重固有値に対する精度保証を中心に書かれているが、 Rump は一般に n 重固有値に対して適用可能な方法を与え、更に実際の計算に対して、大幅なコスト削減を実現している。

第 3 節では、楕円型固有値問題の単純固有値に対する精度保証について述べる。本論文の目的は、楕円型固有値問題の多重固有値に対する精度保証の実現であるが、基本的な固有値問題の検証の手順を明らかにするために、単純固有値についてもふれておく。この節で述べるアルゴリズムは、多重固有値に対しては適用できない。なぜならば、システムが singular になるからである。

そこで、第 4 節では多重固有値に対しても適用できる新たなアルゴリズムを構築する。これは、行列の固有値問題において、固有値、固有ベクトルそのものを包み込むのではなく、固有値と、それに対する不変部分空間の基底を包み込むという Rump の考え方を楕円型固有値問題に対しても適用しようとするものである。すなわち n 重固有値に対して、

$$(-\Delta + q)(y_1, y_2, \ldots, y_n) = (y_1, y_2, \ldots, y_n) \begin{pmatrix} m_{11} & m_{12} & \ldots & m_{1n} \\ m_{21} & m_{22} & \ldots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \ldots & m_{nn} \end{pmatrix}$$

であるような $Y = (y_i)_{1 \le i \le n}$ 及び、 $M = (m_{ij})_{1 \le i,j \le n}$ を精度保証付きで求めることを考える。このシステムに対する解の包み込みは、不動点定式化を用いて、 Banach の不動点定理の成立条件を計算機内で数値的に検証することによって行っている。更に検証付きで求めた M に Gerschgorin の定理を適用して目的の固有値を包み込んでいる。

最後に、第 5 節で結果の考察と今後の検証課題を述べる。

2　行列の固有値に対する Rump の方法

この節では、 Rump[5] による行列の多重固有値及び、近接固有値に対する精度保証付き数値計算についての概要を述べる。

2.1　記号の定義

$\mathbf{K}$ は $\mathbf{R}$ もしくは、 $\mathbf{C}$ とする。また、 $A \in M_n(\mathbf{K})$ を $n \times n$ の行列とし、 $\tilde{X} \in M_{n,k}(\mathbf{K})$ を $A\tilde{X} \approx \tilde{\lambda}\tilde{X}$ （ $\tilde{\lambda}$ は多重または近接固有値の近似値 ）であるような近似不変部分空間の基底とする。以下、 n を行列の次元、 k を不変部分空間の次元を表すとする。

さらに、行列 U と V を以下のように定義する。

$U \in M_{n,n-k}(\mathbf{K})$ は単位行列からある k 個の列を取り去ったものとし、 $V \in M_{n,k}(\mathbf{K})$ は U で取り去ったある k 個の列以外の単位行列の $n-k$ 列で構成される行列とする。これらは $UU^T + VV^T = I_n$ を満たし、 $V^T\tilde{X}$ は $\tilde{X}$ の normalizing part を表す。

$n = 5,\ k = 2$ のときの U, V の例：

$$U = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ V = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

また、 $\mathbf{K}$ 上の n 次元区間ベクトル全体を $\mathbf{IK}^n$、 $p \times q$ 区間行列を $\mathbf{I}M_{p,q}(\mathbf{K})$ と表記することにする。

2.2　行列の多重固有値及び、近接固有値に対する精度保証

ここでは、行列の多重固有値及び、近接固有値に対する精度保証付き数値計算の定式化を述べる。

多重固有値に対し、 Rump[6] によって与えられた単一の正規化条件付きの固有値問題

$$\begin{cases} Ax &= \lambda x \\ x_i &= \xi \end{cases} \tag{1}$$

に対するアルゴリズムは適用できない。なぜならば (1) を、 λ と x とに関する連立非線型方程式として Newton-like method を適用しようとすると、システムが singular になるからである。

そこで考える問題を次の形に拡大し、

$$\begin{cases} AY &= YM \\ V^TY &= V^T\tilde{X} \end{cases} \tag{2}$$

固有値に対する不変部分空間の包み込みを目指すことを考える。ここで、 $V^TY = V^T\tilde{X}$ は上の固有値問題 (1) における拘束条件 $x_i = \xi$ にあたり、 $k \times k$ 個の要素を拘束している。

この問題を考えるに当たって、必要な次の命題をあげておく。

Lemma 2.1. $\mathbf{Z}, \mathbf{X} \in \mathbf{IK}^n,\ \mathbf{C} \in \mathbf{I}M_n(\mathbf{K})$ が与えられたとする。このとき、

$$\mathbf{Z} + \mathbf{C} \cdot \mathbf{X} \subseteq \mathrm{int}(\mathbf{X}) \tag{3}$$

であれば、各 $C \in \mathbf{C}$ は convergent である。すなわち、 $\rho(C) < 1$

この命題と Brouwer の不動点定理から、次の定理が導かれる。

Theorem 2.2. $A \in M_n(\mathbf{K}), \tilde{X} \in M_{n,k}(\mathbf{K}), R \in M_n(\mathbf{K})$, そして $\mathbf{X} \in \mathbf{I}M_{n,k}(\mathbf{K})$ とする。また、U, V を 2.1 で定義したものとする。今、

$$f(\mathbf{X}) := -R(A\tilde{X} - \tilde{\lambda}\tilde{X}) + \{I - R((A - \tilde{\lambda}I)UU^T - (\tilde{X} + UU^T\mathbf{X})V^T)\}\mathbf{X} \tag{4}$$

とし、

$$f(\mathbf{X}) \subseteq \mathrm{int}(\mathbf{X}) \tag{5}$$

であるならば、このとき

$$A\hat{Y} = \hat{Y}\hat{M} \tag{6}$$

であるような $\hat{M} \in \tilde{\lambda}I_k + V^T\mathbf{X} \subset M_k(\mathbf{K}),\ \hat{Y} \in \tilde{X} + UU^T\mathbf{X} \subset M_{n,k}(\mathbf{K})$ が存在する。

上記の定理において、 $\tilde{\lambda}I_k + V^T\mathbf{X}$ の k 個の固有値は、 A のある k 個の固有値と一致する。また、$\tilde{X} + UU^T\mathbf{X}$ の各列ベクトルはそれら k 個の固有値に対する不変部分空間の基底をなす。この定理において区間ベクトル $\mathbf{X}$ を $\tilde{\lambda}, \tilde{X}$ からの変化分として求めているのは、区間計算に対する計算の効率化のためである。

ここで、 $U^T\mathbf{X}$ のなかに $V^TY = V^T\tilde{X}$ として拘束した $k \times k$ 個の要素以外の不変部分空間の基底ベクトルの成分に対応変化分が現れ、また $V^T\mathbf{X}$ のなかに M の $\tilde{\lambda}I_k$ からの変化分が現れていることに注意されたい。このように、解行列 $\mathbf{X}$ の中に $Y,\ M$ の要素を同時に包み込むようなアルゴリズムを考えることにより、 (2) に対し、直接、非線型方程式に対する精度保証のアルゴリズムを適用するよりも、計算時間やメモリ消費量の点で大幅なコスト削減を実現している。(直接、非線型方程式のアルゴリズムを適用する場合は、 $n^2 \times nk$ 次元の大規模な行列を扱うことになるが、 Rump の方法では n 次元の行列を扱うだけですむ。)

さて、以下では $\tilde{X}, \tilde{\lambda}, R$ の具体的な値の決定方法について述べよう。これらには a-priori な仮定がないことに注意する。唯一の仮定が (5) である。特に近似値に関する a-priori な条件はない。よって $A\tilde{X} \approx \tilde{\lambda}\tilde{X}$ なる $\tilde{X}, \tilde{\lambda}$ は適当なアルゴリズムを用いることによって求めればよい。(我々は **MATLAB** の近似固有値及び、固有ベクトルを求めるコマンド "**eig**"を使用することによって計算した。)

行列 R に関しては、以下のようにとる。

$$R \approx ((A - \tilde{\lambda}I)UU^T - \tilde{X}V^T)^{-1} \tag{7}$$

これは、 $H(X) := A(\tilde{X} + UU^TX) - (\tilde{X} + UU^TX)(\tilde{\lambda}I_k + V^TX)$ として Fr'echet 微分 $H'(X)$ の 0 における逆行列になっている。

次に精度保証を実現するために、区間行列 $\mathbf{X}$ を求めるための区間反復を定義する。基本的には、区間反復は $\mathbf{X}^{v+1} = f(\mathbf{X}^v)$ で与えられる。しかしながら、 $\mathbf{X}$ が、 (5) を満たすようにするために

は、 $\mathbf{X}$ にある程度の幅が必要である。このことを達成する適切な方法がいわゆる epsilon-inflation である。これは具体的には、次のアルゴリズムで与えられる。

Algorithm 2.3.

$$
\begin{aligned}
&R = ((A - \tilde{\lambda}I)UU^T - \tilde{X}V^T)^{-1};\\
&\mathbf{Z} = -R(A\tilde{\mathbf{X}} - \tilde{\lambda}\tilde{\mathbf{X}});\\
&\mathbf{C} = I - R((A - \tilde{\lambda}\mathbf{I})UU^T - \tilde{X}V^T);\\
&\mathbf{X} = \mathbf{Z};\ m = 0;\ m\max = 10;\\
&repeat\\
&\quad m = m + 1;\\
&\quad \mathbf{Y} = \mathbf{X} + 0.1 \cdot [-|\mathbf{Z}| - \epsilon, |\mathbf{Z}| + \epsilon];\\
&\quad \mathbf{X} = \mathbf{Z} + \mathbf{CY} + R(UU^T\mathbf{Y}V^T\mathbf{Y});\\
&\quad \text{ready} = (\mathbf{X} \subseteq \text{int}(\mathbf{Y}));\\
&until \text{ ready } or\ (m = m\max);\\
&if \text{ ready}\\
&\quad for\ i = 1 : k\\
&\qquad r = \sum_{j \neq i} |(V^T\mathbf{X})_{ij}|;\\
&\qquad \mathbf{L}_i = \text{complex}(\tilde{\lambda} + (V^T\mathbf{X})_i i + [-r, r]) \subseteq \mathbf{C}
\end{aligned}
$$

以上のアルゴリズムにおいて、 $\tilde{\mathbf{X}}$ は $\tilde{X}$ を包み込むような小さな区間行列、 $\mathbf{I}$ は単位行列 I を包み込むような小さな区間行列を表す。また、反復の初期値 $\mathbf{X}^0$ としては、 $-R(A\tilde{X} - \tilde{\lambda}\tilde{X})$ を包み込むような小さな区間をとることになる。 ϵ はマシンイプシロン (計算機で表現できる最小の正の浮動小数点数) を表す。

各 $\mathbf{L}_i$ は行列 $\mathbf{M} := \tilde{\lambda}I_k + V^T\mathbf{X}$ に対する Gerschgorin Circle を計算している。故に Gerschgorin の定理から、 $\cup_{i=1}^{k}\mathbf{L}_i$ の中に A の k 個の多重または、近接固有値が包まれる。この $\mathbf{L}_i$ は非対称行列の場合、複素区間として計算しなければならない。なぜならその固有値が実かどうかという問題は数値的に決定不可能な問題だからである。

ここで、 normalizing part の選択 (すなわち U, V の選択) の仕方は固定されてはいない。通常は $V^T\tilde{X}$ のスペクトルノルムが最大となるように V を定める。

