

一つのレゾルベントから構成されたフィルタを用いた 実対称定値一般固有値問題に対するフィルタ対角化法の実験

村上 弘^{1,a)}

概要: フィルタ対角化法を用いて、実対称定値一般固有値問題の固有対で固有値が要求された区間にあるものを求める。フィルタに用いる線形作用素はある虚数シフトに於けるレゾルベントの作用の多項式の実部として構成する。この方法がうまく機能するかを調べるために数値実験を行なった。

キーワード: フィルタ対角化, 固有値問題, レゾルベント, 多項式

An experiment of filter diagonalization method for symmetric definite generalized eigenproblem which uses a filter constructed from a resolvent

HIROSHI MURAKAMI^{1,a)}

Abstract: For a real symmetric definite generalized eigenproblem, we use the filter diagonalization method to solve those eigenpairs whose eigenvalues are in a requested interval. The linear operator we use for the filter is constructed as the real part of a polynomial of the action of a resolvent with a imaginary shift. We made some numerical experiments to check if this method works well.

Keywords: filter diagonalization, eigenproblem, resolvent, polynomial

1. 概要

今回の研究は [5], [6], [7] の続きである。最小 2 乗法の系統の方法で係数を決めることにより、「単一の虚シフトのレゾルベント作用の多項式の実部」であるフィルタ作用素を作成し、それを用いたフィルタ対角化法で実対称定値一般固有値問題の中間固有値の固有対を解く数値実験を具体的にに行った結果を示すものである。

いま係数行列 A と B が疎で実対称で B が正定値である一般固有値問題 $Ax = \lambda Bx$ の固有対のうちで固有値が指定した区間にあるものをフィルタ対角化法を用いて解くことにする。これまでは主に、フィルタとして相異なる複数のシフトを持つレゾルベントの線形結合を用いる方法につい

て研究してきた [1], [2], [3], [4]。その際に用いるシフトは虚数で個数は例えば 6 から 16 程度である。シフトが ρ であるレゾルベント $R(\rho) \equiv (A - \rho B)^{-1}B$ の作用は、まず B を乗じた後に $A - \rho B$ を係数とする連立 1 次方程式を解いて実現する。連立 1 次方程式は、たとえば係数が帯行列ならば Gauss 消去法や LU 分解（あるいは係数の対称性を用いると帯用の修正 Cholesky 分解）などの行列分解を用いる直接法で解き、係数が一般の疎行列ならば不完全 LU 分解などを前処理に用いて反復法で解くことになる。すると行列の分解あるいは不完全分解が全体の計算量の中でかなりの割合を占める。また連立 1 次方程式を解き終わるまで分解された因子を保持する記憶量も、この方法の適用上での制約となりがちである。これに対する簡単な対処法としては、

- 一時期に並行に処理するレゾルベントの数を減らす。
- 低次のフィルタを採用して必要なレゾルベントの数を

¹ 首都大学東京・数理情報科学専攻
Department of Mathematics and Information Sciences,
Tokyo Metropolitan University
^{a)} mrkmmhrsh@tmu.ac.jp

減らす。その低次のフィルタを反復適用して特性の悪さがある程度補う。

などが考えられるので、まずそれらについて以下に述べる。

● 並行に処理するレゾルベントの数を減らす方法

複数のレゾルベントをいくつかの組に分けて、各組を順次に（組の内部では並行に）実行すれば、行列分解の保持に必要な記憶量を減らせる。ただし分解にかかる演算の合計量は変わらない。また分割した組の数に比例して経過時間が延びる。この方法は計算の並行度を低減するだけなので、フィルタ特性や得られる近似解は変わらない。

● 使用するレゾルベントの数を減らす方法

フィルタを構成するレゾルベントの数を減らすことで行列分解の計算量を減らし、また分解因子を保持する記憶量も減らす方法である。しかし次数を下げるとフィルタの伝達特性は悪化する。特に、阻止域での減衰率が不十分となるので、低次のフィルタを反復したものに対角化用のフィルタとする。その場合は低次のフィルタを最初に適用した際の行列分解を保持できれば、以降の反復では新たな分解を行う必要はない。

但し、レゾルベントの数をあまり少なくすると、低次のフィルタを阻止域に於ける伝達率が十分小さくなるまで反復して得られるフィルタは、通過域に於ける伝達率の最大最小の比が反復により拡大されて、得られる近似解の精度が悪化する。

今回、記憶量が非常に強い制約条件である場合を想定して上記の方針とは異なり、（虚数シフトの）レゾルベント1個だけを用いてみる。そうしてレゾルベントの作用を単に反復するのではなくて、レゾルベントの作用の多項式を用いる（実際はフィルタ作用素は「単一のレゾルベントの作用の多項式の実部」とする）。レゾルベントの n 次多項式をベクトルに作用させる過程ではレゾルベントを n 回適用するが、レゾルベントの作用を行列分解あるいは不完全行列分解を用いて実現する場合は、最初の分解結果が保持できれば同じ行列の分解を繰り返す必要はないので、分解は全体で1回だけである。

2. レゾルベントの多項式によるフィルタ

いま解きたい実対称定値一般固有値問題を $A\mathbf{v} = \lambda B\mathbf{v}$ とする。ここで行列 A と B は実対称、 B は正定値である。

シフト ρ のレゾルベントを $R(\rho) \equiv (A - \rho B)^{-1}B$ とする。レゾルベント $R(\rho)$ の固有対 $(\lambda, \mathbf{v})$ への作用は $R(\rho)\mathbf{v} = 1/(\lambda - \rho) \cdot \mathbf{v}$ である。いまフィルタ $\mathcal{F}$ がこのレゾルベントの n 次多項式の実部として (c_∞ は実数) 以下のように与えられるとする：

$$\mathcal{F} = c_\infty I + \operatorname{Re} \sum_{k=1}^n \gamma_k \{R(\rho)\}^k.$$

固有対 $(\lambda, \mathbf{v})$ に対してフィルタ $\mathcal{F}$ を作用させると $\mathcal{F}\mathbf{v} = f(\lambda)\mathbf{v}$ であることがわかる。ここで $f(\lambda)$ は λ の実有理関数で

$$f(\lambda) = c_\infty + \operatorname{Re} \sum_{k=1}^n \gamma_k / (\lambda - \rho)^k$$

であり、固有値 λ の固有ベクトルのフィルタによる伝達率（伝達関数）を与える。上記の構成から $f(\lambda)$ の極は ρ とその複素共役の2個だけである。

いま区間 $[a, b]$ に固有値がある固有対を求めたいとする。そのとき $\lambda \in [a, b]$ から $t \in [-1, 1]$ への1次変換 $\lambda = (a+b)/2 + t(b-a)/2$ により、固有値の座標 λ に対応する正規化座標 t を定義する。そうして $t \in [-1, 1]$ は通過域、 $1 < \mu \leq |t|$ は阻止域、途中の $1 < |t| < \mu$ は遷移域とする。正規化座標 t による伝達関数 $g(t)$ を $g(t) \equiv f(\lambda)$ で定義する。伝達関数 $g(t)$ への制約は、通過域に於ける最小値は g_{pass} で、逆に $g(t) \geq g_{\text{pass}}$ となるのは通過域に限られ、そうして阻止域に於ける $|g(t)|$ の上限値が g_{stop} である。いま $g(t)$ を偶関数に限定すれば、 $g(t)$ の複素共役な2個の極は純虚数になる。遷移域と通過域の幅の比である μ を大きくすると、遷移域が広がるので遷移域に入る固有値が増える傾向が増す。遷移域に固有値が多く入ればそれだけフィルタで濾過するベクトルの数を増す必要が生じる。阻止域に於ける伝達率の大きさの上限 g_{stop} が微小でなければ、不変部分空間の近似が悪化する。通過域での伝達率の最大最小比が大きければ、計算で求めた（固有値が通過域にある）近似対の精度が不均一になる可能性がある。

いま純虚数である $g(t)$ の極のうちで虚部が正のものを $\sigma\sqrt{-1}$ ($\sigma > 0$) とすると、

$$g(t) = c'_\infty + \operatorname{Re} \sum_{k=1}^n c_k / (t - \sigma\sqrt{-1})^k$$

である。もとの固有値の座標 λ での伝達関数は

$$f(\lambda) = c_\infty + \operatorname{Re} \sum_{k=1}^n \gamma_k / (\lambda - \rho)^k$$

である。すると関係 $f(\lambda) \equiv g(t)$ から、定数項の間には $c_\infty = c'_\infty$ 、極の位置の間には $\rho = (a+b)/2 + \sigma(b-a)/2 \cdot \sqrt{-1}$ 、多重極の係数の間には $\gamma_k = c_k \{(b-a)/2\}^k$, $k=1, 2, \dots, n$ の関係がある。すると正規化座標での伝達関数 $g(t)$ を決めれば $f(\lambda)$ が決まり、 $f(\lambda)$ からはそれに対応するフィルタ作用素 $\mathcal{F}$ が決まる。

さらに計算の簡便化のため、実数 $\sigma > 0$ による尺度変換で、 $t \equiv \sigma x$ と置き、 $h(x) \equiv g(t)$ と定義する：

$$h(x) = c''_\infty + \operatorname{Re} \sum_{k=1}^n \alpha_k / (1 + x\sqrt{-1})^k.$$

すると係数の間には $c''_\infty = c_\infty$ および $c_k = (-\sigma\sqrt{-1})^k \alpha_k$, $k=1, 2, \dots, n$ の関係がある。また $g(t)$ を偶関数にしたこと

から $h(x)$ も偶関数で、係数 α_k は実数にとることができる。さらに $x \equiv \tan \theta$, $\hat{h}(\theta) \equiv h(x)$ とおくと、

$$\hat{h}(\theta) = c_\infty + \sum_{k=1}^n \alpha_k \cos(k\theta) (\cos \theta)^k.$$

である。これ以降は無遠慮での伝達率は零に限定して $c_\infty = 0$ とする。

ここまですべてを整理すれば (ただし $\sigma = 1$ と簡略化した)。

- 1 次変換 $\lambda = (a+b)/2 + x(b-a)/2$ により $\lambda \in [a, b]$ を $x \in [-1, 1]$ に対応させる。
- 実の有理関数 $h(x)$ は x の偶関数で、純虚数 $x = \pm\sqrt{-1}$ だけに極を持ち、(簡単化のため) 無限遠での値を零とすると、

$$h(x) = \text{Re} \sum_{k=1}^n \alpha_k / (1 + x\sqrt{-1})^k.$$

- 実数パラメタ α_k , $k=1, 2, \dots, n$ をうまく調整して、通過域 $[-1, 1]$ から離れると $|h(x)|$ が急減少となるようにする。
- $h(x)$ の座標を λ に変換した実有理関数を $f(\lambda) \equiv h(x)$ とすると、伝達関数が $f(\lambda)$ のフィルタ作用素 $\mathcal{F}$ はレゾルベントの多項式を用いて

$$\mathcal{F} = \text{Re} \sum_{k=1}^n \gamma_k \{\mathcal{R}(\rho)\}^k$$

である。ただし虚部が正のシフトの値は $\rho \equiv (a+b)/2 + (b-a)/2 \cdot \sqrt{-1}$ であり、レゾルベントの k 乗の項 $\mathcal{R}(\rho)^k$ に対する係数は $\gamma_k \equiv \alpha_k \{(b-a)/2\}^k (-\sqrt{-1})^k$ である。

次数 n を与えて、純虚数 $\pm\sqrt{-1}$ だけに極を持つ偶関数の n 次実有理関数

$$h(x) = \text{Re} \sum_{k=1}^n \alpha_k / (1 + x\sqrt{-1})^k$$

を ($h(x)$ が偶関数であることから係数 α_k を実数にとつて)、係数 α_k , $k=1, 2, \dots, n$ をうまく調整して、 $|h(x)|$ の値が通過域 $[-1, 1]$ では 1 の付近にあり、通過域からある程度離れた阻止域では微小となるようにする。

