

楕円フィルタによる実対称定値一般固有値問題の フィルタ対角化法の実験

村上 弘^{†1}

実対称定値一般固有値問題に対して、フィルタ対角化法では指定区間に固有値が属する固有対だけを選択的に求めることができる。指定区間の近傍に固有値が属する固有ベクトルだけを選択的に透過させる線形作用素をフィルタとして用いる。十分多くの線形独立なベクトルの組を入力としてフィルタに適用し、その出力に特異値分解を行い、特異値が閾値よりも小さな特異ベクトルを除外する。残った特異ベクトルの組の張る部分空間は、指定区間の近傍に固有値が属する固有ベクトルの組で張られた不変部分空間を近似する。そこで、残った特異ベクトルの組に「部分空間法」を適用すれば、指定区間の近傍に固有値が属する固有対の近似が得られる。これがフィルタ対角化法による近似固有対である。

フィルタに（一般化）レゾルベントと恒等作用素の線形結合を用いると、固有ベクトルに対するフィルタの伝達率は固有値の有理関数（各レゾルベントのシフト量と結合係数と恒等作用素の係数をパラメタとして持つ）で表せる。そこで、伝達関数のパラメタをうまく調整して、指定区間の特性関数をよく近似する有理関数にする。区間の特性関数を有理関数で近似する方法としては、アナログ電気回路の周波数フィルタの設計に用いられているものと同じ手法が使える。

本論文では、フィルタ対角化法の実験を楕円フィルタを用いて行った結果を示す。楕円フィルタでは高い弁別能力を少ない個数のレゾルベントで達成できる。

Experiments of filter diagonalization method for real symmetric definite generalized eigenproblems by the use of elliptic filters

HIROSHI MURAKAMI^{†1}

For a real symmetric definite generalized eigenproblem, the filter diagonalization method can solve selectively only those eigenpairs whose eigenvalues belong to the specified interval. The filter is a linear operator which has the property to pass only those eigenvectors whose eigenvalues belong to the neighbor of the specified interval. Sufficiently many linearly independent vectors as the inputs are filtered, then to the output vectors the singular value decomposition is applied and those singular vectors are removed whose singular values are smaller

than a threshold. The invariant subspace spanned by all those eigenvectors whose eigenvalues belong to the neighbor of the specified interval is approximated by the subspace spanned by the set of singular vectors remained. Then, the subspace method is applied to this set of singular vectors to obtain the approximations of eigenpairs whose eigenvalues belong to the neighbor of the specified interval. These approximated eigenpairs are the results of the filter diagonalization method.

When the filter is a linear combination of the resolvents and an identity, then the filter's transfer rate of the eigenvector is expressed by a rational function of the eigenvalue (associated with the parameters the shifts and the coefficients for the resolvents and the coefficient for the identity). Therefore, those parameters are tuned so that the transfer function will be a good rational function approximation of the characteristic function of the specified interval. For the rational function approximation of the characteristic function of an interval, the same method which has been used in the designs of frequency filters of the analog electronic circuit is applicable.

In this paper, some results of experiments are shown of the filter diagonalization method with the elliptic filters which can attain the higher distinction ability with the small number of resolvents.

1. フィルタ対角化法の原理

行列の実対称定値一般固有値問題 $Av = \lambda Bv$ の固有対で、固有値が指定区間 $\mathcal{I} = [\alpha, \beta]$ に属しているものを求めるものとする。フィルタ \mathcal{F} は線形作用素で、固有値が \mathcal{I} に属する固有ベクトルは良く通過させるが、固有値が \mathcal{I} から離れている固有ベクトルは強く減衰させる性質（帯域通過特性）を持つようにうまく構成されているものとする。すると、任意の入力ベクトルをフィルタに通すと、その出力のベクトルからは固有値が \mathcal{I} の近傍に属さない固有ベクトルがほとんど含まれなくなり、固有値が \mathcal{I} 近傍に属する固有ベクトルの組で張られた小さな次元の部分空間への近似的な射影を与える。

十分多くの（ランダムな）ベクトルの組を、まず線形独立性を高めるために計量 B で正規直交化して入力ベクトルの組とする。出力ベクトルの組に対して計量 B による特異値分解を行い、相対的に特異値が小さい特異ベクトルは誤差拡大を防ぐために取り除く。除かれずに残った特異ベクトルの組に対して「部分空間法」を適用すると、固有値が \mathcal{I} の近傍に属する固有対の近似対が得られる。

このフィルタ対角化法の手続きにより得られた近似対は、必要ならば Rayleigh 逆反復も

^{†1} 首都大学東京 数理情報科学専攻

Department of Mathematics and Information Sciences, Tokyo Metropolitan University