以上のように多重または、近接固有値の包み込み $\cup_{i=1}^{k}\mathbf{L}_i$ 及び、それに伴う不変部分空間の基底の包み込み $\tilde{X} + UU^T\mathbf{X}$ を得ることができる。

2.3 一般化固有値問題に対する精度保証の適用

2.2 節の方法は、一般化固有値問題： $Ax = \lambda Bx$ に対して拡張することができる。 $B \in M_n(\mathbf{K})$ を $A\tilde{X} \approx \tilde{\lambda}B\tilde{X}$ なる行列とする。このとき、 Theorem 2.2. を以下のように修正することにより一般化固有値問題に対する精度保証を導く。

Theorem 2.4. $A, B \in M_n(\mathbf{K}), \tilde{\lambda} \in \mathbf{K}, R \in M_n(\mathbf{K})$ そして $\mathbf{X} \in \mathbf{I}M_{n,k}(\mathbf{K})$ が与えられたとする。さらに U, V は上で定義されたものとする。

$$
f(X) = -R(A\tilde{X} - \tilde{\lambda}B\tilde{X}) + \{I - R((A - \tilde{\lambda}B)UU^T - B(\tilde{X} + UU^TX)V^T)\}X \tag{8}
$$

としたとき、

$$f(\mathbf{X}) \subseteq \mathrm{int}\mathbf{X}$$

ならば、

$$A\hat{Y} = B\hat{Y}\hat{M} \tag{9}$$

であるような、 $\hat{M} \in \tilde{\lambda} I_k + V^T\mathbf{X} \subset M_k(\mathbf{K}),\ \hat{Y} \in \tilde{X} + UU^T\mathbf{X} \subset M_{n,k}(\mathbf{K})$ が存在する。

ここで、

$$R \approx ((A - \tilde{\lambda}B)UU^T - B\tilde{X}V^T)^{-1} \tag{10}$$

と選ぶ。

2.4 疎行列に対する精度保証の適用

$((A - \tilde{\lambda}I)UU^T - \tilde{X}V^T)$ の近似逆行列 R は、一般的には密行列になるので、 A が疎行列であるとき、今までの議論を適用するのはあまりよい方法とは言えない。そこで、逆行列をとらないような Theorem 2.2. や Theorem 2.4. の変形を考える。

$G := ((A - \tilde{\lambda}I)UU^T - \tilde{X}V^T)$ と定義し、 G が可逆であるとして $R := G^{-1}$ とする。このとき

$$I - R((A - \tilde{\lambda}I)UU^T - (\tilde{X} + UU^TX)V^T) = RUU^TXV^T \tag{11}$$

これを、 (5) に代入すると

$$f(X) = R(-A\tilde{X} + \tilde{\lambda}\tilde{X} + UU^TXV^TX) \tag{12}$$

このとき、条件 (4) は、 G が可逆で、かつ各線型方程式

$$Gx = b,\ b \in -A\tilde{X} + \tilde{\lambda}\tilde{X} + UU^T\mathbf{X}V^T\mathbf{X} \tag{13}$$

の解が $\mathbf{X}$ の内点に包まれることと同値である。これから次の結果が得られる。

Theorem 2.5. Theorem 2.2. と同じ記号を用い、 $G := ((A - \tilde{\lambda}I)UU^T - \tilde{X}V^T)$ とする。

$$f(\mathbf{X}) := \{y \mid \forall b \in -A\tilde{X} + \tilde{\lambda}\tilde{X} + UU^T\mathbf{X}V^T\mathbf{X}\text{かつ}Gy = b\} \tag{14}$$

とする。このとき、

$$f(\mathbf{X}) \subseteq \mathrm{int}\mathbf{X}$$

ならば、

$$A\hat{Y} = \hat{Y}\hat{M}$$

であるような $\hat{M} \in \tilde{\lambda} I_k + V^T\mathbf{X} \subset M_k(\mathbf{K}),\ \hat{Y} \in \tilde{X} + UU^T\mathbf{X} \subset M_{n,k}(\mathbf{K})$ が存在する。

この方法は、 $(A - \tilde{\lambda}I)UU^T - \tilde{X}V^T$ の逆行列を直接求めないことから、疎行列の疎性を保つ点で、Theorem 2.2. に比べて有利である。

2.5 欠陥のある固有値に対する考察

ここでは、非対称行列で、特に欠陥のある固有値に対する精度保証を述べる。欠陥のある固有値とは幾何的多重度と代数的多重度が異なるような多重固有値のことをいう。

さて、欠陥のある固有値に対して Rump のアルゴリズムを適用するとあまりいい結果は得られなかった。これは $G := ((A - \tilde{\lambda}I)UU^T - \tilde{X}V^T)$ のもつ特性に起因すると考えられる。この行列の構成の仕方により、k 重固有値に伴う k 個の固有ベクトルが G のある k 個の列に現れる。ところが欠陥のある固有値の場合、幾何的多重度を γ、代数的多重度を μ とすると、$\gamma < \mu$ となるので、k 重固有値に対し、それに伴う一次独立な固有ベクトルの個数が k よりも少なくなる。G を構成するためには、少ない分を埋めるために一次従属なベクトルを持ってくることが考えられるが、このとき G は singular となり、それは $R \approx G^{-1}$ を構成できないことを意味する。

そこで、少ない分を埋めるために持ってきた一次従属なベクトルにわずかな摂動を与えて、強制的に G を non-singular にすることを考える。こうすることにより、計算機上では G^{-1} を計算できるようにはなる。但し、ほとんど singular に近いので、非常に不安定であることを避けることはできない。また、欠陥のある固有値をもつ行列に対し、**MATLAB** における **eig** コマンドを適用すると、本来なら対角化はできないが、これを近似的にに対角化するので、このときに出てきた近似固有ベクトルを採用することも 1 つの方法である。

以上の方法をとることにより、一応は Rump のアルゴリズムを適用することはできるが、やはり k が大きいと、どうしても R が不安定であることが原因で検証に失敗する場合が多くなる。

欠陥固有値に対する精度保証の失敗例

行列は以下のようにとった。

$$J = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

とし、さらに W : 5×5 正方行列をランダムに生成して、

$$A = WJW^{-1}$$

と定める。このとき、行列 A は代数的多重度 4、幾何的多重度 1 の欠陥固有値'1' を持つ。この行列 A に対して、精度保証アルゴリズムを適用した。しかし、結果は多くの場合が'Failed' であった。この理由としては、システムがほとんど singular になっていることが原因だと思われる。

そこで、拘束条件 $V^TY = V^T\tilde{X}$ を以下のように内積を用いた拘束条件に変えて、考察を行った。

$$\begin{cases} AY &= YM \\ Y^T\tilde{X} &= I_k \end{cases} \tag{15}$$

これは、Y の形を内積の値を固定することによって拘束したものである。これは $V^TY = V^T\tilde{X}$ とするよりも弱い拘束条件である。この $(n \times k + k \times k)$ 元の非線型方程式を非線型方程式に対する精度保証付き数値計算のアルゴリズムを適用することによって解く。

この方法でいくつかの場合の精度保証には成功したが、 Rump の方法を超えるような結果は得られず、さらに Rump の方法とは違ってシステムが非常に大きくなるために、計算におけるコストの点で不利という結論になった。

以上のように欠陥のある固有値に対する精度保証は非常に困難に思われる。 Rump は欠陥のある固有値に対する問題は、本質的に ill-conditioned な問題であり、適切なアルゴリズムを見出すことは非常に難しいと主張しており、今後の課題であろう。

3　楕円型固有値問題の単純固有値

以下では [2], [3] に基づいた楕円型作用素の単純固有値の精度保証について、その概要を述べる。

3.1　問題と記号の定義

ここでは、以下のような固有値問題

$$\left\{ \begin{array}{rcll} -\Delta u + qu &=& \lambda u & \text{in } \Omega \\ u &=& 0 & \text{on } \partial\Omega \end{array} \right. \tag{16}$$

を考える。但し、 Ω は $\mathbf{R}^2$ 上の有界凸領域、 $q \in L^\infty(\Omega)$ とする。

以下、記号の定義を行う。
$H^m(\Omega)$ を

$$H^m(\Omega) := \{v \mid D_\alpha v \in L^2(\Omega);\ \forall \alpha \text{ s.t. } |\alpha| \le m\} \tag{17}$$

とする。但し、 $\alpha = (\alpha_1, \alpha_2) \in \mathbf{Z}_{\ge 0} \times \mathbf{Z}_{\ge 0}$ で、

$$|\alpha| \equiv \alpha_1 + \alpha_2,\ D^\alpha v \equiv \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2}}$$

である。
また、

$$H_0^1(\Omega) \equiv \{v \in H^1(\Omega) \mid v = 0 \text{ on } \partial\Omega\} \tag{18}$$

とし、 $H_0^1(\Omega)$ の内積を、 $L^2(\Omega)$ の内積 $(\cdot,\cdot)_{L^2(\Omega)}$ を用いて

$$< u, v >_{H_0^1(\Omega)} \equiv (\nabla u, \nabla v)_{L^2} \quad u, v \in H_0^1(\Omega) \tag{19}$$

と定義する。
さらに、 $H_0^1(\Omega) \times \mathbf{R}$ の内積 $< \cdot,\cdot >_{H_0^1(\Omega)\times\mathbf{R}}$ 及び、ノルム $\|\cdot\|_{H_0^1(\Omega)\times\mathbf{R}}$ を以下のように定義する。
但し、 $w_i = (u_i, \lambda_i) \in H_0^1(\Omega) \times \mathbf{R}$ $(i = 1, 2)$ とする。

$$< w_1, w_2 >_{H_0^1(\Omega)\times\mathbf{R}} \equiv (\nabla w_1, \nabla w_2)_{L^2} + \lambda_1\lambda_2 \tag{20}$$

$$\|w_i\|_{H_0^1(\Omega)\times\mathbf{R}} \equiv (\|u_i\|^2_{H_0^1(\Omega)} + |\lambda_i|^2)^{\frac{1}{2}} \tag{21}$$

次に以下の記号を定義しておく。
S_h : パラメータ h に従属するような $H_0^1(\Omega)$ の有限次元部分空間 (h は通常、有限要素のメッシュサイズに対応する。)
$S_h^\perp$: S_h の直交補空間。すなわち、
　　$S_h^\perp = \{\phi \in H_0^1(\Omega) \mid (\nabla\phi, \nabla\psi)_{L^2} = 0,\ \forall\psi \in S_h\}$
P_{h0} : $H_0^1(\Omega)$ から S_h への H_0^1-projection. すなわち、
　　任意の $u \in H_0^1(\Omega)$ に対して、 $(\nabla(u - P_{h0}u), \nabla v)_{L^2} = 0,\ \forall v \in S_h$

また、 projection P_h : $H_0^1(\Omega) \times \mathbf{R} \to S_h \times \mathbf{R}$ を

$$P_h(u, \lambda) \equiv (P_{h0}u, \lambda)$$

で定義する。

3.2 不動点定式化

まずは問題 (16) を次のように正規化する。

$$\begin{aligned} &\text{find } (\hat{u},\lambda)\in H_0^1(\Omega)\times\mathbf{R} \text{ s.t.}\\ &\left\{\begin{array}{rcl} -\Delta\hat{u}+(q-\lambda)\hat{u} &=& 0\\ \int_\Omega \hat{u}^2dx &=& 1 \end{array}\right. \end{aligned} \tag{22}$$