ここからは計算の便宜上、 $x \equiv \tan \theta$ と変数を置換して、 $\hat{h}(\theta) \equiv h(x)$ と置くと

$$\hat{h}(\theta) = \sum_{k=1}^n \alpha_k \cos(k\theta) (\cos \theta)^k.$$

偶関数であることから $\theta \geq 0$ だけを考えれば良いので、通過域についても $\theta \in [0, \pi/4]$ だけを考えれば良い。

以下では、実数 α_k を最小 2 乗法 (の系統) を用いて調整する方法を示す。

2.1 最小 2 乗法 (解法 I)

阻止域と通過域に於いて伝達関数 $\hat{h}(\theta)$ の理想特性からの誤差の 2 乗を (簡単のため重みを 1 として) θ で積分した値をそれぞれ J_{stop} と J_{pass} とする。そうして微小な正数 η を適切に選び目的関数 $J \equiv J_{\text{stop}} + \eta J_{\text{pass}}$ の最小化を行う。

通過域の中央と端点 $x=0$, $x=1$, 阻止域の両端点 $x=\mu$, $x=\infty$ に対応する θ の値をそれぞれ $\theta_0=0$, $\theta_1=\pi/4$, $\theta_\mu=\arctan \mu$, $\theta_\infty=\pi/2$ とすると、

$$J_{\text{stop}} \equiv \int_{\theta_\mu}^{\theta_\infty} \{\hat{h}(\theta)\}^2 d\theta = \sum_{p,q=1}^n \alpha_p A_{p,q} \alpha_q$$

および

$$J_{\text{pass}} \equiv \int_{\theta_0}^{\theta_1} \{1 - \hat{h}(\theta)\}^2 d\theta = \sum_{p,q=1}^n \alpha_p B_{p,q} \alpha_q - 2 \sum_{p=1}^n \alpha_p B_{p,0} + \text{Const}$$

である。ただしここで

$$A_{p,q} \equiv \int_{\theta_\mu}^{\theta_\infty} \cos(p\theta) \cos(q\theta) (\cos \theta)^{p+q} d\theta,$$

$$B_{p,q} \equiv \int_{\theta_0}^{\theta_1} \cos(p\theta) \cos(q\theta) (\cos \theta)^{p+q} d\theta$$

である (注: ここで用いている A や B は、実対称定値一般化固有値問題の係数行列 A や B と記号は同じであるが無関係である)。これら $A_{p,q}$, $B_{p,q}$ を与える定積分の値は解析的な表式を用いて具体的に計算できる。そうして $J \equiv J_{\text{stop}} + \eta J_{\text{pass}}$ の最小化条件は、 $b_p = B_{p,0}$ とおくと、実対称正定値の係数行列を持つ連立 1 次方程式

$$(A + \eta B) \alpha = \eta \mathbf{b}$$

に帰着し、これを解いて α_k , $k=1, 2, \dots, n$ が求まる。

2.2 最小 2 乗法 (解法 II)

α を 2-ノルムが一定のベクトルとみなして、まず先に阻止域での積分

$$J_{\text{stop}} = \sum_{p,q=1}^n \alpha_p A_{p,q} \alpha_q = \alpha^T A \alpha$$

の最小化を試みる。いま実対称正定値行列 A の固有値分解を $A \rightarrow UDU^T$ とする。ただし D は A の固有値を並べた対角行列で、 U は固有値に対応して固有ベクトルを並べた直交行列である。 J_{stop} の値が最小になるのは、 A の最小固有値の固有ベクトルを α に選ぶときであるが、そうすると通過域での伝達特性を調整する自由度がないので、以下のような変更を加えて自由度を入れる。

いま微小な正数 ϵ を適切に決める。固有値が ϵ 以下の A の固有対が ℓ 個存在したとする。そのときそれら ℓ 個の固有ベクトルで張られた部分空間を $\mathcal{S}^{(\ell)}$ とする。もしも $\mathcal{S}^{(\ell)}$

内にベクトル α があれば, $J_{\text{stop}} \leq \epsilon \|\alpha\|_2^2$ である. そこで α を部分空間 $\mathcal{S}^{(\ell)}$ 内に制限した上で J_{pass} の値を最小化する. これは ℓ 次に縮小された実対称正定値行列を係数とする連立 1 次方程式に帰着できることが示せる.

通過域に於ける特性を (伝達率の最大最小の比を減らし) 改善すべく部分空間を拡大すると, 阻止域に於ける特性が (伝達率の上限が増大して) 悪化する. 逆に, 阻止域に於ける特性を改善すべく部分空間を縮小すると, 通過域に於ける特性が悪化する. そのようなトレードオフの関係がある.

3. フィルタ対角化の数値実験

上述した方法でいくつかのフィルタを作成し, それを用いてフィルタ対角化法の数値計算を実験してみる.

3.1 フィルタの伝達関数の係数作成

まず, 単一のレゾルベントの多項式 (の実部) によるフィルタを作成する. 固有値の通過域 $\lambda \in [a, b]$ が $x \in [-1, 1]$ に対応する正規化座標 x により, $\pm\sqrt{-1}$ だけに極を持ち偶関数である (半) 次数 n のフィルタの伝達関数を $h(x) = \text{Re} \sum_{k=1}^n \alpha_k / (1 + x\sqrt{-1})^k$ とする. 係数 α_k , $k=1, 2, \dots, n$ は実数にとれる.

今回の数値実験に用いたのは, 最小 2 乗法の系統による方法を用いて作成した以下の 3 種類のフィルタである.

- フィルタ (その 1): これは $n = 15$ で, フィルタの形状は $\mu = 2.0$, $g_{\text{pass}} = 2.3 \times 10^{-4}$, $g_{\text{stop}} = 1.1 \times 10^{-15}$ である.
- フィルタ (その 2): これも $n = 15$ であるが, フィルタの形状は $\mu = 1.5$, $g_{\text{pass}} = 5.46 \times 10^{-5}$, $g_{\text{stop}} = 5.85 \times 10^{-13}$ である. (フィルタ (その 1) よりも μ の値が小さいが, 他方では g_{pass} の値が下がり, g_{stop} の値が上がっている).
- フィルタ (その 3): これは $n = 20$ で, フィルタの形状は $\mu = 2.0$, $g_{\text{pass}} = 1.273 \times 10^{-2}$, $g_{\text{stop}} = 2.6 \times 10^{-15}$ である (次数 n を上げたことによる自由度の増加から, フィルタ (その 1) よりも g_{pass} の値が 1 に近づいている).

求めたこれらの 3 種類のフィルタについて, それぞれ実数係数 α_k , $k=1, 2, \dots, n$ の値を倍精度に丸めたものを表 1, 表 2, 表 3 に示す. フィルタ対角化法の数値実験にはこの表中の数値を実際に用いている.

また伝達関数の大きさ $|h(x)|$ の値を (偶関数性から正側の x だけについて) 両対数, 縦対数, 横対数の各グラフにプロットしたものを

- フィルタ (その 1) では図 1, 図 2, 図 3 に,
 - フィルタ (その 2) では図 4, 図 5, 図 6 に,
 - フィルタ (その 3) では図 7, 図 8, 図 9 に,
- それぞれ示す.

表 1 フィルタ (その 1): 正規座標での係数

$n=15$, $\mu=2$, $g_{\text{pass}}=2.3 \times 10^{-4}$, $g_{\text{stop}}=1.1 \times 10^{-15}$

k	α_k
1	3.10422 91727 23495 E-1
2	3.10422 91727 25609 E-1
3	2.85453 67519 83506 E-1
4	2.35515 19113 67395 E-1
5	1.64913 99494 59607 E-1
6	8.22631 58940 55446 E-2
7	-6.57520 79352 44120 E-4
8	-7.11802 27019 60262 E-2
9	-1.18756 19212 14338 E-1
10	-1.37828 28527 33139 E-1
11	-1.29654 88587 73316 E-1
12	-1.01680 66293 50991 E-1
13	-6.60360 83956 00963 E-2
14	-3.26587 11429 62141 E-2
15	-1.19174 53737 97113 E-2

表 2 フィルタ (その 2): 正規座標での係数

$n=15$, $\mu=1.5$, $g_{\text{pass}}=5.46 \times 10^{-5}$, $g_{\text{stop}}=5.85 \times 10^{-13}$

k	α_k
1	2.96820 21545 20158 E-1
2	2.96820 21559 16071 E-1
3	2.75088 15974 67332 E-1
4	2.31624 08572 14527 E-1
5	1.69794 56003 35121 E-1
6	9.63363 20742 38457 E-2
7	2.05451 20416 48405 E-2
8	-4.71689 11183 01840 E-2
9	-9.79849 96401 27541 E-2
10	-1.24548 37945 26314 E-1
11	-1.29956 87350 27408 E-1
12	-1.07402 42743 15133 E-1
13	-8.79229 80353 17280 E-2
14	-4.04631 99059 03723 E-2
15	-3.14108 29390 18306 E-2

3.2 フィルタ作用素 $\mathcal{F}$ の構成とその作用の計算法

表の数値から区間 $[a, b]$ を通過域とするフィルタ $\mathcal{F} = \text{Re} \sum_{k=1}^n \gamma_k \{\mathcal{R}(\rho)\}^k$ を作るには, 今の場合は $\tau = \sqrt{-1}$ としたこと, レゾルベントに対するシフトは $\rho = (a+b)/2 + (b-a)/2 \cdot \sqrt{-1}$ で, レゾルベント $\mathcal{R}(\rho)$ の k 乗の項の係数は $\gamma_k = \alpha_k \cdot \{(\beta - \alpha)/2\}^k (-\sqrt{-1})^k$ で与えられる (γ_k は k が偶数なら実数で k が奇数なら純虚数になることが容易に分かる).

いま X と Y が実の縦ベクトルを m 個列方向に並べた $N \times m$ の「実」行列であるときに, X にフィルタを作用させて Y を作る計算 $Y \leftarrow \mathcal{F} X$ を行うには, たとえば作業用の $N \times m$ の「複素」行列 W と Z を用いて,

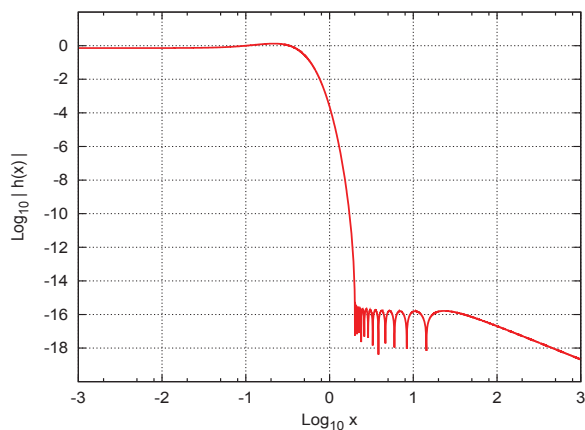


図 1 フィルタ (その 1) : 伝達関数の大きさ $|h(x)|$ (両軸対数)

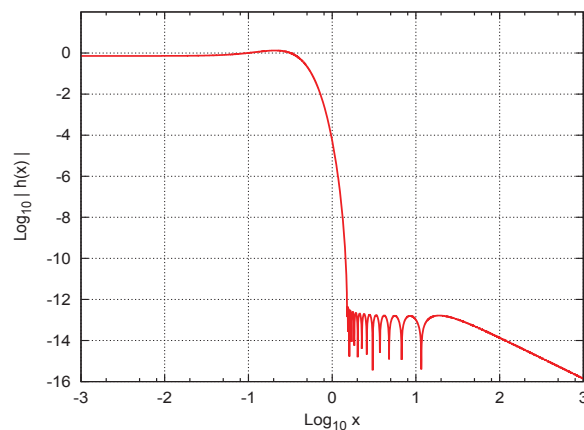


図 4 フィルタ (その 2) : 伝達関数の大きさ $|h(x)|$ (両軸対数)