しくは Ritz 同時逆反復の数回程度の適用で、精度を改良できる。

2. レゾルベントの線形結合によるフィルタ

いま実対称定値の一般固有値問題に対応する(一般化)レゾルベントを $R(\lambda) \equiv (A - \lambda B)^{-1} B$ とする。フィルタは $2n$ 個のレゾルベントの線形結合とするが、後の導出との関係で、他に恒等演算子 I の項も含めて定義しておく： $\mathcal{F} \equiv c_\infty I + \sum_{p=1}^{2n} c_p R(\lambda_p)$ 。調整可能なパラメータは、整数 n 、分点 $\lambda_p, p=1, 2, \dots, 2n$ 、係数 $c_p, p=1, 2, \dots, 2n$ 、および係数 c_∞ である。係数や分点は一般に複素数値である。

固有対 $(\lambda^{(\nu)}, \mathbf{v}^{(\nu)})$ に対して $\mathcal{F} \mathbf{v}^{(\nu)} = \mathbf{v}^{(\nu)} \cdot f(\lambda^{(\nu)})$ 、但し $f(\lambda)$ は有理関数 $f(\lambda) \equiv c_\infty + \sum_{p=1}^{2n} c_p / (\lambda - \lambda_p)$ で、固有値 λ を持つ固有ベクトルに対するフィルタ \mathcal{F} の伝達率(出力/入力)の比)を与える。固有値に縮重があっても、固有ベクトルの伝達率は実対称定値一般固有値問題の場合には、その固有値だけで決まる。 $f(\lambda)$ をフィルタ \mathcal{F} の「伝達関数」とよぶ。

2.1 パラメータを決める指針

指定された実区間を $\mathcal{I} = [\alpha, \beta]$ とするとき、実数 λ の有理関数である伝達関数 $f(\lambda)$ が、区間 \mathcal{I} の特性関数 $\chi_{\mathcal{I}}(\lambda)$ ($\lambda \in \mathcal{I}$ なら値 1, $\lambda \notin \mathcal{I}$ なら値 0) の良い近似となるように、フィルタのパラメータを決める。

指定区間の特性関数の有理関数近似を利用する手法がアナログ電子回路の周波数フィルタの設計に用いられてきた。その手法を模倣すると、フィルタ対角化法により実対称定値一般固有値問題の指定区間に固有値を持つ固有対を解くためのフィルタの設計ができる。フィルタ設計の数理的基礎の参考として、文献 1) が極めて有用であった。

2.2 フィルタの設計

簡単化のために、指定された区間 $\lambda \in [\alpha, \beta]$ と標準区間 $t \in [-1, 1]$ との間の線形変換 $\lambda = \mathcal{L}(t)$ を用いて、伝達関数 $f(\lambda)$ の引数 λ を正規化座標 t に変更して $g(t) = f(\lambda)$ を定義する。この線形変換の式は $\lambda = \mathcal{L}(t) \equiv (\beta + \alpha)/2 + (\beta - \alpha)/2 \cdot t$ により与えられる。

フィルタ設計に良く用いられる減衰率関数 (attenuation function) $A(t)$ は伝達関数 $g(t)$ の逆数である： $A(t) \equiv 1/g(t)$ 。いま $g(t)$ は有理関数なので $A(t)$ も有理関数である。

いま有理関数 $g(t)$ の複素数範囲での部分分数展開が $g(t) = c_\infty + \sum_{p=1}^{2n} c_p / (t - t_p)$ の形を持つとき、対応するレゾルベントの線形結合のフィルタ作用素は $\mathcal{F} = c_\infty I + \sum_{p=1}^{2n} c_p R(\lambda_p)$ となる。但し $\lambda_p = \mathcal{L}(t_p)$ である。

2.3 フィルタ特性の形状

正規化座標でのフィルタの通過帯域 (passband) を $|t| \leq 1$ 、阻止帯域 (stopband) を $|t| \geq \mu$ とする。但し μ は通過域の中央から阻止帯域の端までの距離と通過帯域の端までの距離との形状比で、1 より大きい値を持つ。通過帯域と阻止帯域の途中の間を遷移帯域 (transition band) という。

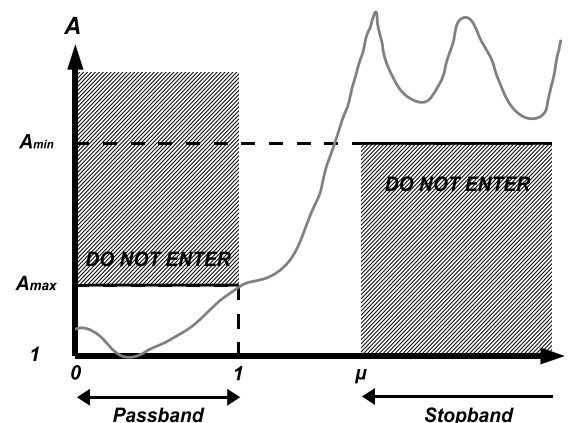


図 1 減衰率関数の形状パラメータ。
Fig. 1 Shape Parameters for Attenuation Function.