ここで、以下の議論に必要な、定数に関する仮定と命題を挙げる。

Assumption 3.1. $u\in H^2(\Omega)\cap H_0^1(\Omega)$ に対して、次を満たす h に依存しない、具体的に数値として与えられる定数 C_1 が存在する。([4] を参照)

$$\inf_{\chi\in S_h}\|u-\chi\|_{H_0^1(\Omega)}\le C_1h|u|_{H^2(\Omega)} \tag{23}$$

ここで、$|u|_{H^2(\Omega)}$ は次で定義される $H^2(\Omega)$ 上のセミノルムを表す。

$$|u|^2_{H^2(\Omega)}\equiv\sum_{i,j=1}^{2}\|\frac{\partial^2u}{\partial x_i\partial x_j}\|^2_{L^2(\Omega)} \tag{24}$$

Lemma 3.2. $\forall\psi\in L^2(\Omega)$ に対して、$\exists!\phi\in H^2(\Omega)\cap H_0^1(\Omega)$ s.t.

$$\left\{\begin{array}{rcll} -\Delta\phi &=& \psi & \text{in } \Omega\\ \phi &=& 0 & \text{on } \partial\Omega \end{array}\right. \tag{25}$$

かつ、

$$|\phi|_{H^2(\Omega)}\le C_2\|\psi\|_{L^2(\Omega)} \tag{26}$$

を満たす、h に従属しない定数 C_2 が存在する。(実際には $C_2=1$ とする。[4] を参照)

ここで、$\hat{w}_h=(\hat{u}_h,\hat{\lambda}_h)$ を

$$\left\{\begin{array}{rcl} (\nabla\hat{u}_h,\nabla\phi_i)_{L^2(\Omega)} &=& ((\hat{\lambda}_h-q)\hat{u}_h,\phi_h)_{L^2(\Omega)}\ \forall\phi_h\in S_h\\ \int_\Omega\hat{u}_h^2dx &=& 1 \end{array}\right. \tag{27}$$

つまり、(22) の有限要素近似解とする。以下ではこの近似解の近傍に、真の固有値及び、固有関数が存在することを精度保証付き数値計算で検証する。このとき、行列の固有値に関する精度保証でしたように、真の固有値及び、固有関数を直接包むのではなく、近似解 $\hat{w}_h$ と真の解との差を包む込むようにする。この近似解との差の取り方に関しては $P_{h0}\bar{u}=\hat{u}_h$ となる $(\bar{u},\hat{\lambda}_h)$ と真の解との差を考える。こうすることにより特別な方法をとることができて、誤差評価がよくなることが知られている。よって、この $(\bar{u},\hat{\lambda}_h)$ の近傍に、(22) の真の解 $(\hat{u},\lambda)$ を探すことにする。

さて、

$$\left\{\begin{array}{rcll} -\Delta\bar{u}+(q-\hat{\lambda}_h)\hat{u}_h &=& 0 & \text{in } \Omega\\ \bar{u} &=& 0 & \text{on } \partial\Omega \end{array}\right. \tag{28}$$

を満たす $\bar{u}\in H^2(\Omega)\cap H_0^1(\Omega)$ 及び $\lambda\in\mathbf{R}$ を考えよう。このとき、

$$\begin{aligned} (\nabla\hat{u}_h-\nabla\bar{u},\nabla\phi_h)_{L^2(\Omega)} &= (\nabla\hat{u}_h,\nabla\phi_h)_{L^2(\Omega)}-(\nabla\bar{u},\nabla\phi_h)_{L^2(\Omega)}\\ &= ((\hat{\lambda}-q)\hat{u}_h,\phi_h)_{L^2(\Omega)}-((\hat{\lambda}-q)\hat{u}_h,\phi_h)_{L^2(\Omega)}\\ &= 0 \end{aligned}$$

故に、 $\hat{u}_h = P_{h0}\bar{u}$ となり、さらに

$$P_h(\bar{u}, \hat{\lambda}_h) = (P_{h0}\hat{u}, \hat{\lambda}_h) = (\hat{u}_h, \hat{\lambda}_h) = \hat{w}_h$$

となる。

(22)、 (28) より

$$\left\{ \begin{array}{rcl} -\Delta(\hat{u} - \bar{u}) & = & (\lambda - q)\hat{u} - (\hat{\lambda}_h - q)\hat{u}_h \\ \int_\Omega \hat{u}^2 dx & = & 1 \end{array} \right. \tag{29}$$

ここで、 $v_0 = \bar{u} - \hat{u}_h$ とおくと、 $v_0 = \bar{u} - P_{h0}\bar{u} = (I - P_{h0})\bar{u} \in S_h^\perp$ であるから、

$$\bar{u} = \hat{u}_h + v_0, \quad \hat{u}_h \in S_h,\ v_0 \in S_h^\perp$$

と表せる。

次に v_0 のノルムを評価する。 (23)、 (26)、 (28) から

$$\begin{array}{rcll} \|v_0\|_{H_0^1(\Omega)} & = & \|\bar{u} - P_{h0}\bar{u}\|_{H_0^1(\Omega)} & \\ & \leq & C_1 h |\hat{u}|_{H^2(\Omega)} & \\ & \leq & C_1 C_2 h \|(\hat{\lambda}_h - q)\hat{u}_h\|_{L^2(\Omega)} & \\ & = & C_0 h \|(\hat{\lambda}_h - q)\hat{u}_h\|_{L^2(\Omega)} & (C_0 \equiv C_1 C_2) \end{array} \tag{30}$$

と評価できる。また L^2- ノルムに関しては、次のように評価できる。

$$\|v_0\|_{L^2(\Omega)} \leq C_0^2 h^2 \|(\hat{\lambda}_h - q)\hat{u}_h\|_{L^2(\Omega)} \tag{31}$$

これは、 Aubin-Nitsche's trick を用いることによって導かれるものである。

ここで、

$$\hat{u} = \bar{u} + \tilde{u},\ \lambda = \hat{\lambda}_h + \tilde{\lambda}$$

とおく。つまり、近似解 $(\bar{u}, \hat{\lambda}_h)$ と真の解 $(\hat{u}, \lambda)$ との差の部分である $(\tilde{u}, \tilde{\lambda})$ に関して包み込みを行うことを考えると、 (28) は、 $(\tilde{u}, \tilde{\lambda})$ に関して以下のように書き直される。

$$\left\{ \begin{array}{rcl} -\Delta\tilde{u} & = & (\hat{\lambda}_h + \tilde{\lambda} - q)(\tilde{u} + \hat{u}_h + v_0) - (\hat{\lambda}_h - q)\hat{u}_h \\ \int_\Omega (\tilde{u} + \hat{u}_h + v_0)^2 dx & = & 1 \end{array} \right. \tag{32}$$

ここで、 $F : H_0^1(\Omega) \times \mathbf{R} \to H_0^1(\Omega) \times \mathbf{R}$ を

$$F : \left(\begin{array}{c} \tilde{u} \\ \tilde{\lambda} \end{array} \right) \equiv \left(\begin{array}{c} (-\Delta)^{-1}\{(\hat{\lambda}_h + \tilde{\lambda} - q)(\tilde{u} + \hat{u}_h + v_0) - (\hat{\lambda}_h - q)\hat{u}_h\} \\ \tilde{\lambda} + \int_\Omega (\tilde{u} + \hat{u}_h + v_0)^2 dx - 1 \end{array} \right) \tag{33}$$

と定義すると、 (28) は $w = (\tilde{u}, \tilde{\lambda})$ について、不動点方程式

$$w = F(w) \tag{34}$$

の形で書くことができる。

3.3 検証条件とアルゴリズム

まずは次を仮定する。

Assumption 3.3. $\rho \equiv (-v_0, 0)$ とおく。 F の ρ における Fréchet 微分 $F'(\rho)$ が存在するとき、

$$P_h[I - F'(\rho)] \ : \ H_0^1(\Omega) \times \mathbf{R} \to S_h \times \mathbf{R}$$

を $S_h \times \mathbf{R}$ に制限したものは、逆作用素

$$[I - F'(\rho)]_h^{-1} \ : \ S_h \times \mathbf{R} \to S_h \times \mathbf{R}$$

を持つ。

ここで、 (34) を考える。 (34) は有限次元部分と無限次元部分とを次のように表せる。

$$\begin{cases} P_h w &= P_h F(w) \\ (I - P_h)w &= (I - P_h)F(w) \end{cases} \tag{35}$$

(35) のうち、 $P_h w = P_h F(w)$ のほうにのみ、 Newton-like method を適用する。つまり、

$$N_h(w) \equiv P_h w - [I - F'(\rho)]_h^{-1}(P_h w - P_h F(w)) \tag{36}$$

として、

$$T(w) \equiv N_h(w) + (I - P_h)F(w) \tag{37}$$

とすると、 T はコンパクト作用素となり、次の命題が成り立つ。

Proposition 3.4.

$$w = T(w) \iff w = F(w) \tag{38}$$

以上の命題により、 $T(W) \subset W$ なる $W \in H_0^1(\Omega) \times \mathbf{R}$: 有界凸閉集合 が存在すれば Schauder の不動点定理により

$$\exists w \in W \ s.t. \ w = T(w) \ (\text{i.e. } w = F(w))$$

となる。 $W = W_h \oplus W_\perp$, $W_h \subset S_h \times \mathbf{R}$, $W_\perp \subset S_h^\perp \times \{0\}$ とすれば、検証条件は以下のように表すことができる。

$$\begin{cases} N_h(W) \subset W_h \\ (I - P_h)F(W) \subset W_\perp \end{cases} \tag{39}$$

次に計算機における精度保証のアルゴリズムについて述べる。まず、

$$\{\phi_j\}_{j=0}^{M-1} \ : \ S_h\text{の基底}$$

$$\chi_h \equiv \{(\sum_{j=0}^{M-1} A_j \phi_j, A_M) \mid A_j \subset \mathbf{IR}\}$$

とおく。
また、 $\alpha \in \mathbf{R}^+$ に対して

$$[\alpha] \equiv \{\phi \in S_h^\perp \mid \|\phi\|_{H_0^1(\Omega)} \leq \alpha\} \times \{0\}$$

とする。
この χ_h と $[\alpha]$ の元に対して、反復列 $W_{(n)} \equiv w_h^{(n)} \oplus [\alpha_n]$ を考える。

Algorithm 3.5.