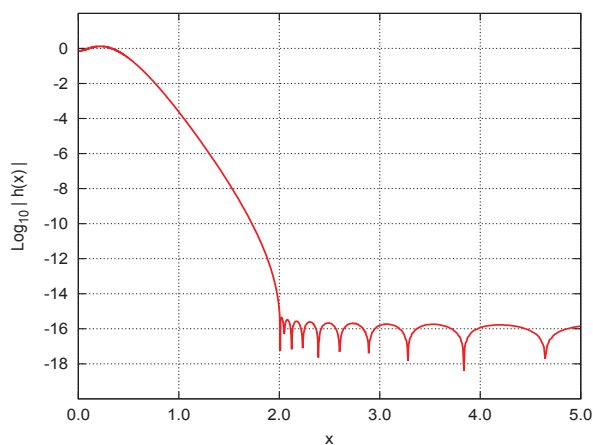


図 2 フィルタ (その 1) : 伝達関数の大きさ $|h(x)|$ (縦軸対数)

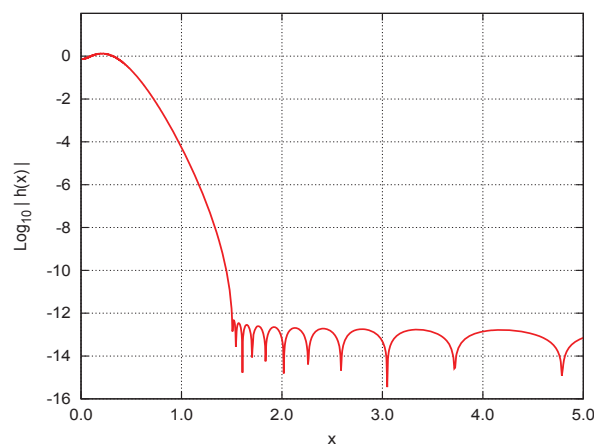


図 5 フィルタ (その 2) : 伝達関数の大きさ $|h(x)|$ (縦軸対数)

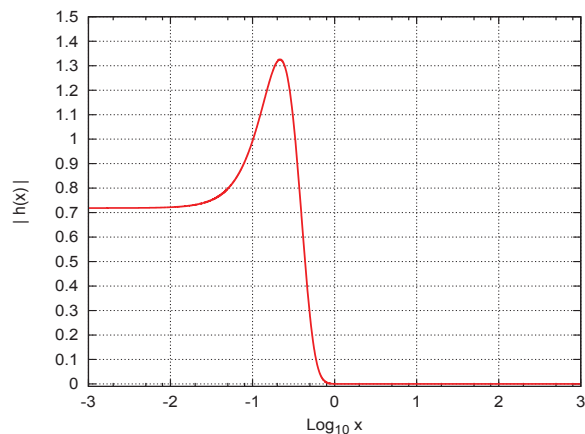


図 3 フィルタ (その 1) : 伝達関数の大きさ $|h(x)|$ (横軸対数)

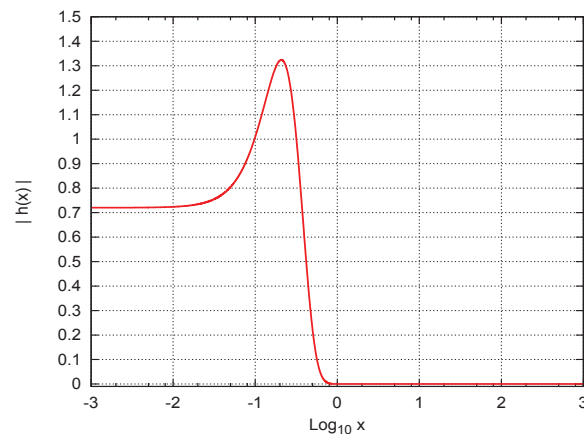


図 6 フィルタ (その 2) : 伝達関数の大きさ $|h(x)|$ (横軸対数)

```

W ← X;
Y ← 0;
for k := 1 to n do begin
  Z ← R(ρ) W;
  Y ← Y + Re(γk Z);
  W ← Z;
end

```

のようにする ($\text{Re}(\gamma_k Z)$ の計算では k が偶か奇により γ_k

が実数か純虚数になることを用いれば複素数の乗算が省ける, 代入 $W \leftarrow Z$ は最後の $k = n$ のときには省ける, また k が偶か奇により W と Z を入れ替えるコードを書くことで代入による毎回の値の移動を省く, などの小さな改良が可能である).

レゾルベントの作用の実現

上記中でのレゾルベントの作用 $Z \leftarrow \mathcal{R}(\rho) W$ の計算は, 与えられた W に対して係数 $C = A - \rho B$ の連立 1 次方程式 $CZ = BW$ の右辺 BW をまず作り, それから Z を解く

表 3 フィルタ (その 3): 正規座標での係数

$n=20, \mu=2, g_{\text{pass}}=1.273 \times 10^{-2}, g_{\text{stop}}=2.6 \times 10^{-15}$

k	α_k
1	4.83711 51618 67720 E-1
2	4.83711 51618 86980 E-1
3	3.89953 63967 72419 E-1
4	2.02437 88771 47818 E-1
5	-4.47810 12123 49263 E-2
6	-2.83593 39733 50968 E-1
7	-4.27656 83262 24258 E-1
8	-4.04019 57859 22469 E-1
9	-1.91913 38100 55309 E-1
10	1.49822 57109 15564 E-1
11	4.82023 35016 46190 E-1
12	6.49169 20356 99877 E-1
13	4.90263 91392 85137 E-1
14	1.69552 59243 54134 E-1
15	-5.78530 56855 20654 E-1
16	-3.72065 97434 12249 E-1
17	-1.44479 33647 97014 E-0
18	7.85556 89830 56973 E-1
19	-1.07607 55888 09609 E-0
20	1.70217 32211 25582 E-0

ことで行う。複素行列 C は $C^T = C$ なので対称で (複素エルミート対称では無い), A と B が帯行列のときは C も帯行列になる。

複素行列 C の対称性を利用しない場合は, 通常の帯用の LU 分解を用いて C を分解し, 分解を保持して再利用する。その際には一般に行ピボット交換を行う方が数値的な精度が良い (ただし今の場合は問題の性質からシフト ρ が虚数ならば C は数学的には正則である。シフトが実軸上の固有値に近づくとき行列 C が特異に近づくが, そのときには欲しい固有値の範囲付近の固有ベクトルの成分だけが強く拡大されるので, フィルタを構成するのが目的の場合にはピボット交換を省いてもあまり悪影響がないのではないかとと思われるが, このことは今後検討すべきである)。

複素行列 C の対称性を用いると複素版の Cholesky 法や修正 Cholesky 法が使用できる。これらについても対称性を保つピボット交換を行う場合と行わない場合があるが, 対称ピボット交換を行なうと帯幅が広がるなどにより計算の能率が悪くなるので, それならばむしろピボット交換付きの LU 分解を使うのが良いであろう (注意: Cholesky 法や修正 Cholesky 法の難点は, 行列が正則で良条件の場合でもたとえば正定値であるなどの強い条件がなければ分解途中で対角ピボットに零が生じて計算が破綻するリスクがあることである。たとえ対称なピボット交換を導入しても, よく知られた 2 次の例 $\begin{bmatrix} 0 & z \\ z & 0 \end{bmatrix}$ のように, 破綻回避が不可能な状況に遭遇する可能性もある。零でなくてもピボットに微小な値が現れると数値誤差が拡大する可能性がある。

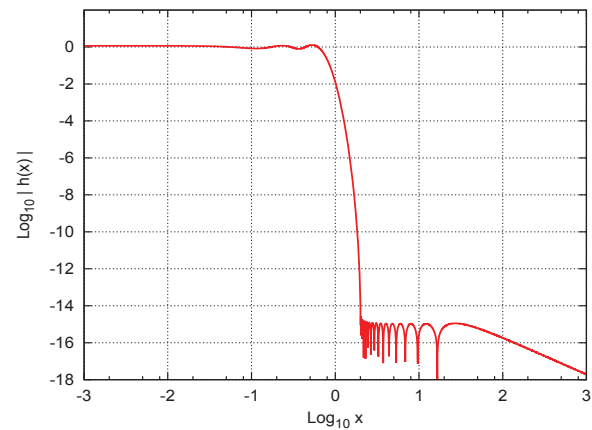


図 7 フィルタ (その 3): 伝達関数の大きさ $|h(x)|$ (両軸対数)

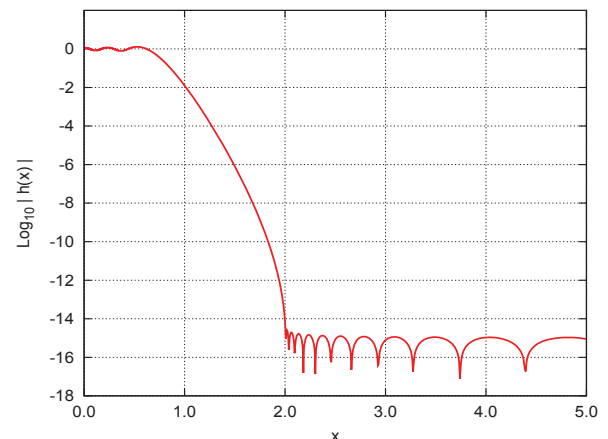


図 8 フィルタ (その 3): 伝達関数の大きさ $|h(x)|$ (縦軸対数)

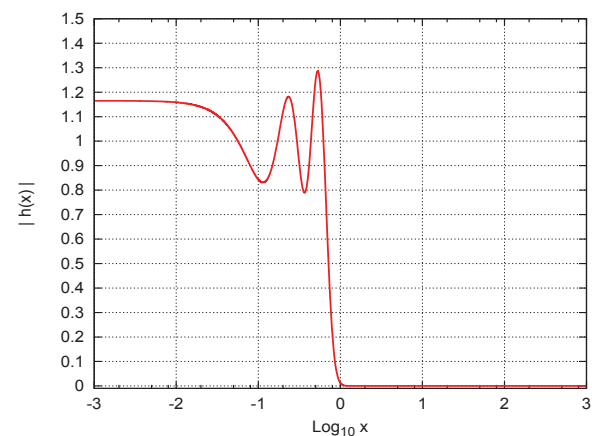


図 9 フィルタ (その 3): 伝達関数の大きさ $|h(x)|$ (横軸対数)

正則な行列でも計算が破綻する可能性がある Cholesky 法や修正 Cholesky 法の難点を回避して行列の対称性を維持しながら分解を行う方法として Bunch の方法があるが, 一般には対称ピボット交換をするたびに帯幅が広がっていくので, それならば元の対称行列の半帯幅 h に対して下帯幅 h , 上帯幅 $2h$ だけの記憶場所を確保すれば必ず実行ができる行ピボット交換付きの LU 分解を用いる方が簡単で便利であろう。

以下の実験例ではすべての場合について (上述の計算破

綻等の「リスク」はあるが), 複素行列 C が帯で対称であることから, 行列分解にはピボット交換を行わない複素版の帯用の修正 Cholesky 法を用いた。(確実な数値的安全性のためには, C の対称性を利用せずに行ピボット交換を入れた帯 LU 分解を用いるか, あるいは対称性を利用した場合には分解の残差を検査すべきであろう.)