$n = 0$ のとき

$$w_h^{(0)} = \{0\},\ \alpha_0 = 0 \text{ i.e. } W_{(0)} = \{0\}$$

$n \geq 1$ のとき、 $0 < \delta \ll 1$ なる δ に対して

$$\left\{ \begin{array}{rcl} \tilde{w}_h^{(n-1)} & = & w_h^{(n-1)} + (\sum_{j=0}^{M-1}[-1,1]\delta\phi_j, [-1,1]\delta) \\ \tilde{\alpha}_{n-1} & = & \alpha_{n-1} + \delta \end{array} \right. \tag{40}$$

とし $\tilde{W}^{(n-1)} = \tilde{w}_h^{n-1} \oplus [\tilde{\alpha}_{n-1}]$ に対して、 $w_h^{(n)},\ \alpha_n$ を次のように定義する。

$$\left\{ \begin{array}{rcl} w_h^{(n)} & \supset & N_h(\tilde{W}^{(n-1)}) \\ \alpha_n & = & C_0 h \sup_{v=(\hat{v},\mu)\in\tilde{W}^{(n-1)}} \|(\tilde{\lambda}_h + \mu - q)(\hat{v} + v_0 + \hat{u}_h) - (\hat{\lambda}_h - q)\hat{u}_h\|_{L^2(\Omega)} \end{array} \right. \tag{41}$$

ここで、 $C_0 \equiv C_1 C_2$ である。

以上のアルゴリズムに対し、以下の定理が解の存在を保証する。

Teorem 3.6. $\exists N \in \mathbf{N}$ s.t. $w_h^{(N)} \subset \tilde{w}_h^{(N-1)},\ \alpha_N \leq \tilde{\alpha}_{N-1}$
$\Rightarrow \exists w \in w_h^{(N)} \oplus [\tilde{\alpha}_{N-1}]$ s.t. $w = F(w)$

さて、次は $\tilde{W}^{(n-1)} = \tilde{w}_h^{(n-1)} \oplus [\tilde{\alpha}_{n-1}]$ に対して、 (41) を満たす $w_h^{(n)},\ \alpha_n$ を具体的に求める方法を示す。
まずは

$$w_h^{(n)} = (\sum_{j=0}^{M-1} B_j^{(n)}\phi_j,\ B_M^{(n)}) \supset N_h(\tilde{W}^{(n-1)}) \tag{42}$$

を満たす $B_i^{(n)} \in \mathbf{IR},\ (i = 0, 1, \ldots, M)$ を求めることを考える。

$$w_h^{(n-1)} = (\sum_{j=0}^{M-1} B_j^{(n-1)}\phi_j,\ B_M^{(n-1)}) \supset N_h(\tilde{W}^{(n-1)}) \tag{43}$$

とおいて、 $\forall v \in \tilde{W}^{(n-1)}$ を $v = \hat{w} + \hat{\alpha},\ \hat{w} \in \tilde{w}_h^{(n-1)},\ \hat{\alpha} = (\hat{\phi}, 0) \in [\tilde{\alpha}_{n-1}]$ と表すことにすると、

$$\begin{array}{rcl} N_h(v) & = & N_h(\hat{w} + \hat{\alpha}) \\ & = & P_h(\hat{w} + \hat{\alpha}) - [I - F'(\rho)]_h^{-1}(P_h(\hat{w} + \hat{\alpha}) - P_h F(\hat{w} + \hat{\alpha})) \\ & = & \hat{w} - [I - F'(\rho)]_h^{-1}(\hat{w} - P_h F(\hat{w} + \hat{\alpha}) \end{array} \tag{44}$$

であるから、これを整理すると

$$[I - F'(\rho)]_h N_h(v) = P_h F(\hat{w} + \hat{\alpha}) - P_h F'(\rho)\hat{w} \tag{45}$$

となる。

ここで、(42) を満たすような $B_i^{(n)} \in \mathbf{IR},\ (i = 0, 1, \ldots, M)$ を求めるために、以下の集合 X を考えることにする。

$$\begin{aligned} X \equiv\ & \{x \in H_0^1(\Omega) \times \mathbf{R} \mid \\ & ([I - F'(\rho)]_h x) \odot (\phi_i, 1) = (P_h F(\hat{w} + \hat{\alpha}) - P_h F'(\rho)\hat{w}) \odot (\phi_i, 1), \\ & \forall \hat{w} \in \tilde{w}_h^{(n-1)} \text{ and } \hat{\alpha} \in [\tilde{\alpha}_{n-1}],\ i = 0, \ldots, M-1\} \end{aligned} \tag{46}$$

但し、上の X の定義において、$w_i = (u_i, \lambda_i) \in H_0^1(\Omega) \times \mathbf{R}\ (i = 1, 2)$ に対して、演算子 $\odot$ を次で定義するものとする。

$$w_1 \odot w_2 \equiv (\int_\Omega \nabla u_1 \cdot \nabla u_2 dx, \lambda_1 \lambda_2) \tag{47}$$

実際に計算機を用いて計算するときは、(47) の右辺を区間として解くことになるので、$X \in \mathbf{X}$ なる区間包 $\mathbf{X}$ を得る。よって、

$$(\sum_{j=0}^{M-1} B_j^{(n)} \phi_j,\ B_M^{(n)}) \equiv \mathbf{X} \tag{48}$$

とすることができる。以下では実際に計算機による $\mathbf{X}$ の計算法を示す。

$\phi_i\ (0 \le i \le M-1)$ と、$x \equiv (\sum_{j=0}^{M-1} x_j^{(n)} \phi_j, x_M^{(n)})$ に対して、

$$\begin{aligned} ([I - F'(\rho)]_h x) \odot (\phi_i, 1) =\ & \textstyle\sum_{j=0}^{M-1} x_j^{(n)} (\int_\Omega \nabla\phi_j \cdot \nabla\phi_i dx + \int_\Omega q\phi_j\phi_i dx \\ & -\hat{\lambda}_h \textstyle\int_\Omega \phi_j\phi_i dx,\ -2\int_\Omega \hat{u}_h\phi_j dx) \\ & -(x_M^{(n)} \textstyle\int_\Omega \hat{u}_h\phi_i dx, 0) \end{aligned} \tag{49}$$

となり、また $\hat{w} \equiv (\sum_{j=0}^{M-1} b_j^{(n-1)} \phi_j, x_M^{(n-1)})$ に対して、

$$\begin{aligned} & (P_h F(\hat{w} + \hat{\alpha} - F'(\rho)\hat{w}) \odot (\phi_i, 1) \\ = \ & (\int_\Omega \{(\hat{\lambda}_h - q)(\hat{\phi} + v_0) + b_M^{(n-1)} (\textstyle\sum_{j=0}^{M-1} b_j^{(n-1)} \phi_j + \hat{\phi} + v_0)\} \phi_i dx, \\ & \int_\Omega \{(\textstyle\sum_{j=0}^{M-1} b_j^{(n-1)} \phi_j + \hat{\phi} + v_0)^2 + 2(\hat{\phi} + v_0)\hat{u}_h\} dx) \end{aligned} \tag{50}$$

となることがわかる。そこで、$(M+1) \times (M+1)$ 行列 $G \equiv (g_{ij})_{0 \le i,j \le M}$ を

$$\begin{aligned} g_{ij} &= (\nabla\phi_i, \nabla\phi_j) + (q\phi_i, \phi_j) - \hat{\lambda}_h(\phi_i, \phi_j) \quad (0 \le i, j \le M-1) \\ g_{iM} &= -(\hat{u}_h, \phi_i) \quad (0 \le i \le M-1) \\ g_{Mj} &= -2(\hat{u}_h, \phi_j) \quad (0 \le j \le M-1) \\ g_{MM} &= 0 \end{aligned}$$

と定義し、さらに $(M+1)$ 次の区間ベクトル $K^{(n-1)} \equiv (K_i^{(n-1)})_{0 \le i \le M}$ を

$$\begin{aligned} K_i^{(n-1)} &= \int_\Omega \{(\hat{\lambda}_h - q)(\hat{\phi} + v_0) + B_M^{(n-1)} (\textstyle\sum_{j=0}^{M-1} B_j^{(n-1)} + \hat{\phi} + v_0)\} \phi_i dx \quad (0 \le i \le M-1) \\ K_M^{(n-1)} &= \int_\Omega \{(\textstyle\sum_{j=0}^{M-1} B_j^{(n-1)} \phi_j + \hat{\phi} + v_0)^2 + 2(\hat{\phi} + v_0)\hat{u}_h\} dx \end{aligned}$$

で定義すると、(49) 及び、(50) から、(42) における区間係数 $B_i^{(n)}$ に対し、$B^{(n)} \equiv (B_i^{(n)}) \quad (0 \le i \le M)$ として

$$GB^{(n)} = K^{(n-1)}$$

となる。また、 Assumption 3.3. により、 G は逆行列を持つので、区間係数 $B_i^{(n)}$ は

$$B^{(n)} = G^{-1}K^{(n-1)} \tag{51}$$

と決めることができる。

α_n に関しては $\|(\hat{\lambda}_h + \mu - q)(\hat{v} + v_0 + \hat{u}_h) - (\hat{\lambda}_h - q)\hat{u}_h\|_{L^2(\Omega)}$ を三角不等式などを用いて分けた後、前に挙げたした v_0 のノルムの評価などを適用することにより、決めることができる。

以上のようにして、 (16) の単純固有値及び、それに伴う固有関数を検証することができる。

4　楕円型固有値問題の多重固有値

4.1　多重固有値に対する不動点定式化

考える問題は、第 3 節と同じで

$$\left\{ \begin{array}{rcll} -\Delta u + qu &=& \lambda u & \text{in } \Omega \\ u &=& 0 & \text{on } \partial\Omega \end{array} \right.$$

である。
ところが、多重固有値に対して第 3 節の議論は適用できない。なぜなら行列の多重固有値に対する精度保証のときと同様にシステムが singular になるからである。そこで、第 2 節で考えたのと同様に、固有値とそれに対する不変部分空間の基底関数に対する検証を行うことを考える。ここで、議論の簡単のために以下では二重固有値に対するアルゴリズムを考えることにする。つまり、以下のような問題を考える。一般に n 重固有値に対しても、全く同様に適用することが可能である。

$$\begin{array}{l} \text{find } y_i \in H_0^1(\Omega) \quad (i=1,2),\ m_{ij} \in \mathbf{R} \quad (i,j=1,2) \\ \left\{ \begin{array}{rcl} (-\Delta+q)(y_1,y_2) &=& (y_1,y_2) \left(\begin{array}{cc} m_{11} & m_{12} \\ m_{21} & m_{22} \end{array} \right) \\ (y_1,\tilde{\phi}_1) &=& (\tilde{x}_1,\tilde{\phi}_1) \\ (y_2,\tilde{\phi}_1) &=& (\tilde{x}_2,\tilde{\phi}_1) \\ (y_1,\tilde{\phi}_2) &=& (\tilde{x}_1,\tilde{\phi}_2) \\ (y_2,\tilde{\phi}_2) &=& (\tilde{x}_2,\tilde{\phi}_2) \end{array} \right. \end{array} \tag{52}$$

ここで (52) の下 4 つの等式は拘束条件である。第 2 節の行列の多重固有値に対する問題における $V^TY = V^T\tilde{X}$ に相当する。ここで、 $\tilde{x}_i \quad (i=1,2)$ は

$$(-\Delta+q)(\tilde{x}_1,\tilde{x}_2) \approx \tilde{\lambda}(\tilde{x}_1,\tilde{x}_2)$$

であるようにとる。つまり、近似固有値 $\tilde{\lambda}$ に伴う近似固有関数である。この $\tilde{x}_i \quad (i=1,2)$ の求め方に関しては後述する。また、 $\tilde{\phi}_i \quad (i=1,2)$ は一次独立な関数で、その定義についても後で述べる。

さて、以下ではまず、 (52) の Fourier-Galerkin 近似解を求める。ここで、自然数 $M,\ N$ に対して、S_h を以下の基底関数によって張られる $H_0^1(\Omega)$ の部分空間とする。従ってこの場合、 h をメッシュサイズに対応させるとすれば、 $h \approx \max\{1/M,\ 1/N\}$ となる。

$$\{\frac{2}{\pi}\sin mx \sin ny\}_{1\le m\le M,\ 1\le n\le N} \subset H_0^1(\Omega) \tag{53}$$

また、表現を簡単にするために、 $(M\times N)$ 個の基底関数を

$$\phi_i = \frac{2}{\pi}\sin m_i x \sin n_i y \quad (1 \le i \le MN)$$

と書くことにする。また、考える領域 Ω は

$$\Omega \equiv (0,\pi)\times(0,\pi)$$

である。

ここで、(52) の固有方程式の部分を

$$\left\{ \begin{array}{rcl} (-\Delta+q)y_1 & = & m_{11}y_1+m_{21}y_2 \\ (-\Delta+q)y_2 & = & m_{12}y_1+m_{22}y_2 \end{array} \right. \tag{54}$$

と書く。この Fourier-Galerkin 近似解を $\hat{y}_1^h,\ \hat{y}_2^h \in S_h$ 及び、$\hat{m}_{ij}^h \in \mathbf{R} \quad (1 \le i,j \le 2)$ とおいて、

$$\left\{ \begin{array}{rcl} \hat{y}_1^h & = & \sum_{j=1}^n a_j\phi_j \\ \hat{y}_2^h & = & \sum_{j=1}^n b_j\phi_j \end{array} \right. \qquad (n=MN)$$

と表し、係数 $\{a_j\}_{1\le j\le n},\ \{b_j\}_{1\le j\le n}$ を求めることにする。すなわち、

$$\left\{ \begin{array}{rcl} & (\nabla\sum_{j=1}^n a_j\phi_j, \nabla\phi_i)+(q\sum_{j=1}^n a_j\phi_j,\phi_i) & \\ = & \hat{m}_{11}^h(\sum_{j=1}^n a_j\phi_j,\phi_i)+\hat{m}_{21}^h(\sum_{j=1}^n b_j\phi_j,\phi_i) & \\ & (\nabla\sum_{j=1}^n b_j\phi_j, \nabla\phi_i)+(q\sum_{j=1}^n b_j\phi_j,\phi_i) & \\ = & \hat{m}_{12}^h(\sum_{j=1}^n a_j\phi_j,\phi_i)+\hat{m}_{22}^h(\sum_{j=1}^n b_j\phi_j,\phi_i) & \end{array} \right. \qquad (1\le i\le n) \tag{55}$$

を満たすような $\{a_j\}_{1\le j\le n},\ \{b_j\}_{1\le j\le n},\ \hat{m}_{ij}$ を求める。ここで、
$A=(A_{ij})_{1\le i,j\le n},\ A_{ij}=(\nabla\phi_j,\nabla\phi_i)$
$B=(B_{ij})_{1\le i,j\le n},\ B_{ij}=(q\phi_j,\phi_i)$
$D=(D_{ij})_{1\le i,j\le n},\ D_{ij}=(\phi_j,\phi_i)$
また、
$\mathbf{a}=(a_1,a_2,\ldots,a_n)^T$
$\mathbf{b}=(b_1,b_2,\ldots,b_n)^T$
とおくと、(55) は以下のように書き直される。

$$\left\{ \begin{array}{rcl} A\mathbf{a}+B\mathbf{a} & = & \hat{m}_{11}^h D\mathbf{a}+\hat{m}_{21}^h D\mathbf{b} \\ A\mathbf{b}+B\mathbf{b} & = & \hat{m}_{12}^h D\mathbf{a}+\hat{m}_{22}^h D\mathbf{b} \end{array} \right.$$

よって、これをまとめると

$$(A+B)Y=DYM,\ M=\left(\begin{array}{cc} \hat{m}_{11}^h & \hat{m}_{12}^h \\ \hat{m}_{21}^h & \hat{m}_{22}^h \end{array} \right),\ Y=(\mathbf{a},\mathbf{b})$$

よって、$(A+B)Y=DYM$ を満たす Y,M を求めればよい。実際に計算機を用いて計算する場合、第 2 節での議論を用いて精度保証付きで求めることも可能であるが、ここではその必要はなく、単に

$$(A+B)\tilde{X}\approx\tilde{\lambda}D\tilde{X}$$

なる $\tilde{X},\ \tilde{\lambda}$ を求め、(我々は **MATLAB** の **eig** コマンドを使用することによって求めている。)

$$Y=\tilde{X},\ M=\left(\begin{array}{cc} \tilde{\lambda} & 0 \\ 0 & \tilde{\lambda} \end{array} \right)$$

とすれば十分である。

ここで、上記の $\tilde{x}_1,\ \tilde{x}_2$ に関しては、$\tilde{x}_1=\hat{y}_1^h,\ \tilde{x}_2=\hat{y}_2^h$ ととる。また、$\tilde{\phi}_1$ としては $\hat{y}_1^h$ の中で、係数の絶対値が最大となるような ϕ_i を、$\tilde{\phi}_2$ としては $\hat{y}_2^h$ の中で、係数の絶対値が最大になるような ϕ_i をとる。

さて、以下の議論でも、第 3 節と同様に直接、真の解を包むのではなく、近似解と真の解の差を包み込むことを考える。そこで、 (52) の真の解を $\hat{y}_1,\ \hat{y}_2 \in H_0^1(\Omega) \quad \hat{m}_{ij} \in \mathbf{R}\ (i,j=1,2)$ として、

$$\begin{aligned} \tilde{y}_i &\equiv \hat{y}_i - \hat{y}_i^h \quad (i=1,2) \\ \tilde{m}_{ij} &\equiv \hat{m}_{ij} - \hat{m}_{ij}^h \quad (i,j=1,2) \end{aligned}$$

と定める。このとき (52) は以下のように書き換えられる。

$$\left\{ \begin{array}{rcl} -\Delta\tilde{y}_1 &=& \Delta\hat{y}_1^h - q(\hat{y}_1^h+\tilde{y}_1) + (\hat{m}_{11}^h+\tilde{m}_{11})(\hat{y}_1^h+\tilde{y}_1) + (\hat{m}_{21}^h+\tilde{m}_{21})(\hat{y}_2^h+\tilde{y}_2) \\ -\Delta\tilde{y}_2 &=& \Delta\hat{y}_2^h - q(\hat{y}_2^h+\tilde{y}_2) + (\hat{m}_{12}^h+\tilde{m}_{12})(\hat{y}_1^h+\tilde{y}_1) + (\hat{m}_{22}^h+\tilde{m}_{22})(\hat{y}_2^h+\tilde{y}_2) \\ \tilde{m}_{11} &=& \tilde{m}_{11} + (\hat{y}_1^h+\tilde{y}_1-\tilde{x}_1,\ \tilde{\phi}_1) \\ \tilde{m}_{12} &=& \tilde{m}_{12} + (\hat{y}_2^h+\tilde{y}_2-\tilde{x}_2,\ \tilde{\phi}_1) \\ \tilde{m}_{21} &=& \tilde{m}_{21} + (\hat{y}_1^h+\tilde{y}_1-\tilde{x}_1,\ \tilde{\phi}_2) \\ \tilde{m}_{22} &=& \tilde{m}_{22} + (\hat{y}_2^h+\tilde{y}_2-\tilde{x}_2,\ \tilde{\phi}_2) \end{array} \right. \tag{56}$$

そこで、 $F : (H_0^1(\Omega))^2 \times \mathbf{R}^4 \to (H_0^1)^2 \times \mathbf{R}^4$ を

$$F\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \tilde{m}_{11} \\ \tilde{m}_{12} \\ \tilde{m}_{21} \\ \tilde{m}_{22} \end{pmatrix} = \begin{pmatrix} (-\Delta)^{-1}\{-q\tilde{y}_1 + \tilde{m}_{11}(\hat{y}_1^h+\tilde{y}_1) + \tilde{m}_{21}(\hat{y}_2^h+\tilde{y}_2) + \hat{m}_{11}^h\tilde{y}_1 + \hat{m}_{21}^h\tilde{y}_2 + v_0^1\} \\ (-\Delta)^{-1}\{-q\tilde{y}_2 + \tilde{m}_{12}(\hat{y}_1^h+\tilde{y}_1) + \tilde{m}_{22}(\hat{y}_2^h+\tilde{y}_2) + \hat{m}_{12}^h\tilde{y}_1 + \hat{m}_{22}^h\tilde{y}_2 + v_0^2\} \\ \tilde{m}_{11} + (\hat{y}_1^h+\tilde{y}_1-\tilde{x}_1,\ \phi_1) \\ \tilde{m}_{12} + (\hat{y}_2^h+\tilde{y}_2-\tilde{x}_2,\ \phi_1) \\ \tilde{m}_{21} + (\hat{y}_1^h+\tilde{y}_1-\tilde{x}_1,\ \phi_2) \\ \tilde{m}_{22} + (\hat{y}_2^h+\tilde{y}_2-\tilde{x}_2,\ \phi_2) \end{pmatrix} \tag{57}$$

但し、

$$v_0^1 \equiv -\Delta\hat{y}_1^h + q\hat{y}_1^h + \hat{m}_{11}^h\hat{y}_1^h + \hat{m}_{21}^h\hat{y}_2^h,\ v_0^2 \equiv -\Delta\hat{y}_2^h + q\hat{y}_2^h + \hat{m}_{12}^h\hat{y}_1^h + \hat{m}_{22}^h\hat{y}_2^h$$

と定義すると、 (52) は $w = (\tilde{y}_1, \tilde{y}_2, \tilde{m}_{11}, \tilde{m}_{12}, \tilde{m}_{21}, \tilde{m}_{22})$ に関して、コンパクト作用素 F の不動点方程式

$$w = F(w) \tag{58}$$

の形でかける。

4.2 検証条件とアルゴリズム

まずは $(S_h)^2 \times \mathbf{R}^4 \subset (H_0^1(\Omega))^2 \times \mathbf{R}^4$ の基底を以下のように定める。

$$\begin{aligned} \Phi_i^1 &= (\phi_i,0,0,0,0,0) \quad (1 \le i \le n) \\ \Phi_i^2 &= (0,\phi_i,0,0,0,0) \quad (1 \le i \le n) \\ \Phi_{2n+1} &= (0,0,1,0,0,0) \\ \Phi_{2n+2} &= (0,0,0,1,0,0) \\ \Phi_{2n+3} &= (0,0,0,0,1,0) \\ \Phi_{2n+4} &= (0,0,0,0,0,1) \end{aligned} \tag{59}$$

ここで、 $\Phi_i = \Phi_i^1,\ \Phi_{n+i} = \Phi_i^2 \quad (1 \le i \le n)$ と定めておく。
また $(H_0^1(\Omega))^2 \times \mathbf{R}^4$ における内積 $< \cdot,\cdot >_{(H_0^1(\Omega))^2\times\mathbf{R}^4}$ を
$w_i = (u_i^1, u_i^2, \lambda_i^1, \lambda_i^2, \lambda_i^3, \lambda_i^4) \quad (i=1,2)$ に対して

$$< w_1, w_2 >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} \equiv (\nabla u_1^1, \nabla u_2^1)_{L^2(\Omega)} + (\nabla u_1^2, \nabla u_2^2)_{L^2(\Omega)} + \lambda_1^1\lambda_2^1 + \lambda_1^2\lambda_2^2 + \lambda_1^3\lambda_2^3 + \lambda_1^4\lambda_2^4$$

と定める。

ここで次を仮定しておく。

Assumption 4.1.
$\rho \equiv (0,0,0,0,0,0)$ とする。 F の ρ での Fréchet 微分 $F'(\rho)$ が存在するとき、

$$P_h[I - F'(\rho)] \ : \ (H_0^1(\Omega))^2 \times \mathbf{R}^4 \to (H_0^1(\Omega))^2 \times \mathbf{R}^4$$