なお, フィルタ対角化法で得られた近似対を Rayleigh 商逆反復法系の方法で改良する際には, Rayleigh 商でシフトされた行列は実対称であるが, 不定値であり極めて特異に近いので, 安全のために対称性を利用しないで行ピボット交換を行なう(実数の)帯 LU 分解を用いて解いている。

実験に用いた計算機システムの諸元

今回の実験に用いた計算機システムの環境は, CPU は Intel Core i7-5960X (3.0GHz, 8 コア, 20MB の共有 L3 キャッシュで, Turbo モードと Hyperthread 機能は BIOS からオフにしてある) で, 主記憶は Quad Channel 動作で 8G バイトのモジュール (DDR4-17000) が 8 本で合計容量 64G バイト。OS は CentOS 7.0 (64bit アドレス版) である。コンパイラには Intel Fortran ver.15.0.0 を用いた。コンパイラへのオプションは, 最適化には `-fast` を, OpenMP を用いたスレッド並列化には `-openmp` を与えた。但し今回の実験は計算法がうまく働くかを確認するのが主な目的であったので, OpenMP によるスレッド並列化やマルチコア向けのブロック分割化などの算法の並列化や最適化は十分には行っていない。計算に用いた浮動小数点数とその演算は IEEE-754 の 64 ビット倍精度である。(CPU は Haswell アーキテクチャーで Intel AVX2 拡張命令セットの融合乗加算命令を持つ)。

3.3 例題: 3 次元 FEM による Laplacian の固有値問題

今回数値実験で扱った例題は, 3 次元立方体領域 $[0, \pi] \times [0, \pi] \times [0, \pi]$ に於ける零 Dirichlet 境界条件での (符号を逆にした) Laplacian の固有値問題 $-\nabla^2 \Psi(x, y, z) = \lambda \Psi(x, y, z)$ を有限要素法 (FEM) により離散化して得られた実対称定値一般固有値問題 $A\mathbf{v} = \lambda B\mathbf{v}$ に対して, 指定した区間に固有値 λ がある固有対 $(\lambda, \mathbf{v})$ を求める問題である。有限要素内の基底関数として各座標軸方向の区分線形関数の直積を用いて, 立方体領域の各座標軸方向をそれぞれ $n_1+1=21$, $n_2+1=31$, $n_3+1=41$ に等分した有限要素を用いた。離散化で得られる一般固有値問題の係数行列 A , B の次数は $N=n_1 n_2 n_3=20 \times 30 \times 40=24,000$ である。また FEM の基底関数に適切に番号を付けると (対角線を含めない) 下帯幅は $1+n_1+n_1 n_2=1+20+20 \times 30=621$ になる (帯行列 A と B はその帯の内部はほとんど零で疎であるが, 以下の計算ではそれらをあたかも密な帯行列であるかのように扱っている。 B を乗じる場合も密な帯行列として扱っている)。そうして, 固有値が区間 $[200, 210]$ にある近似固有対 (区間内の固有対の正しい個数は 87 個) を求

めてみた。この固有値問題の固有値の正確な値は簡単な数式を用いて計算できる (付録 A.1)。各次元方向の要素分割数が異なるのでこの FEM で離散化された一般固有値問題の固有値には縮退が無い。

3.4 フィルタ対角化法について

これまでフィルタ対角化法では, 用いるフィルタ作用素の $\mathcal{F}$ の伝達関数 $f(\lambda)$ の特性を利用することで, フィルタの入力ベクトルの組 X と出力ベクトルの組 Y をうまく分析することにより, 求めたい固有値の区間に対する不変部分空間の近似空間の基底を構成することを行ってきた (論文 [1], [2], [3], [4], [8])。

本節では概略ではあるが, フィルタとして今回の「単一のレゾルベントの作用の多項式の実部」を用いる場合も, 従来の「複数のレゾルベントの作用の線形結合」を用いた場合とほとんど同様の手順で, 不変部分空間を近似する空間の基底が構成できることを述べる。

フィルタ対角化法は, (特に事前の情報が無ければ) 乱数から作った B -正規直交化された m 個のベクトルをフィルタへの入力組 X とし, フィルタを X に作用させてベクトルの組 Y を作る。そうしてこの組 X と組 Y とを合わせて, さらにフィルタの特性を考慮して必要な範囲の固有値と対応する不変部分空間を近似する空間の基底 Z を構成し, 得られた不変部分空間の近似基底 Z を Rayleigh-Ritz 法に与えて固有値問題の近似対をとりだす。

今回のフィルタ対角化法に用いた 3 種類のフィルタであるフィルタ (その 1), フィルタ (その 2), フィルタ (その 3) は, いずれも単一の虚数をシフトとするレゾルベントの多項式の実部として構成されたもので, その構成法は既に上で述べたものである。

m 次の実行列 $\beta = X^T B Y$ が対称であることは, $Y = \mathcal{F} X$ であることと, $\mathcal{F}$ がレゾルベントの多項式の実部であることから $B\mathcal{F}$ が実対称である, つまり $B\mathcal{F} = \mathcal{F}^T B$ であることを用いて示せる。またさらに β の全ての固有値が伝達関数の値域に含まれることも示せる。

β の有効階数が落ちていることは, 入力ベクトルの組 X をフィルタで濾過して得た Y の有効階数が落ちていることを意味し, フィルタへの入力として与えたベクトルの個数 m がフィルタの通過域または遷移域に固有値を持つ固有対の個数を越えた (十分な飽和に達した) ことを意味する。であるから β の固有値が閾値で切断されない (有効階数が落ちていない) 場合には, 入力ベクトルの個数 m は十分ではない。もしも閾値を大きい値に変更すれば切断が生じるのであれば, それに見合う低精度の演算では飽和に達していることになるが, それは実効的にフィルタの阻止域の定義を変更して g_{stop} の値を大きい値に変更したことと同様で, 精度の良い近似解は得られない。

m 次の行列 $\alpha = Y^T B Y$ は実対称半正定値であり, それ

と β から m 次の (実対称一般) 固有値方程式 $\alpha \mathbf{u} = \phi \beta \mathbf{u}$ を作る. すると元の固有値問題のある固有対 $(\lambda, \mathbf{v})$ の固有ベクトル $\mathbf{v}$ が $\text{span}(Y)$ に含まれるという近似が成立すると $\mathbf{v} = Y\mathbf{u}$ と書ける. そのことから $\phi = f(\lambda)$ の関係が出てくる. λ の通過域 $[a, b]$ に対する伝達関数の値域 $f([a, b])$ (閾値が g_{pass} である) に含まれる ϕ と対応する m 次の固有値問題の固有ベクトル $\mathbf{u}$ を全て並べた行列を U とし, それから YU を作ると, YU は元の固有値問題の通過域に対応する不変部分空間の基底の近似になる.

実際には数値誤差や近似の程度を考慮して安全マージンを入れて $f([a, b])$ を拡大するようにフィルタの通過域に於ける伝達率の最小値 g_{pass} よりも少し小さい値を ϕ の閾値に設定する. その閾値は固有値 ϕ の値分布を見てから決めることができ, ある程度広い幅を持った空隙を ϕ の値分布の中に見つけて閾値をそこに設定すれば良い (空隙の上側にある値を持つ ϕ とそれに対応する固有ベクトル $\mathbf{u}$ を選択する).

3.5 数値実験例

「単一の虚数シフトのレゾルベントの作用の多項式の実部」で構成したフィルタを用いた対角化法によりうまく近似固有対を構成できるかについて調べるために数値実験を行った.

以下の各実験の例では, 図 (グラフ) をそれぞれ 4 枚示しているが, あらかじめここでそれらのグラフの内容について説明しておく.

(1) 図「 β の固有値の分布」のグラフ:

第 3.4 節中の「 m 次行列 β 」の固有値を, 減少順につけた順位を横軸にとり, 縦軸にはその絶対値の対数をとって, 固有値の符号が正のものは赤色で, 負のものは青色でプロットしている. 緑色の水平線は, 有効階数を設定するための閾値で, 倍精度計算では適当であると思われるマシンイプシロンの 100 倍の値 2.2×10^{-14} を示している.

(2) 図「 $|\phi|$ の値の分布」のグラフ:

第 3.4 節中の「 m 次に縮小された固有値問題 $\alpha \mathbf{u} = \phi \beta \mathbf{u}$ 」の固有値 ϕ を減少順につけた順位を横軸にとり, 縦軸にはその ϕ の大きさの対数をとって, ϕ の符号が正のものは赤色で, 負のものは青色でプロットしている. 緑色の水平線は通過域の値域よりも少し小さい値として設定された閾値である.

(3) 図「近似対の残差のノルム Δ 」のグラフ:

あらかじめベクトルが B -正規化されている近似固有対 $(\lambda, \mathbf{v})$ の残差 $\mathbf{r} = (A - \lambda B)\mathbf{v}$ に対する一般化された Wilkinson の限界 $\Delta = \sqrt{\mathbf{r}^T B^{-1} \mathbf{r}}$ を「ノルム」に用いている. これは近似対の近似固有値の誤差限界を与える. 計算法は, 例えば正定値対称行列 B の Cholesky 分解を $B = LL^T$ とすると, $B^{-1} = L^{-T}L^{-1}$ であるか

ら $\Delta = \sqrt{\mathbf{r}^T B^{-1} \mathbf{r}} = \|L^{-1}\mathbf{r}\|_2$ となる (但し, 残差が微小な場合には, 残差 $\mathbf{r}$ を通常よりも高精度演算で求め, 残差から Δ を求める過程も高精度で計算をしないと, 丸め誤差の影響が最後に平方根で拡大されて評価が過大になるであろう. 但し今回の数値計算では, 残差のノルム Δ の計算は通常の倍精度演算だけを用いて行なった). 赤色の折れ線 IT0 はフィルタ対角化法で得られた近似対の残差のノルムを横軸を近似固有値にとってプロットしたものである. 緑色の折れ線 IT1, 青色の折れ線 IT2 はそれぞれ IT0 の近似対から逆反復法により 1 回, 2 回の修正を行って得られた近似対に対する残差のノルムを同様にプロットしたものである.

(4) 図「近似固有対の絶対誤差」のグラフ:

厳密値を与える式から計算して求めた真の固有値とフィルタ対角化法で求めた近似固有値との差の大きさを縦軸に, 各近似固有値を横軸にとってプロットしたものである.

3.5.1 フィルタ (その 1) による数値実験例

● $m = 100$ の場合

フィルタ (その 1) を用いて, 乱数から作成した $m = 100$ 個のベクトルを入力として計算を行った例を図 10, 図 11, 図 12, 図 13 に示す.

「 β の固有値の分布」のグラフ (図 10) から, $m = 100$ ではまだ飽和に達していないことがわかる.

「 $|\phi|$ の値の分布」のグラフ (図 11) では, 切断は 2.3×10^{-4} で行われ, 閾値の上側の ϕ の値は真の固有対の個数と一致する 87 個であった.

また「近似対の残差のノルム Δ 」のグラフ (図 12) を見ればフィルタ対角化法による近似固有値の誤差が区間の中央付近では 10^{-4} 程度以下であり, 少なくとも 6 桁程度以上の精度があるはずであるが, 区間の端付近では 10^{-0} 程度以下であり, 少なくとも 2 桁程度以上の精度があるはずであることがわかる.

また真の固有値の値を知らなければ描けないグラフではあるが「近似固有値の絶対誤差」(図 13) から区間の中央部では絶対誤差は 10^{-10} 程度で近似固有値の実際の精度は 12 桁程度であるが, 区間の端では絶対誤差が 10^{-2} 程度あって近似固有値の実際の精度は 4 桁程度しかないことがわかる.

● $m = 150$ の場合

フィルタ (その 1) を用いて, 乱数から作成した $m = 150$ 個のベクトルを入力として計算を行った例を図 14, 図 15, 図 16, 図 17 に示す.

「 β の固有値の分布」のグラフ (図 14) から, $m = 150$ では既に飽和に達していることがわかる.

「 $|\phi|$ の値の分布」のグラフ (図 15) で, 切断は 2.3×10^{-4}

で行われ、閾値の上側の ϕ の値は真の固有対の個数と一致する 87 個であった。

「近似対の残差のノルム Δ 」のグラフ (図 16) を見ればフィルタ対角化法による近似対の近似固有値の誤差が区間の端では 10^{-6} 程度以下であり、(固有値の値自身が 2 桁の数値であるから) 少なくとも 8 桁程度の精度があるはずで、区間の中央付近では 10^{-10} 程度以下であり、少なくとも 12 桁程度の精度があるはずであることがわかる。

また真の固有値の値を知らなければ描けないグラフではあるが「近似固有値の絶対誤差」(図 17) から絶対誤差で 10^{-13} 程度、つまり固有値が約 15 桁の精度で求まっていることが確認できる。

- $m = 200$ の場合

フィルタ (その 1) を用いて、乱数から作成した $m = 200$ 個のベクトルを入力として計算を行った例を図 18, 図 19, 図 20, 図 21 に示す。

「 β の固有値の分布」のグラフ (図 18) から、 $m = 200$ では既に飽和に達していることがわかる。

「 $|\phi|$ の値の分布」のグラフ (図 19) で、切断は 2.3×10^{-4} で行われ、閾値の上側の ϕ の値は真の固有対の個数と一致する 87 個であった。

「近似対の残差のノルム Δ 」のグラフ (図 20) を見ればフィルタ対角化法による近似対の近似固有値の誤差が区間の端では 10^{-7} 程度以下で少なくとも 9 桁程度の精度があり、区間の中央付近では 10^{-12} から 10^{-11} 程度以下で、少なくとも 13 桁程度の精度があることがわかる。

また真の固有値の値を知らなければ描けないグラフではあるが「近似固有値の絶対誤差」(図 21) から絶対誤差は 10^{-13} 程度、つまり近似固有値は約 15 桁の精度で求まっていることが確認できる。

フィルタ (その 1) を用いた計算処理の経過時間の内訳を、 $m = 100$, $m = 150$, $m = 200$ のそれぞれの場合について表 4 に掲げる。

3.5.2 フィルタ (その 2) による数値実験例

- $m = 100$ の場合

フィルタ (その 2) を用いて、乱数から作成した $m = 100$ 個のベクトルを入力として計算を行った例を図 22, 図 23, 図 24, 図 25 に示す。

「 β の固有値の分布」のグラフ (図 22) から、 $m = 100$ ではまだ飽和に達していないことがわかる。

「 $|\phi|$ の値の分布」のグラフ (図 23) で、切断は 5.46471×10^{-5} で行われ、閾値の上側の ϕ の値は真の固有対の個数と一致する 87 個であった。

また「近似対の残差のノルム Δ 」のグラフ (図 24) を見ればフィルタ対角化法による近似固有値の誤差が区間の中央付近では 10^{-6} 程度以下で、少なくとも 8 桁

程度以上の精度があるが、区間の端付近では 10^{-1} 程度以下で、少なくとも 3 桁程度以上の精度しかないことがわかる。

また真の固有値の値を知らなければ描けないグラフではあるが「近似固有値の絶対誤差」(図 25) から区間の中央部では絶対誤差は 10^{-12} 程度で、近似固有値の実際の精度が 14 桁程度あるが、区間の端では絶対誤差が 10^{-3} 程度あり、近似固有値の実際の精度は 5 桁程度しかないことがわかる。

- $m = 150$ の場合

フィルタ (その 2) を用いて、乱数から作成した $m = 150$ 個のベクトルを入力として計算を行った例を図 26, 図 27, 図 28, 図 29 に示す。

「 β の固有値の分布」のグラフ (図 26) から、 $m = 150$ では既に飽和に達していることがわかる。

「 $|\phi|$ の値の分布」のグラフ (図 27) で、切断は 5.46471×10^{-5} で行われ、閾値の上側の ϕ の値は真の固有対の個数と一致する 87 個であった。

「近似対の残差のノルム Δ 」のグラフ (図 28) を見ればフィルタ対角化法による近似対の近似固有値の誤差が区間の端では 10^{-6} 程度以下で少なくとも 8 桁程度の精度があり、区間の中央付近では 10^{-10} 程度以下で少なくとも 12 桁程度の精度があることがわかる。

また真の固有値の値を知らなければ描けないグラフではあるが「近似固有値の絶対誤差」(図 29) から絶対誤差は 10^{-13} 程度で、つまり近似固有値が実際に約 15 桁の精度で求まっていることが確認できる。

- $m = 200$ の場合

フィルタ (その 2) を用いて、乱数から作成した $m = 200$ 個のベクトルを入力として計算を行った例を図 30, 図 31, 図 32, 図 33 に示す。

「 β の固有値の分布」のグラフ (図 30) から、 $m = 200$ では既に飽和に達していることがわかる。

「 $|\phi|$ の値の分布」のグラフ (図 31) で、切断は 5.46471×10^{-5} で行われ、閾値の上側の ϕ の値は真の固有対の個数と一致する 87 個であった。

「近似対の残差のノルム Δ 」のグラフ (図 32) を見ればフィルタ対角化法による近似対の近似固有値の誤差が区間の端では 10^{-6} 程度以下で少なくとも 8 桁程度の精度があり、区間の中央付近では 10^{-10} 程度以下で、少なくとも 12 桁程度の精度があることがわかる。また真の固有値の値を知らなければ描けないグラフではあるが「近似固有値の絶対誤差」(図 33) から絶対誤差は 10^{-13} 程度で、つまり近似固有値は実際に約 15 桁の精度で求まっていることが確認できる。

フィルタ (その 2) を用いた計算処理の経過時間の内訳を、 $m = 100$, $m = 150$, $m = 200$ のそれぞれの場合について表 5 に掲げる。

3.5.3 フィルタ（その 3）による数値実験例

- $m = 100$ の場合

フィルタ（その 3）を用いて、乱数から作成した $m = 100$ 個のベクトルを入力として計算を行った例を図 34, 図 35, 図 36, 図 37 に示す。

「 β の固有値の分布」のグラフ（図 34）から、 $m = 100$ ではまだ飽和に達していないことがわかる。

「 $|\phi|$ の値の分布」のグラフ（図 35）で、切断は 1.27268×10^{-2} で行われ、閾値の上側の ϕ の値は真の固有対の個数と一致する 87 個であった。

また「近似対の残差のノルム Δ 」のグラフ（図 36）を見ればフィルタ対角化法による近似固有値の誤差が区間の中央付近では 10^{-3} から 10^{-2} 程度以下で少なくとも 4 桁から 5 桁程度以上の精度があることになるが、区間の端付近では 10^{-0} 程度以下で、少なくとも 2 桁程度以上の精度であることがわかる。

また真の固有値の値を知らなければ描けないグラフではあるが「近似固有値の絶対誤差」（図 37）から区間の中央部では絶対誤差は 10^{-6} 程度で、近似固有値の実際の精度が 8 桁程度あることがわかるが、区間の端では絶対誤差が 10^{-2} 程度あり、近似固有値の実際の精度が 4 桁程度しかないことがわかる。

- $m = 150$ の場合

フィルタ（その 3）を用いて、乱数から作成した $m = 150$ 個のベクトルを入力として計算を行った例を図 38, 図 39, 図 40, 図 41 に示す。

「 β の固有値の分布」のグラフ（図 38）から、 $m = 150$ では既に飽和に達していることがわかる。

「 $|\phi|$ の値の分布」のグラフ（図 39）で、切断は 1.27268×10^{-2} で行われ、閾値の上側の ϕ の値は真の固有対の個数と一致する 87 個であった。

「近似対の残差のノルム Δ 」のグラフ（図 40）を見ればフィルタ対角化法による近似対の近似固有値の誤差が区間の端では 10^{-7} 程度以下で少なくとも 9 桁程度の精度があり、区間の中央付近では 10^{-9} 程度以下で少なくとも 11 桁程度の精度があることがわかる。

また真の固有値の値を知らなければ描けないグラフではあるが「近似固有値の絶対誤差」（図 41）から絶対誤差は 10^{-13} 程度で、つまり近似固有値は実際には約 15 桁の精度で求まっていることが確認できる。

- $m = 200$ の場合

フィルタ（その 3）を用いて、乱数から作成した $m = 200$ 個のベクトルを入力として計算を行った例を図 42, 図 43, 図 44, 図 45 に示す。

「 β の固有値の分布」のグラフ（図 42）から、 $m = 200$ では既に飽和に達していることがわかる。

「 $|\phi|$ の値の分布」のグラフ（図 43）で、切断は 1.27268×10^{-2} で行われ、閾値の上側の ϕ の値は真

の固有対の個数と一致する 87 個であった。

「近似対の残差のノルム Δ 」のグラフ（図 44）を見ればフィルタ対角化法による近似対の近似固有値の誤差が区間の端では 10^{-10} 程度以下で、少なくとも 12 桁程度の精度があり、区間の中央付近では 10^{-12} から 10^{-11} 程度以下で、少なくとも 13 桁程度の精度があることがわかる。

また真の固有値の値を知らなければ描けないグラフではあるが「近似固有値の絶対誤差」（図 45）から絶対誤差はどれも 10^{-13} 程度で、つまり近似固有値が実際に約 15 桁の精度で求まっていることが確認できる。

フィルタ（その 3）を用いた計算処理の経過時間の内訳を、 $m = 100$, $m = 150$, $m = 200$ のそれぞれの場合について表 6 に掲げる。

表 4 フィルタ（その 1）：処理の経過時間の内訳（秒）

処理ステップ	$m=100$	$m=150$	$m=200$
フィルタ対角化全体	13.32	19.08	24.26
- 乱数ベクトル生成	0.01	0.02	0.03
- 入力 B -正規直交化	0.31	0.48	0.65
- フィルタの適用	11.35	16.19	21.12
- 不変部分空間の基底作成	1.11	1.85	1.93
- Rayleigh-Ritz 法	0.53	0.53	0.53
Rayleigh 商逆反復各 2 回	1,040.83	1,046.21	1,048.69

表 5 フィルタ（その 2）：処理の経過時間の内訳（秒）

処理ステップ	$m=100$	$m=150$	$m=200$
フィルタ対角化全体	13.31	19.12	24.50
- 乱数ベクトル生成	0.01	0.02	0.03
- 入力 B -正規直交化	0.31	0.48	0.65
- フィルタの適用	11.34	16.21	21.12
- 不変部分空間の基底作成	1.12	1.88	2.17
- Rayleigh-Ritz 法	0.53	0.53	0.53
Rayleigh 商逆反復各 2 回	1,051.75	1,047.60	1,050.45

表 6 フィルタ（その 3）：処理の経過時間の内訳（秒）

処理ステップ	$m=100$	$m=150$	$m=200$
フィルタ対角化全体	17.46	23.72	30.73
- 乱数ベクトル生成	0.01	0.02	0.03
- 入力 B -正規直交化	0.31	0.48	0.65
- フィルタの適用	14.96	21.38	27.94
- 不変部分空間の基底作成	1.65	1.32	1.58
- Rayleigh-Ritz 法	0.53	0.53	0.53
Rayleigh 商逆反復各 2 回	1,051.19	1,051.60	1,052.23

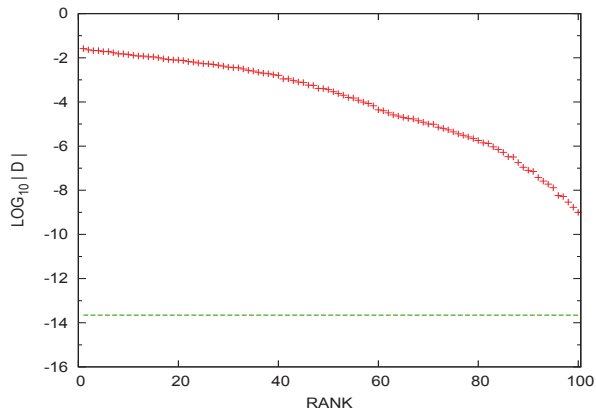


図 10 (その 1) β の固有値の分布 ($m=100$, 閾値 2.2×10^{-14})

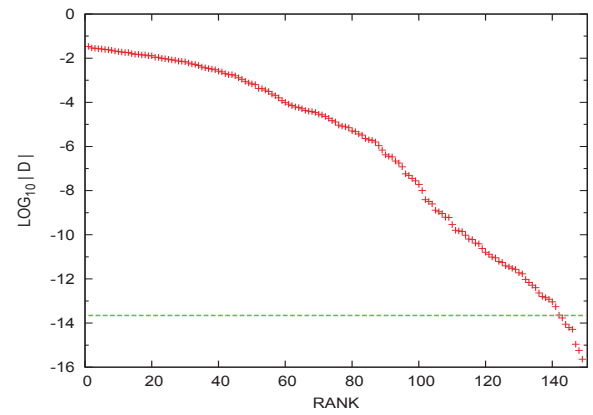


図 14 (その 1) β の固有値の分布 ($m=150$, 閾値 2.2×10^{-14})