を $(S_h)^2 \times \mathbf{R}^4$ に制限したものは、逆作用素

$$[I - F'(\rho)]_h^{-1} \ : \ (S_h)^2 \times \mathbf{R}^4 \to (S_h)^2 \times \mathbf{R}^4$$

を持つ。ここで $P_h \ : \ (H_0^1(\Omega))^2 \times \mathbf{R}^4 \to (S_h)^2 \times \mathbf{R}^4$ は、

$$P_h(u_1, u_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4) \equiv (P_{h0}u_1, P_{h0}u_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4) \quad (u_i \in H_0^1(\Omega),\ \lambda_i \in \mathbf{R})$$

で定義するするものとする。ここで $P_h 0$ は、前節で定義した $H_0^1(\Omega) \to S_h$ なる H_0^1-projection である。

さて、第 3 節の議論と同様に (58) は有限次元部分と無限次元部分とに分けて

$$\left\{ \begin{array}{rcl} P_h w & = & P_h F(w) \\ (I - P_h)w & = & (I - P_h)F(w) \end{array} \right.$$

とかくことができ、

$$N_h(w) \equiv P_h w - [I - F'(\rho)]_h^{-1}(P_h w - P_h F(w))$$

とおいて、

$$T(w) \equiv N_h(w) + (I - P_h)F(w)$$

とすると、 T はコンパクト作用素となり、次が成り立つ。

Proposition 4.2.

$$w = T(w) \iff w = F(w) \tag{60}$$

proof:
$((S_h)^2 \times \mathbf{R}^4)^\perp = (S_h^\perp)^2 \times \{0\}^4$ であるから、

$$(H_0^1(\Omega))^2 \times \mathbf{R}^4 = (S_h)^2 \times \mathbf{R}^4 \oplus ((S_h)^2 \times \mathbf{R}^4)^\perp = (S_h)^2 \times \mathbf{R}^4 \oplus (S_h^\perp)^2 \times \{0\}^4$$

と書くことができる。また、

$$N_h(w) \in (S_h)^2 \times \mathbf{R}^4,\ (I - P_h)F(w) \in (S_h^\perp)^2 \times \{0\}^4$$

であるから、 $N_h(w)$ と $(I - P_h)F(w)$ とは直交する。

また、 $w = P_h w + (I - P_h)w$ と書くことができるので、

$$
\begin{aligned}
w = T(w) \quad &\Longleftrightarrow \quad \left\{ \begin{array}{rcl} P_h w & = & N_h(w) \\ (I - P_h)w & = & (I - P_h)F(w) \end{array} \right. \\
&\Longleftrightarrow \quad \left\{ \begin{array}{rcl} [I - F'(\rho)]_h^{-1}(P_h w - P_h F(w)) & = & 0 \\ (I - P_h)w & = & (I - P_h)F(w) \end{array} \right. \\
&\Longleftrightarrow \quad \left\{ \begin{array}{rcl} P_h w & = & P_h F(w) \\ (I - P_h)w & = & (I - P_h)F(w) \end{array} \right. \\
&\Longleftrightarrow \quad w = F(w)
\end{aligned}
$$

■

ここでは、解の一意性を保証する Banach の不動点定理に基づく保証手順を用いる ([2] を参照)。
この場合、 $T(0) + T'(W)W \subset \mathbf{int}(W)$ であるような $W \in (H_0^1(\Omega))^2 \times \mathbf{R}^4$: 有界閉凸集合を求めることができれば

$$\exists 1 w \in W \text{ s.t. } w = T(w) \text{ (i.e. } w = F(w))$$

となる。ここで $T'(W)W := \{T'(\tilde{w})w \mid w,\ \tilde{w} \in W\}$ であり、 T' は T の Fréchet 微分を表す。
$W = W_h \oplus W_\perp \quad (W_h \subset (S_h)^2 \times \mathbf{R}^4,\ W_\perp \subset (S_h^\perp)^2 \times \{0\}^4)$ とおくと、検証条件は以下のように書くことができる。

$$
\left\{ \begin{array}{rcl} N_h(0) + N_h'(W)W & \subset & \mathbf{int}(W_h) \\ (I - P_h)F(0) + (I - P_h)F'(W)W & \subset & \mathbf{int}(W_\perp) \end{array} \right. \tag{61}
$$

ここで、 $N_h(w)$ の定義から、

$$[I - F'(\rho)]_h N_h(w) = P_h F(w) - P_h F'(\rho)w \tag{62}$$

である。
今、 $N_h(0) + N_h'(W^{(n-1)})W^{(n-1)} \subset W_h^{(n)}$ であるような $W_h^{(n)}$ を計算することを考えよう。
(62) から、

$$[I - F'(\rho)]_h N_h(0) = P_h F(0) \tag{63}$$

である。よって、まず

$$N_h(0) = (\sum_{j=1}^{n} T_j^1 \phi_j, \sum_{j=1}^{n} T_j^2 \phi_j, T_{2n+1}, T_{2n+2}, T_{2n+3}, T_{2n+4})$$

とおいて、 (63) の左辺と Φ_i との内積を考える。

$$
\begin{aligned}
& < N_h(0), \Phi_i^1 >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} - < P_hF'(\rho)N_h(0), \Phi_i^1 >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} \\
= \;& \textstyle\sum_{j=1}^n T_j^1(\nabla\phi_j, \nabla\phi_i)_{L^2(\Omega)} - ((\hat{m}_{11}-q)\sum_{j=1}^n T_j^1\phi_j + \hat{m}_{21}\sum_{j=1}^n T_j^2\phi_j + T_{2n+1}\hat{y}_1^h + T_{2n+3}\hat{y}_2^h, \phi_i)_{L^2(\Omega)} \\
= \;& \textstyle\sum_{j=1}^n T_j^1\{(\nabla\phi_j, \nabla\phi_i)_{L^2(\Omega)} + (q-\hat{m}_{11}^h)(\phi_j, \phi_i)_{L^2(\Omega)}\} - \sum_{j=1}^n T_j^2(\hat{m}_{21}^h\phi_j, \phi_i)_{L^2(\Omega)} \\
& -T_{2n+1}(\hat{y}_1^h, \phi_i)_{L^2(\Omega)} - T_{2n+3}(\hat{y}_2^h, \phi_i)_{L^2(\Omega)}
\end{aligned}
$$

$$
\begin{aligned}
& < N_h(0), \Phi_i^2 >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} - < P_hF'(\rho)N_h(0), \Phi_i^2 >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} \\
= \;& \textstyle\sum_{j=1}^n T_j^2(\nabla\phi_j\nabla\phi_i)_{L^2(\Omega)} - (\hat{m}_{12}\sum_{j=1}^n T_j^1\phi_j + (\hat{m}_{22}-q)\sum_{j=1}^n T_j^2\phi_j + T_{2n+2}\hat{y}_1^h + T_{2n+4}\hat{y}_2^h, \phi_i)_{L^2(\Omega)} \\
= \;& \textstyle -\sum_{j=1}^n T_j^1(\hat{m}_{12}^h\phi_j, \phi_i)_{L^2(\Omega)} + \sum_{j=1}^n T_j^2\{(\nabla\phi_j, \nabla\phi_i)_{L^2(\Omega)} + (q-\hat{m}_{22}^h)(\phi_j, \phi_i)_{L^2(\Omega)}\} \\
& -T_{2n+2}(\hat{y}_1^h, \phi_i)_{L^2(\Omega)} - T_{2n+4}(\hat{y}_2^h, \phi_i)_{L^2(\Omega)}
\end{aligned}
$$

$$
\begin{aligned}
& < N_h(0), \Phi_{2n+1} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} - < P_hF'(\rho)N_h(0), \Phi_{2n+1} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} \\
= \;& T_{2n+1} - \{\textstyle\int_\Omega \tilde{\phi}_1\sum_{j=1}^n T_j^1\phi_j dx + T_{2n+1}\} \\
= \;& \textstyle -\sum_{j=1}^n T_j^1(\phi_j, \tilde{\phi}_1)
\end{aligned}
$$

$$
\begin{aligned}
& < N_h(0), \Phi_{2n+2} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} - < P_hF'(\rho)N_h(0), \Phi_{2n+2} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} \\
= \;& T_{2n+2} - \{\textstyle\int_\Omega \tilde{\phi}_1\sum_{j=1}^n T_j^2\phi_j dx + T_{2n+2}\} \\
= \;& \textstyle -\sum_{j=1}^n T_j^2(\phi_j, \tilde{\phi}_1)
\end{aligned}
$$

$$
\begin{aligned}
& < N_h(0), \Phi_{2n+3} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} - < P_hF'(\rho)N_h(0), \Phi_{2n+3} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} \\
= \;& T_{2n+3} - \{\textstyle\int_\Omega \tilde{\phi}_2\sum_{j=1}^n T_j^1\phi_j dx + T_{2n+3}\} \\
= \;& \textstyle -\sum_{j=1}^n T_j^1(\phi_j, \tilde{\phi}_2)
\end{aligned}
$$

$$
\begin{aligned}
& < N_h(0), \Phi_{2n+4} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} - < P_hF'(\rho)N_h(0), \Phi_{2n+4} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} \\
= \;& T_{2n+4} - \{\textstyle\int_\Omega \tilde{\phi}_2\sum_{j=1}^n T_j^2\phi_j dx + T_{2n+4}\} \\
= \;& \textstyle -\sum_{j=1}^n T_j^2(\phi_j, \tilde{\phi}_2)
\end{aligned}
$$

であることがわかる。但し、 $1 \le i \le n$ である。

また、同様に右辺と Φ_i との内積を考えると、

$$< P_hF(0), \Phi_i^1 >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} = (v_0^1, \phi_i)_{L^2(\Omega)}$$

$$< P_hF(0), \Phi_i^2 >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} = (v_0^2, \phi_i)_{L^2(\Omega)}$$

$$< P_hF(0), \Phi_{2n+1} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} = (\hat{y}_1^h - \tilde{x}_1, \tilde{\phi}_1)_{L^2(\Omega)}$$

$$< P_hF(0), \Phi_{2n+2} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} = (\hat{y}_2^h - \tilde{x}_2, \tilde{\phi}_1)_{L^2(\Omega)}$$

$$< P_hF(0), \Phi_{2n+3} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} = (\hat{y}_1^h - \tilde{x}_1, \tilde{\phi}_2)_{L^2(\Omega)}$$

$$< P_hF(0), \Phi_{2n+4} >_{(H_0^1(\Omega))^2\times\mathbf{R}^4} = (\hat{y}_2^h - \tilde{x}_2, \tilde{\phi}_2)_{L^2(\Omega)}$$