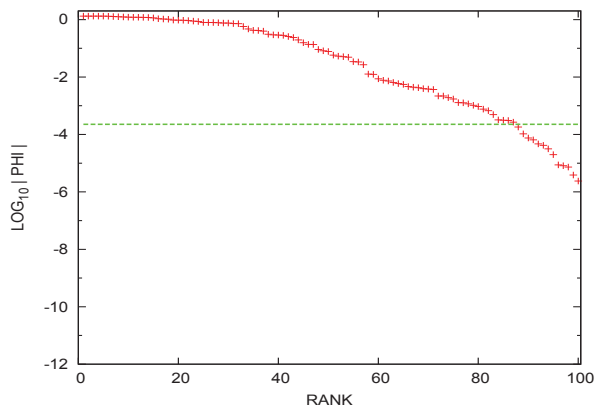


図 11 (その 1) $|\phi|$ の値の分布 ($m=100$, 閾値 2.3×10^{-4})

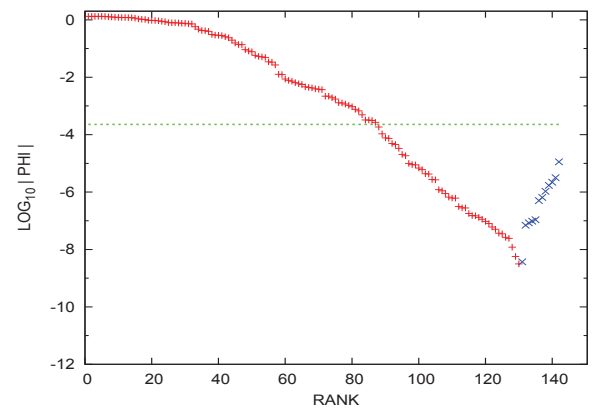


図 15 (その 1) $|\phi|$ の値の分布 ($m=150$, 閾値 2.3×10^{-4})

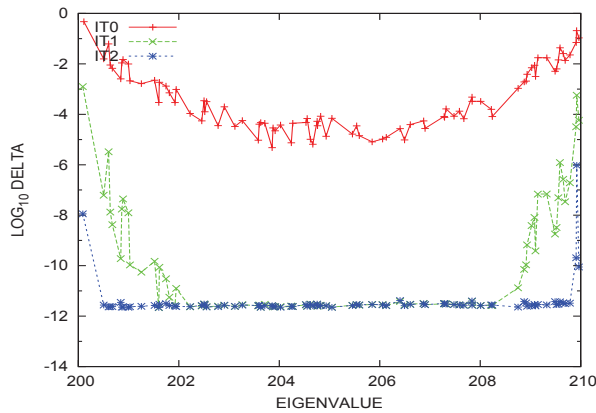


図 12 (その 1) 近似対の残差のノルム Δ (区間 $[200, 210]$, $m=100$)

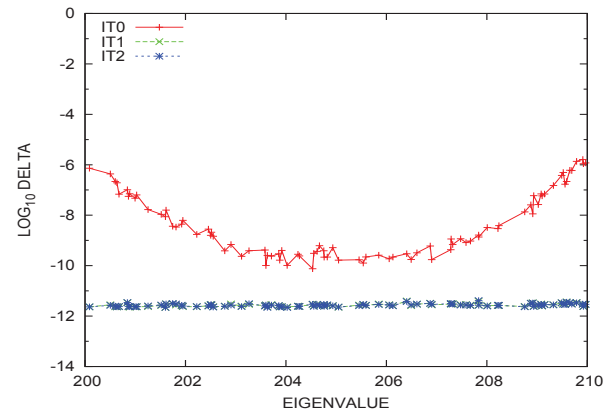


図 16 (その 1) 近似対の残差のノルム Δ (区間 $[200, 210]$, $m=150$)

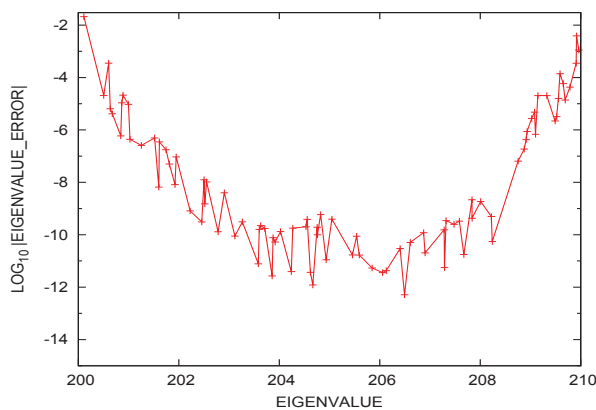


図 13 (その 1) 近似固有値の絶対誤差 ($m=100$, ϕ の閾値 2.3×10^{-4})

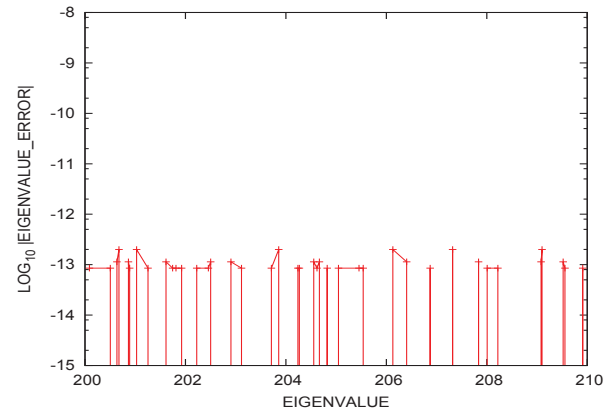


図 17 (その 1) 近似固有値の絶対誤差 ($m=150$, ϕ の閾値 2.3×10^{-4})

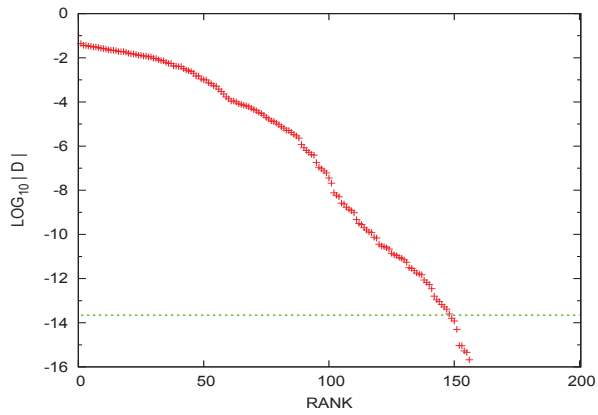


図 18 (その 1) β の固有値の分布 ($m=200$, 閾値 2.2×10^{-14})

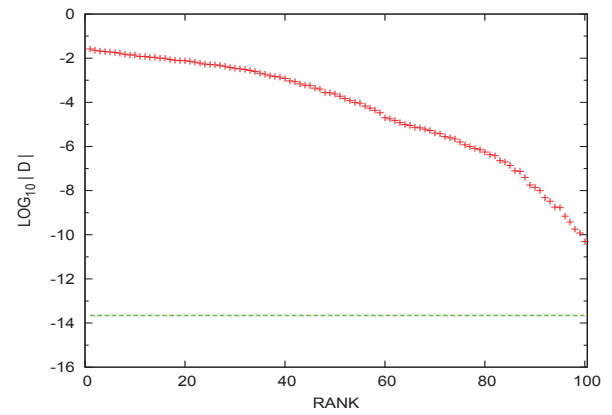


図 22 (その 2) β の固有値の分布 ($m=100$, 閾値 2.2×10^{-14})

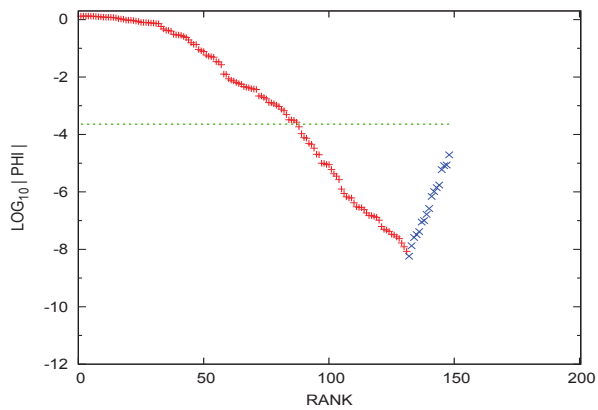


図 19 (その 1) $|\phi|$ の値の分布 ($m=200$, 閾値 2.3×10^{-4})

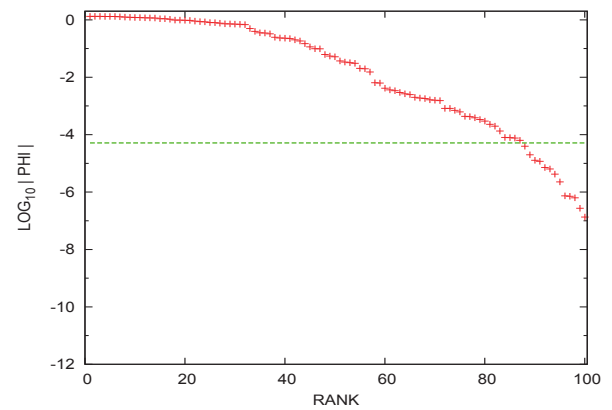


図 23 (その 2) $|\phi|$ の値の分布 ($m=100$, 閾値 5.465×10^{-5})

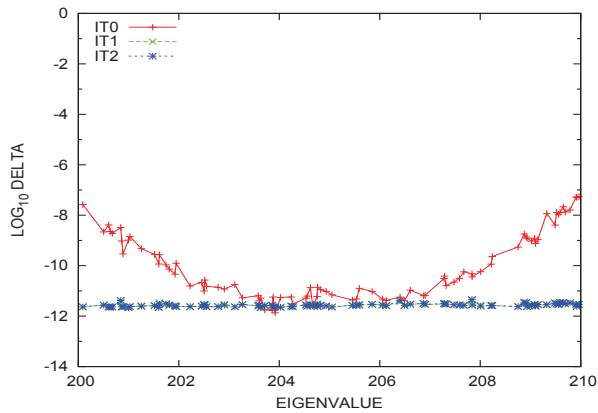


図 20 (その 1) 近似対の残差のノルム Δ (区間 $[200, 210]$, $m=200$)

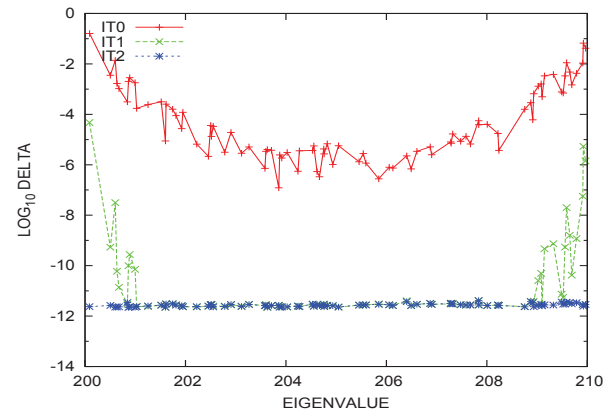


図 24 (その 2) 近似対の残差のノルム Δ (区間 $[200, 210]$, $m=100$)

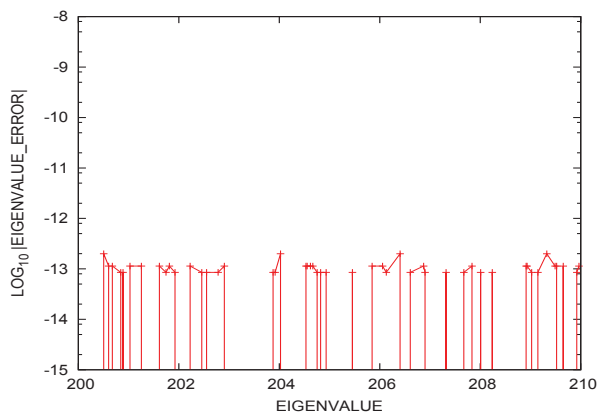


図 21 (その 1) 近似固有値の絶対誤差 ($m=200$, ϕ の閾値 2.3×10^{-4})

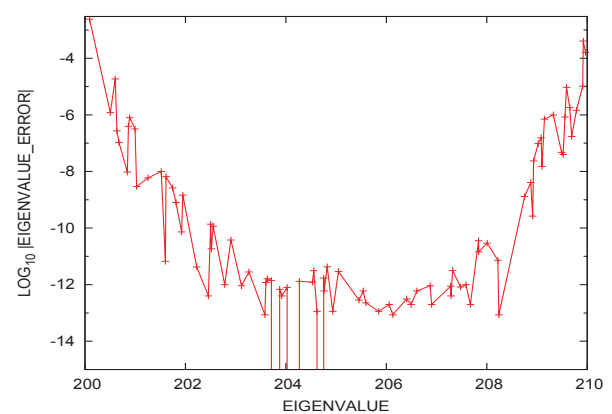


図 25 (その 2) 近似固有値の絶対誤差 ($m=100$, ϕ の閾値 5.465×10^{-5})

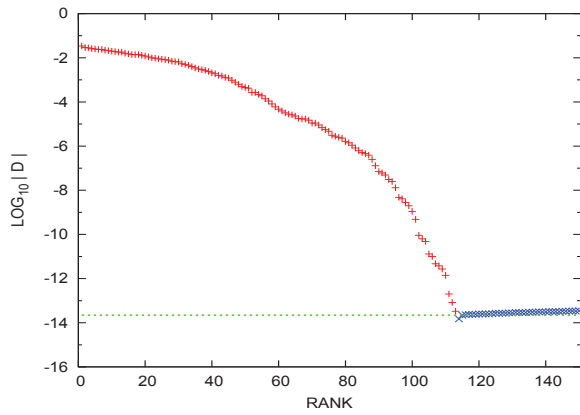


図 26 (その 2) β の固有値の分布 ($m=150$, 閾値 2.2×10^{-14})

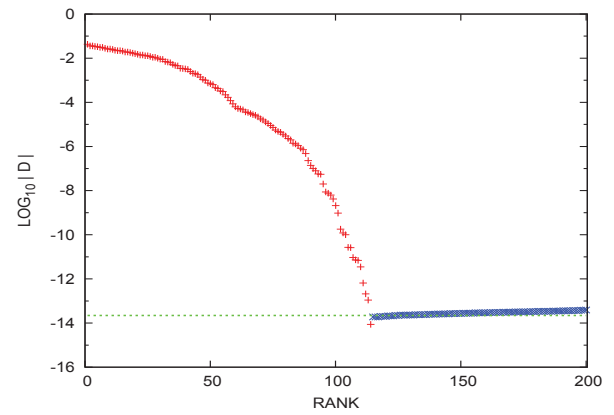


図 30 (その 2) β の固有値の分布 ($m=200$, 閾値 2.2×10^{-14})

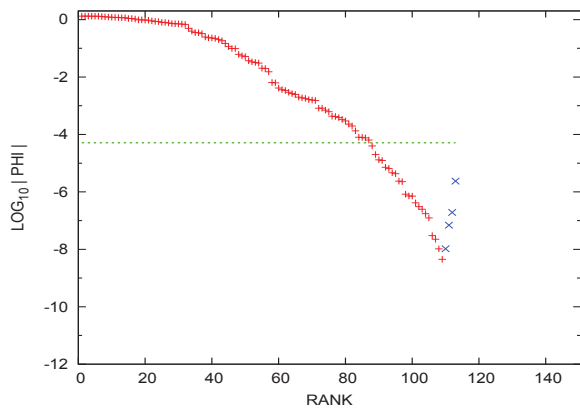


図 27 (その 2) $|\phi|$ の値の分布 ($m=150$, 閾値 5.465×10^{-5})

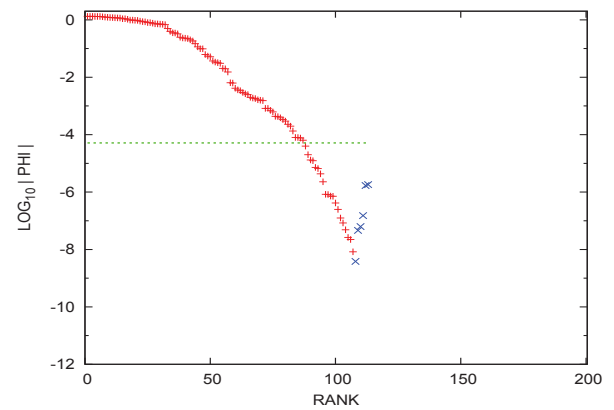


図 31 (その 2) $|\phi|$ の値の分布 ($m=200$, 閾値 5.465×10^{-5})

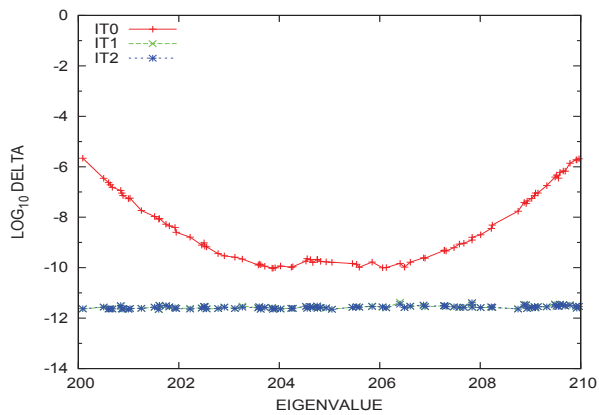


図 28 (その 2) 近似対の残差のノルム Δ (区間 $[200, 210]$, $m=150$)

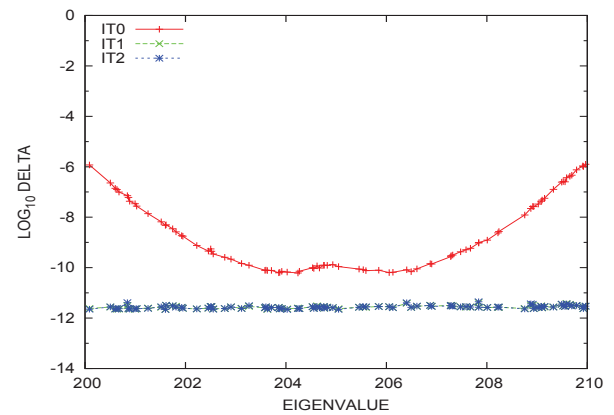


図 32 (その 2) 近似対の残差のノルム Δ (区間 $[200, 210]$, $m=200$)

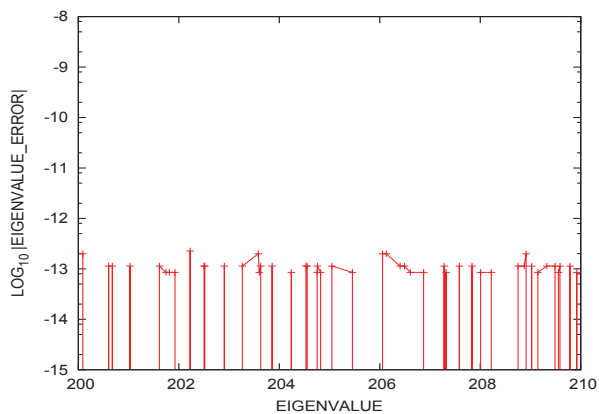


図 29 (その 2) 近似固有値の絶対誤差 ($m=150$, ϕ の閾値 5.465×10^{-5})

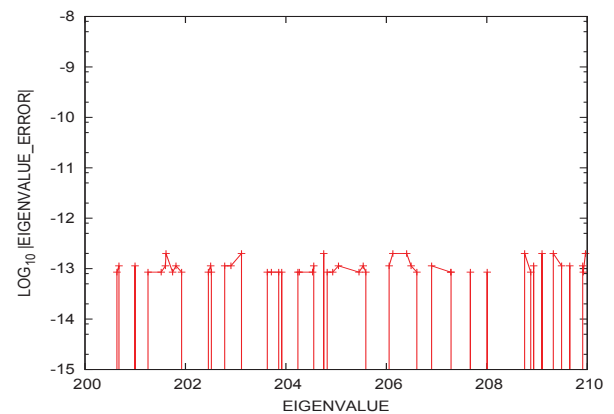


図 33 (その 2) 近似固有値の絶対誤差 ($m=200$, ϕ の閾値 5.465×10^{-5})

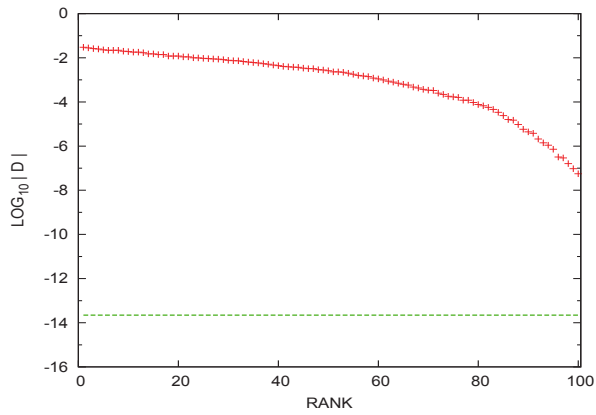


図 34 (その 3) β の固有値の分布 ($m=100$, 閾値 2.2×10^{-14})

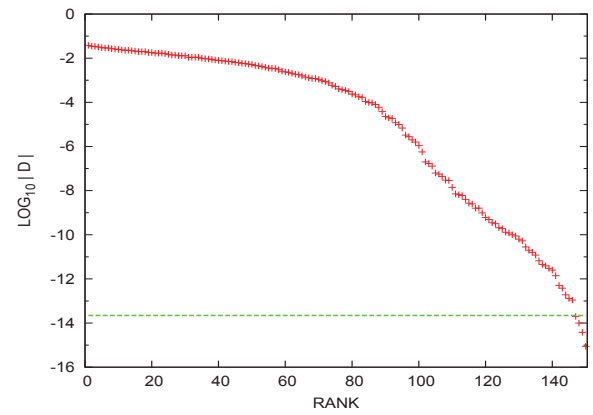


図 38 (その 3) β の固有値の分布 ($m=150$, 閾値 2.2×10^{-14})

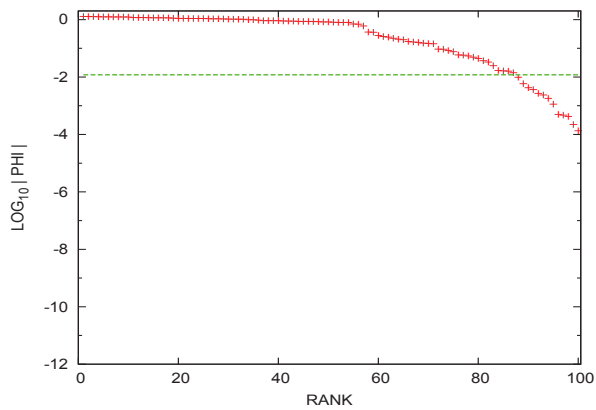


図 35 (その 3) $|\phi|$ の値の分布 ($m=100$, 閾値 1.273×10^{-2})

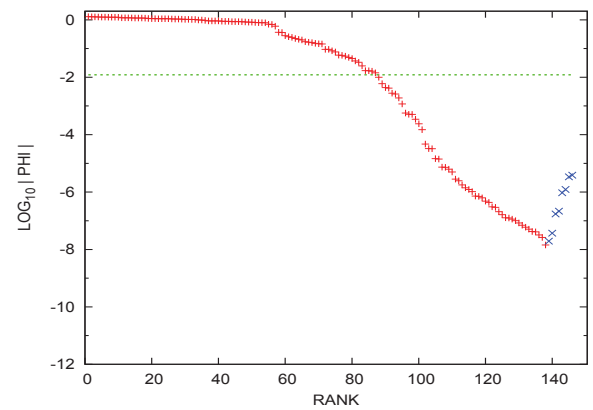


図 39 (その 3) $|\phi|$ の値の分布 ($m=150$, 閾値 1.273×10^{-2})