となることがわかる。但し、 $1 \le i \le n$ である。

そこで、$(2n+4)\times(2n+4)$ 行列 $G:=(g_{ij})_{1\le i,j\le 2n+4}$ を次のように定める。

$$
\begin{aligned}
g_{ij} &= (\nabla\phi_j,\nabla\phi_i)_{L^2(\Omega)}+((q-\hat{m}^h_{11})\phi_j,\phi_i)_{L^2(\Omega)} && (1\le i\le n,\ 1\le j\le n)\\
g_{ij} &= (\nabla\phi_j,\nabla\phi_i)_{L^2(\Omega)}+((q-\hat{m}^h_{22})\phi_j,\phi_i)_{L^2(\Omega)} && (n+1\le i\le 2n,\ n+1\le j\le 2n)\\
g_{ij} &= -(\hat{m}^h_{21}\phi_j,\phi_i)_{L^2(\Omega)} && (1\le i\le n\ , n+1\le j\le 2n)\\
g_{ij} &= -(\hat{m}^h_{12}\phi_j,\phi_i)_{L^2(\Omega)} && (n+1\le i\le 2n\ , 1\le j\le n)\\
g_{i,2n+1} &= -(\hat{y}^h_1,\phi_i)_{L^2(\Omega)} && (1\le i\le n)\\
g_{i,2n+3} &= -(\hat{y}^h_2,\phi_i)_{L^2(\Omega)} && (1\le i\le n)\\
g_{i,2n+2} &= -(\hat{y}^h_1,\phi_i)_{L^2(\Omega)} && (n+1\le i\le 2n)\\
g_{i,2n+4} &= -(\hat{y}^h_2,\phi_i)_{L^2(\Omega)} && (n+1\le i\le 2n)\\
g_{2n+1,j} &= -(\phi_j,\hat{\phi}_1)_{L^2(\Omega)} && (1\le j\le n)\\
g_{2n+2,j} &= -(\phi_j,\hat{\phi}_1)_{L^2(\Omega)} && (n+1\le j\le 2n)\\
g_{2n+3,j} &= -(\phi_j,\hat{\phi}_2)_{L^2(\Omega)} && (1\le j\le n)\\
g_{2n+4,j} &= -(\phi_j,\hat{\phi}_2)_{L^2(\Omega)} && (n+1\le j\le 2n)\\
g_{ij} &= 0 && (\text{上記以外の}i,j)
\end{aligned}
$$

また、$(2n+4)$ 次ベクトル $K_0:=((K_0)_i)_{1\le i\le 2n+4}$ を

$$
\begin{aligned}
(K_0)_i &= (v^1_0,\phi_i)_{L^2(\Omega)} && (1\le i\le n)\\
(K_0)_i &= (v^2_0,\phi_i)_{L^2(\Omega)} && (n+1\le i\le 2n)\\
(K_0)_{2n+1} &= (\hat{y}^h_1-\tilde{x}_1,\tilde{\phi}_1)_{L^2(\Omega)}\\
(K_0)_{2n+2} &= (\hat{y}^h_2-\tilde{x}_2,\tilde{\phi}_1)_{L^2(\Omega)}\\
(K_0)_{2n+3} &= (\hat{y}^h_1-\tilde{x}_1,\tilde{\phi}_2)_{L^2(\Omega)}\\
(K_0)_{2n+4} &= (\hat{y}^h_2-\tilde{x}_2,\tilde{\phi}_2)_{L^2(\Omega)}
\end{aligned}
$$

とし、$\tilde{T}^h_1:=(T^1_1,\dots,T^1_n,T^2_1,\dots,T^2_n,T_{2n+1},T_{2n+2},T_{2n+3},T_{2n+4})^T$ とすると

$$G\tilde{T}^h_1=K_0$$

と書くことができる。ここで、 Assumption 4.1. により、G は可逆であるので

$$\tilde{T}^h_1=G^{-1}K_0 \tag{64}$$

と定めることができる。

次に $N'(\tilde{w})w,\ (w,\tilde{w}\in W^{(n-1)})$ について考えよう。(62) から

$$[I-F'(\rho)]_hN'_h(\tilde{w})w=P_hF(\tilde{w})w-F'(\rho)P_hw$$

であるから、

$$N'_h(\tilde{w})w=(\sum_{j=1}^n S^1_j\phi_j,\sum_{j=1}^n S^2_j\phi_j,S_{2n+1},S_{2n+2},S_{2n+3},S_{2n+4})$$

とおいて、左辺及び、右辺と Φ_i との内積を考える。

$$\tilde{T}^h_2=(S^1_1,\dots,S^1_n,S^2_1,\dots,S^2_n,S_{2n+1},S_{2n+2},S_{2n+3},S_{2n+4})^T$$

とおくと、左辺に関しては $G\tilde{T}^h_2$ と書くことができる。

そこで右辺との内積を考える。まずは $f : (H_0^1(\Omega))^2 \times \mathbf{R}^4 \to (L^2(\Omega))^2 \times \mathbf{R}^4$ を以下のように定義しておく。

$$f \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \tilde{m}_{11} \\ \tilde{m}_{12} \\ \tilde{m}_{21} \\ \tilde{m}_{22} \end{pmatrix} = \begin{pmatrix} -q\tilde{y}_1 + \tilde{m}_{11}(\hat{y}_1^h + \tilde{y}_1) + \tilde{m}_{21}(\hat{y}_2^h + \tilde{y}_2) + \hat{m}_{11}^h \tilde{y}_1 + \hat{m}_{21}^h \tilde{y}_2 + v_0^1 \\ -q\tilde{y}_2 + \tilde{m}_{12}(\hat{y}_1^h + \tilde{y}_1) + \tilde{m}_{22}(\hat{y}_2^h + \tilde{y}_2) + \hat{m}_{12}^h \tilde{y}_1 + \hat{m}_{22}^h \tilde{y}_2 + v_0^2 \\ \tilde{m}_{11} + (\hat{y}_1^h + \tilde{y}_1 - \tilde{x}_1,\ \phi_1) \\ \tilde{m}_{12} + (\hat{y}_2^h + \tilde{y}_2 - \tilde{x}_2,\ \phi_1) \\ \tilde{m}_{21} + (\hat{y}_1^h + \tilde{y}_1 - \tilde{x}_1,\ \phi_2) \\ \tilde{m}_{22} + (\hat{y}_2^h + \tilde{y}_2 - \tilde{x}_2,\ \phi_2) \end{pmatrix} \equiv \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \end{pmatrix}$$

さて、 $w = (w_1, w_2, w_3, w_4, w_5, w_6)^T,\ \tilde{w} = (\tilde{w}_1, \tilde{w}_2, \tilde{w}_3, \tilde{w}_4, \tilde{w}_5, \tilde{w}_6)^T \in (H_0^1(\Omega))^2 \times \mathbf{R}^4$ とおく。

$$< P_h F(\tilde{w}) w, \Phi_i^1 >_{(H_0^1(\Omega))^2 \times \mathbf{R}^4} = (f_1'(\tilde{w})w - f_1'(\rho) P_h w, \phi_i)_{L^2(\Omega)}$$

$$< P_h F(\tilde{w}) w, \Phi_i^2 >_{(H_0^1(\Omega))^2 \times \mathbf{R}^4} = (f_2'(\tilde{w})w - f_2'(\rho) P_h w, \phi_i)_{L^2(\Omega)}$$

$$< P_h F(\tilde{w}) w, \Phi_{2n+1} >_{(H_0^1(\Omega))^2 \times \mathbf{R}^4} = \int_\Omega (1 - P_{h0}) w_1 \tilde{\phi}_1 dx$$

$$< P_h F(\tilde{w}) w, \Phi_{2n+2} >_{(H_0^1(\Omega))^2 \times \mathbf{R}^4} = \int_\Omega (1 - P_{h0}) w_2 \tilde{\phi}_1 dx$$

$$< P_h F(\tilde{w}) w, \Phi_{2n+3} >_{(H_0^1(\Omega))^2 \times \mathbf{R}^4} = \int_\Omega (1 - P_{h0}) w_1 \tilde{\phi}_2 dx$$

$$< P_h F(\tilde{w}) w, \Phi_{2n+4} >_{(H_0^1(\Omega))^2 \times \mathbf{R}^4} = \int_\Omega (1 - P_{h0}) w_2 \tilde{\phi}_2 dx$$

であることがわかる。但し $1 \le i \le n$ である。

よって $(2n+4)$ 次ベクトル $K_1 := ((K_1)_i)_{1 \le i \le 2n+4}$ を

$$\begin{aligned} (K_1)_i &= (f_1'(\tilde{w})w - f_1'(\rho) P_h w, \phi_i)_{L^2(\Omega)} \quad (1 \le i \le n) \\ (K_1)_i &= (f_2'(\tilde{w})w - f_2'(\rho) P_h w, \phi_i)_{L^2(\Omega)} \quad (n+1 \le i \le 2n) \\ (K_1)_{2n+1} &= \int_\Omega (1 - P_{h0}) w_1 \tilde{\phi}_1 dx \\ (K_1)_{2n+2} &= \int_\Omega (1 - P_{h0}) w_2 \tilde{\phi}_1 dx \\ (K_1)_{2n+3} &= \int_\Omega (1 - P_{h0}) w_1 \tilde{\phi}_2 dx \\ (K_1)_{2n+4} &= \int_\Omega (1 - P_{h0}) w_2 \tilde{\phi}_2 dx \end{aligned}$$

と定義すると、

$$G \tilde{T}_2^h = K_1$$

となり、

$$\tilde{T}_2^h = G^{-1} K_1 \tag{65}$$

と定めることができる。ここで各 $\tilde{T}_1^h,\ \tilde{T}_2^h$ は区間として計算することになるので、 $W_h^{(n)} = G^{-1}K_0 + G^{-1}K_1$ と決めることができる。

次は無限次元について考える。まずは $\|u - P_h u\|_{H_0^1(\Omega)}$ の評価をする。今、 u は

$$u = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} A_{ij} \sin ix \sin jy$$

としているので、

$$P_h u = \sum_{i=1}^{M}\sum_{j=1}^{N} A_{ij} \sin ix \sin jy$$

である。このとき、以下の評価を得ることができる。

Proposition 4.3.

$$\|u - P_h u\|^2_{H_0^1(\Omega)} \leq \frac{1}{(\bar{N}+1)^2+1} \|\Delta u\|^2_{L_2(\Omega)} \tag{66}$$

但し、 $\bar{N} \equiv \min\{M, N\}$ である。

proof:
$k = \|\sin ix \sin jy\|_{L^2(\Omega)}$ とおく。

$$\begin{aligned}
\|u - P_h u\|^2_{H_0^1(\Omega)} &= \|\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} A_{ij} \sin ix \sin jy - \sum_{i=1}^{M}\sum_{j=1}^{N} A_{ij} \sin ix \sin jy\|^2_{H_0^1(\Omega)} \\
&= k^2 \sum_{i=M+1}^{\infty}\sum_{j=1}^{N} \frac{(i^2+j^2)^2 A_{ij}^2}{i^2+j^2} + k^2 \sum_{i=1}^{M}\sum_{j=N+1}^{\infty} \frac{(i^2+j^2)^2 A_{ij}^2}{i^2+j^2} \\
&+ k^2 \sum_{i=M+1}^{\infty}\sum_{j=N+1}^{\infty} \frac{(i^2+j^2)^2 A_{ij}^2}{i^2+j^2} \\
&\leq \max\{b_1, b_2, b_3\} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} k^2 (i^2+j^2)^2 A_{ij}^2
\end{aligned}$$

ここで、

$$b_1 = \max_{M+1\leq i,\ 1\leq j\leq N} \frac{1}{i^2+j^2},\ b_2 = \max_{1\leq i\leq M,\ N+1\leq j} \frac{1}{i^2+j^2},\ b_3 = \max_{M+1\leq i,\ N+1\leq j} \frac{1}{i^2+j^2}$$

また、

$$\|\Delta u\|^2_{L^2(\Omega)} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} k^2 (i^2+j^2)^2 A_{ij}^2$$