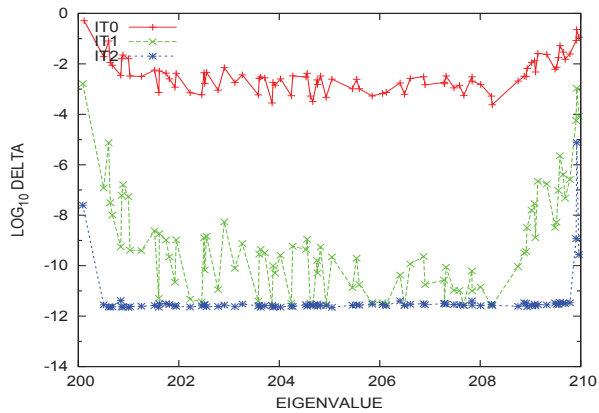


図 36 (その 3) 近似対の残差のノルム Δ (区間 $[200, 210]$, $m=100$)

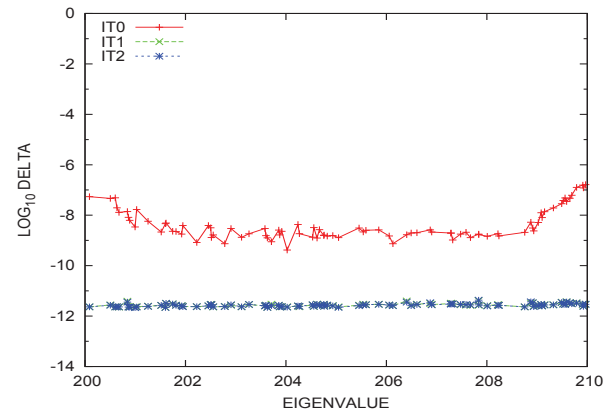


図 40 (その 3) 近似対の残差のノルム Δ (区間 $[200, 210]$, $m=150$)

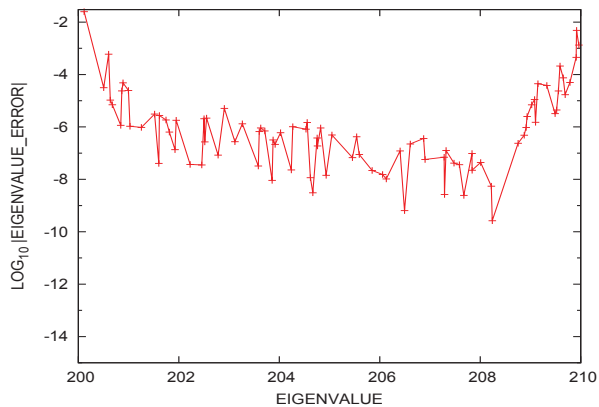


図 37 (その 3) 近似固有値の絶対誤差 ($m=100$, ϕ の閾値 1.273×10^{-2})

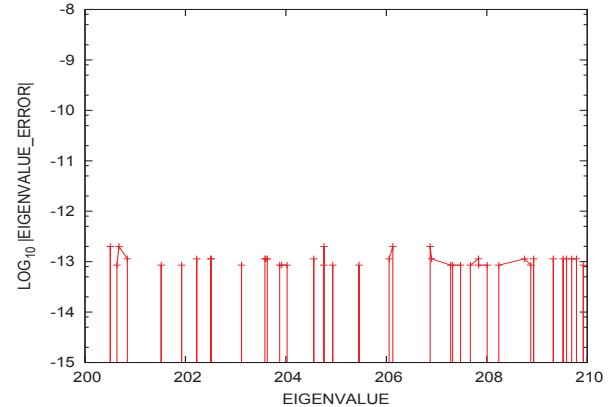


図 41 (その 3) 近似固有値の絶対誤差 ($m=150$, ϕ の閾値 1.273×10^{-2})

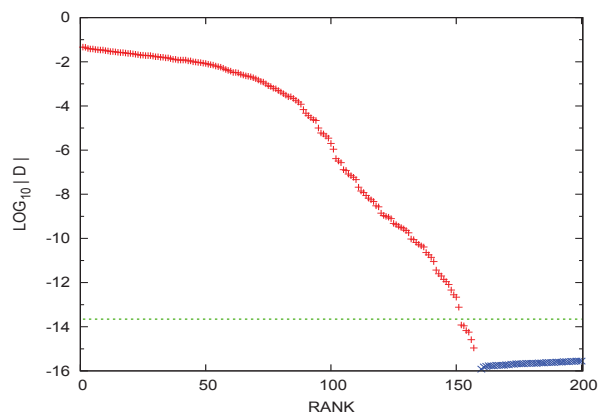


図 42 (その 3) β の固有値の分布 ($m=200$, 閾値 2.2×10^{-14})

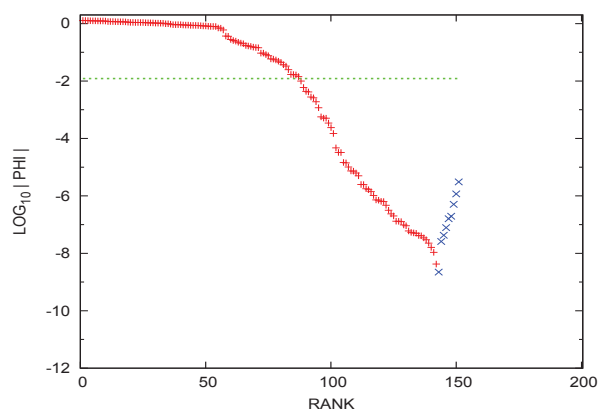


図 43 (その 3) $|\phi|$ の値の分布 ($m=200$, 閾値 1.273×10^{-2})

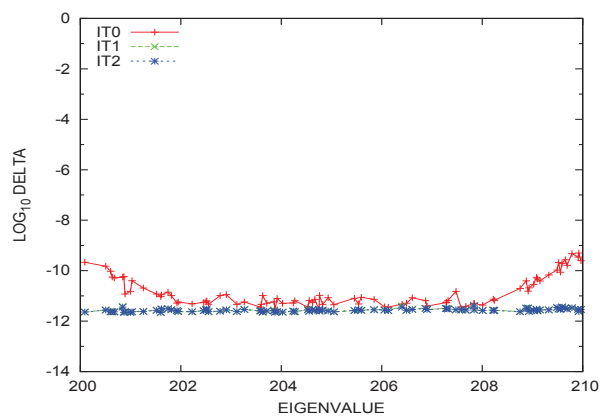


図 44 (その 3) 近似対の残差のノルム Δ (区間 $[200, 210]$, $m=200$)

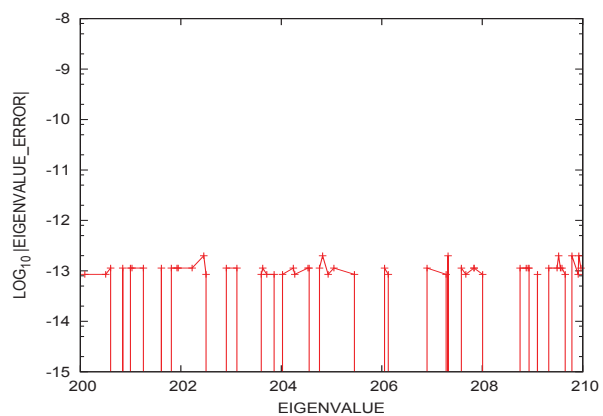


図 45 (その 3) 近似固有値の絶対誤差 ($m=200$, ϕ の閾値 1.273×10^{-2})

4. おわりに

実対称定値一般固有値問題の指定された区間に固有値がある固有対をフィルタ対角化法を用いて求める際に必要となる記憶量と演算量の低減を狙って、シフトが虚数のレゾルベントを1つだけ用いて、レゾルベントの多項式の実部としてフィルタ作用素を構成する方法を示し、そのフィルタを実際に用いた数値対角化の実験を行って妥当な結果を得た。

この方法は複数の異なるシフトを持つレゾルベントの線形結合をフィルタとして用いる場合に比べると、潜在的な並列性は低下するが、レゾルベントを与える連立1次方程式を直接法で解く場合に必要となる行列の分解、あるいは反復法で解く場合の行列の不完全分解がただ1通りで済み、1度行った行列の分解あるいは不完全分解の結果を再利用することでレゾルベントの多項式的作用を実現できるので、記憶量と共に計算量も少なく済むことが期待できる。

今回の実験で用いたレゾルベントの多項式の次数は $n = 15$ あるいは $n = 20$ であり、これらの多項式の係数の数値を決めるのには最小2乗法の系統の手法を用いて、悪条件の連立1次方程式を切断を導入して解いたものである(この部分の計算では四倍精度演算を利用して求めている)。今後は最小2乗近似の代わりに最良近似に基づく係数の決定法を試すことや、単なる数値による最適化よりも見通しの良い数学的な方法が無いかなどについても考察し、さらに高次のレゾルベントの多項式を用いて特性の良いフィルタを構成できないかについても調べてみたい。またほかの方向性として、伝達関数の持つ極を今回のような共役な虚数1対だけとする場合から複数の対を極とする場合に拡張し、複数のレゾルベントの多項式的作用によりフィルタを構成することなども考えられる。

付 録

A.1 FEM による 3 次元 Laplacian の固有値

1次元の区間 $[0, \pi]$ に於ける零 Dirichlet 境界条件の Laplacian (但し符号を逆にした $-\nabla^2$) の FEM 離散化による固有値は、区間 $[0, \pi]$ を $n+1$ 等分した幅 $h = \pi/(n+1)$ の小区間による分割の区分線形関数を基底とする場合は、 $E(n, k)$, $k=1, 2, \dots, n$ で与えられる。ここで、 $\theta_k = hk$ とおいて $E(n, k) = k^2 \{ \sin(\theta_k)/\theta_k \}^2 / \{ (1 + \cos \theta_k)(2 + \cos \theta_k)/6 \}$ である。そうして3次元領域 $[0, \pi]^3$ で各方向の区間をそれぞれ n_1+1 , n_2+1 , n_3+1 に等分して各方向の1次元の基底の直積型の基底関数を用いる場合の FEM では、 $k_i=1, 2, \dots, n_i$ とするときに (k_1, k_2, k_3) により識別される3次元 Laplacian ($-\nabla^2$) の固有値 $E((n_1, n_2, n_3), (k_1, k_2, k_3))$ の値は1次元の場合の固有値の和の形で、数式 $E(n_1, k_1) + E(n_2, k_2) + E(n_3, k_3)$ により与えられる。

参考文献

- [1] 村上弘: 固有値が指定された区間内にある固有対を解くための対称固有値問題用のフィルタの設計, **情報処理学会論文誌: コンピューティングシステム (ACS31)**, Vol.3, No.3 (2010 年 9 月), pp.1-21.
- [2] 村上弘: 対称一般固有値問題のフィルタ作用素を用いた不変部分空間の近似構成, **情報処理学会論文誌: コンピューティングシステム (ACS35)**, Vol.4, No.4 (2011 年 10 月), pp.1-14.
- [3] 村上弘: Hermite 対称な定値一般固有値問題のフィルタ対角化法について, **情報処理学会研究報告**, Vol.2012-HPC-134, No.1 (2012 年 6 月), pp.1-8.
- [4] 村上弘: レゾルベントの線形結合をフィルタに用いたエルミート定値一般固有値問題のフィルタ対角化法, **情報処理学会論文誌: コンピューティングシステム (ACS45)**, Vol.7, No.1 (2014 年 3 月), pp.57-72.
- [5] 村上弘: フィルタ対角化法について, **日本応用数学会 2014 年度年会予稿集** (2014 年 8 月), pp.329-330.
- [6] 村上弘: レゾルベントの多項式をフィルタとして用いる対角化法について, **情報処理学会研究報告**, Vol.2014-HPC-146, No.13 (2014 年 9 月), pp.1-4.
- [7] 村上弘: 実対称定値一般固有値問題に対するレゾルベントの多項式によるフィルタの構成法の検討, **情報処理学会研究報告**, Vol.2014-HPC-147, No.2 (2014 年 12 月), pp.1-10.
- [8] 村上弘: 実数シフトのレゾルベントを組み合わせたフィルタによる実対称定値一般固有値問題の下端付近の固有値を持つ固有対の解法, **HPCS2015 シンポジウム論文集**, Vol.2015 (2015 年 5 月), pp.38-51.