であるから、

$$\|u - P_h u\|^2_{H_0^1(\Omega)} \leq \frac{1}{(\bar{N}+1)^2+1} \|\Delta u\|^2_{L^2(\Omega)}$$

を得ることができる。 ■

(66) から、

$$\|T_\perp^1(0)\|_{H_0^1(\Omega)} = \|F_1(0) - P_h F_1(0)\|_{H_0^1(\Omega)} \leq \frac{1}{\sqrt{(\bar{N}+1)^2+1}} \|f_1(0)\|_{L_2(\Omega)} \tag{67}$$

が導かれる。ここで $f_1(0) = \Delta \hat{y}_1^h + (\hat{m}_{11}^h - q)\hat{y}_1^h + \hat{m}_{21}^h \hat{y}_2^h = v_0^1$ である。
同様に

$$\|T_\perp^2(0)\|_{H_0^1(\Omega)} \leq \frac{1}{\sqrt{(\bar{N}+1)^2+1}} \|f_2(0)\|_{L_2(\Omega)} \tag{68}$$

が導かれる。ここで $f_2(0) = \Delta\hat{y}_2^h + (\hat{m}_{12}^h - q)\hat{y}_1^h + \hat{m}_{22}^h\hat{y}_2^h = v_0^2$ である。
但し、 $(I - P_h)F(w) \equiv T_\perp(w) \equiv (T_\perp^1(w), T_\perp^2(w), 0, 0, 0, 0)$ とおいた。また、

$$\|T_\perp(w)\|_{H_0^1(\Omega)^2\times\mathbf{R}^4} = (\|T_\perp^1(w)\|_{H_0^1(\Omega)}^2 + \|T_\perp^2(w)\|_{H_0^1(\Omega)}^2)^{\frac{1}{2}}$$

である。

次に $w, \tilde{w} \in W$ に対して、

$$\|T_\perp^{1\prime}(\tilde{w})w\|_{H_0^1(\Omega)} \leq \frac{1}{\sqrt{(\bar{N}+1)^2+1}}\|f_1'(\tilde{w})w\|_{L_2(\Omega)} \tag{69}$$

$$\|T_\perp^{2\prime}(\tilde{w})w\|_{H_0^1(\Omega)} \leq \frac{1}{\sqrt{(\bar{N}+1)^2+1}}\|f_2'(\tilde{w})w\|_{L_2(\Omega)} \tag{70}$$

を得ることができる。ここで、 $(I - P_h)F'(\tilde{w})w \equiv T_\perp'(\tilde{w})w \equiv (T_\perp^{1\prime}(\tilde{w})w, T_\perp^{2\prime}(\tilde{w})w, 0, 0, 0, 0)$ とおいた。

さて、

$$[\alpha] \equiv \{\phi \in S_h^\perp \mid \|\phi\|_{h_0^1(\Omega)} \leq \alpha\} \times \{0\} \times \{0\}^4 \subset (H_0^1(\Omega))^2 \times \mathbf{R}^4$$

$$[\beta] \equiv \{0\} \times \{\phi \in S_h^\perp \mid \|\phi\|_{h_0^1(\Omega)} \leq \beta\} \times \{0\}^4 \subset (H_0^1(\Omega))^2 \times \mathbf{R}^4$$

と定義しておく。検証に関するアルゴリズムをまとめると以下のようになる。

Algorithm 4.4.
$n = 0$ のとき

$$w_h^{(0)} = \{0\} \subset (S_h)^2 \times \mathbf{R}^4,\ \alpha_0 = 0 \in \mathbf{R}_{\geq 0},\ \beta_0 = 0 \in \mathbf{R}_{\geq 0}$$

とし、 $W^{(0)} = w_h^{(0)} \oplus [\alpha_0] \oplus [\beta_0]$ とおく。

$n \geq 1$ のとき、 $0 < \epsilon \ll 1$ であるような ϵ に対して

$$\left\{\begin{array}{rcl} \tilde{w}_h^{(n-1)} & = & w_h^{(n-1)} + (\sum_{j=1}^n [-1,1]\epsilon\phi_j, \sum_{j=1}^n [-1,1]\epsilon\phi_j, [-1,1]\epsilon, [-1,1]\epsilon, [-1,1]\epsilon, [-1,1]\epsilon) \\ \tilde{\alpha}_{n-1} & = & \alpha_{n-1} + \epsilon \\ \tilde{\beta}_{n-1} & = & \beta_{n-1} + \epsilon \end{array}\right. \tag{71}$$

として、

$$\tilde{W}^{(n-1)} = \tilde{w}_h^{(n-1)} \oplus [\tilde{\alpha}_{n-1}] \oplus [\tilde{\beta}_{n-1}]$$

に対して、 $w_h^{(n)},\ \alpha_n,\ \beta_n$ を次のように定義する。

$$\left\{\begin{array}{rcl} w_h^{(n)} & \supset & N_h(0) + N_h'(\tilde{W}^{(n-1)})\tilde{W}^{(n-1)} \\ \alpha_n & = & \sup_{w,\tilde{w}\in\tilde{W}^{(n-1)}} \frac{1}{\sqrt{(\bar{N}+1)^2+1}}(\|v_0^1\|_{L^2(\Omega)} + \|f_1'(\tilde{w})w\|_{L^2(\Omega)}) \\ \beta_n & = & \sup_{w,\tilde{w}\in\tilde{W}^{(n-1)}} \frac{1}{\sqrt{(\bar{N}+1)^2+1}}(\|v_0^2\|_{L^2(\Omega)} + \|f_2'(\tilde{w})w\|_{L^2(\Omega)}) \end{array}\right. \tag{72}$$

以上のアルゴリズムに対し、以下の定理が解の存在を保証する。

Theorem 4.5.

$$\begin{aligned} & \exists n \in \mathbf{N} \text{ s.t. } w_h^{(n)} \subset \tilde{w}_h^{(n-1)},\ \alpha_n \leq \tilde{\alpha}_{n-1},\ \beta_n \leq \tilde{\beta}_{n-1} \\ \Rightarrow\quad & \exists w \in w_h^{(n)} \oplus [\alpha_n] \oplus [\beta_n] \text{ s.t. } w = F(w) \end{aligned}$$

以上のアルゴリズムにより、精度保証付きで (52) の解

$$(-\Delta + q)[Y] = [Y][M]$$

を得ることができるが、第 2 節における議論と同様に $[M]$ に対して、 Gerschgorin の定理を適用することにより、固有値の包み込みを得ることができる。

以上のように、多重固有値と不変部分空間の基底関数に対する精度保証を実現することができる。

4.3 数値例

数値例 1：

以下では、 $q = \sin x \sin y$ としたときの例を挙げる。 S_h の基底関数の個数は $M = 11,\ N = 11$ として、 121 個の基底を使っている。また近似値を計算した結果、多重 (二重) 固有値の 1 つとして、 5.5724 がそうであるかに見えた。よって、これに対して精度保証のアルゴリズムを適用する。このとき、

反復回数 =5 で、

$w_h^{(5)} \subset \tilde{w}_h^{(4)},\ \alpha_5 \leq \tilde{\alpha}_4,\ \beta_5 \leq \tilde{\beta}_4$

となった。ここで、

$\max |w_h^{(5)}|$ =3.7087e-04

$\max |\tilde{w}_h^{(4)}|$ =3.7512e-04

であった。また、

$\|v_0^1\|_{L^2(\Omega)} = 0.0026$

$\|v_0^2\|_{L^2(\Omega)} = 0.0040$

$\sup_{w,\tilde{w} \in \tilde{W}^{(4)}} \|f_1'(\tilde{w})w\|_{L^2(\Omega)} = 0.0042$

$\sup_{w,\tilde{w} \in \tilde{W}^{(4)}} \|f_2'(\tilde{w})w\|_{L^2(\Omega)} = 0.0137$

であった。また

$\lambda \in 5.5724 + [-0.6260\text{e-}03,\ 0.6260\text{e-}03]$

を得ることができた。

数値例 2：

数値例 1 と同様の問題を考える。但し、基底関数の個数は $M = 8,\ N = 8$ として、 64 個の基底を使った。このとき、近似値計算の結果、 10.524940396647 と 10.584986363725 が二重固有値であるかのように見えた。そこで、これに対し、精度保証のアルゴリズムを適用する。すると、

$\lambda_1 \in 10.524940396647 + [-0.5685\text{e-}03,\ 0.5685\text{e-}03]$

$\lambda_2 \in 10.584986363725 + [-0.5862\text{e-}03,\ 0.5862\text{e-}03]$

と包み込みが完全に分離され、実は二重ではなく、近接固有値であることがわかった。

5　結果の考察と今後の課題

以上述べたように、行列の固有値に対する Rump の方法を、楕円型固有値問題の精度保証に拡張できた。また、数値例 2 に示されるように、一見すると多重であるかのような固有値を、単純固有値として保証付きで分離できることがわかった。

次に今後の課題であるが、最も改善しなければならないのは、計算に関するコストの削減であろう。行列の固有値に対する Rump の方法では、うまく固有値と固有ベクトルとを 1 つにまとめることによって、大幅なコスト削減を実現していたが (第 2 節を参照) 、それに習い、楕円型固有値問題に対しても、うまくコスト削減を実現できないだろうかということが第 1 の課題であるように思う。

第 2 に、固有関数の包み込みについてである。上のアルゴリズムでは、固有値及び、それに伴う不変部分空間の基底の包み込みはできているが、固有関数の包み込みはできてはいない。固有関数の包み込みを得るためには、精度保証付きで得られた

$$(-\Delta + q)[Y] = [Y][M]$$

の $[M] = \begin{pmatrix} [m_{11}] & [m_{12}] \\ [m_{21}] & [m_{22}] \end{pmatrix}$ において、

$$P^{-1}[M]P = \begin{pmatrix} [\lambda_1] & 0 \\ 0 & [\lambda_2] \end{pmatrix}$$

となるような P を見つけなければならない。しかし、これは難しい問題と思われ、別の観点からの考察が必要であろう。

参考文献

[1] J.J.Dongarra, C.B.Moler, J.H.Wilkinson, Improving the accuracy of computed eigenvalues and eigenvectors, SIAM J. Numer. Anal. Vol.20, No.1 : 23-45, 1983.

[2] K.Nagatou, A numerical method to verify the elliptic eigenvalue problems including a uniqueness property, Computing 63 : 109-130, 1999.

[3] M.T.Nakao, N.Yamamoto, K.Nagatou, Numerical verifications for eigenvalues of second-order eliptic operators, Japan J. Industrial and Applied Mathematics 16, N0.3 : 307-320, 1999.

[4] 中尾充宏, 山本野人, 精度保証付き数値計算 −コンピュータによる無限への挑戦−, 日本評論社, 1998.

[5] S.M.Rump, Computational error bounds for multiple or nearly multiple eigenvalues, to appear in Journal of Linear Algebra and its applications.

[6] S.M.Rump, Verification Methods for dense and sparse systems of equations, Topics in Validated Computations : 63-134, 1994

[7] S.M.Rump, INTLAB - INTerval LABoratory, Developments in Reliable Computing : 77-104, 1999

[8] S.Oishi, Fast enclosure matrix eigenvalues and singular values via rounding mode controlled computation, preprint.

[9] Y.Saad, Numerical methods for large eigenvalue problems, Manchester university press, Manchester, 1992.