さらに減衰率関数 A の値が、通過帯域では A_{\max} 以下、阻止帯域では A_{\min} 以上、となることを要求する。これらのフィルタ特性の形状の概念を、横軸を正規化座標 t 、縦軸を減衰率 A にとって描いたものを図 1 に示す。

議論の簡単化のため、 A は通過帯域に於いて実際に上限値 A_{\max} の値をとる(上限値が tight となっている)と仮定しておく。また、ここで後で用いる値 $L_{\min} \equiv \sqrt{(A_{\min} - 1)/(A_{\max} - 1)}$ を予め定義しておく。

注記：通信工学や電子回路理論では、信号の伝達率や減衰率の大きさを表すのに、しばしば真値の代わりに、常用対数値の 10 倍を [dB] (デシベル) という単位をつけて表記する。例えば 3[dB] は約二倍の減衰率、10[dB] は十倍の減衰率、60[dB] は百万倍の減衰率を表す。

フィルタ特性の形状の選択に関する考察

A_{\max} が大きすぎると通過帯域内の減衰率の均一性が悪くなり、得られる固有対の精度が不揃いになり、極端な場合は必要な固有対が抜け落ちることも起こり得る。小さくし過ぎれば次数 n が増える。そこで通常の使用では A_{\max} を 3[dB] 程度にとることにする。

A_{\min} は十分に大きくないと、不要な固有ベクトルがフィルタで十分に除去されず、また特異値の相対閾値の設定が難しくなる。大きくすれば次数 n が増える。但し、演算の有効精度が D 桁のとき、 $10D$ [dB] を超えた値を指定しても、丸め誤差により減衰効果は実現できない。例えば倍精度演算では 150[dB] 程度とする(あるいはそれ以下 100[dB] などにする)。

$\mu (> 1)$ の値は小さくしないと、遷移帯域が広くなり、フィルタを通過する固有値の範囲が

広がり、フィルタが作り出す部分空間の次元が増え、それに比例して入力ベクトルの個数を増す必要性が生じ、また特異値分布の明瞭な境界が失われる。1.1 以下の小さな値 (1.1 とか 1.01 など) が適当であろう。μ を 1 に近づけると、必要な次数 n が増すが、楕円フィルタではその増加は極めて緩やかである。μ を 1 に近づける場合の利点は、あいまいで不都合な遷移帯域に入る固有値の存在確率が減らせることである。もしも遷移帯域に固有値が存在しなければ、特異値分析は容易で、フィルタ対角化法単独で得られる固有対の精度も十分に高い。フィルタの生成する部分空間の次元数も真に必要な固有ベクトルの個数に近づく。そのため、必要な入力ベクトル、出力ベクトルの個数が減る。但し、必要な固有値の範囲が最初から正確に分かっていることは稀であろう。

2.4 フィルタ設計の手順

フィルタ設計の手順は以下ようになる。

- (1) 三個の形状パラメータ A_{\max} , A_{\min} , μ をフィルタ特性への要求事項として与える。
- (2) フィルタの種別を選ぶ。今回のフィルタの種別は楕円 (elliptic) とする。すると、実軸上で 1 以上の実数値をとる減衰率関数 $A(t)$ の関数形が、パラメータ n, ϵ だけを残して具体的に決まる (楕円フィルタ以外の典型的なフィルタには、バターワース (Butterworth), チェビシェフ (Chebyshev), 逆チェビシェフ (inverse Chebyshev) があるが、楕円フィルタが同じ特性要求を最も小さい次数 n で達成する¹⁸⁾。 n はフィルタを構成するレゾルベントの個数 (の半分) である。)
- (3) フィルタ特性の三個の形状パラメータから、 $|t| \leq 1$ のときには $A \leq A_{\max}$ (この制約条件は tight と仮定した), $|t| \geq \mu$ のときには $A \leq A_{\min}$, を満たすように、 $A(t)$ が持つパラメータである n, ϵ を決める。
- (4) 複素平面内での伝達関数 $g(t) \equiv 1/A$ の部分分数展開の極の位置とその係数を求める。(典型的な 4 種類のフィルタに関しては、伝達関数の極の位置とその係数は解析的表式で計算できる。)
- (5) 得られた部分分数展開の極とその係数に対応して、レゾルベントの線形結合型のフィルタのレゾルベントの分点 (シフト量) と結合係数が決まり、フィルタ対角化用のフィルタ作用素が決定される。

2.5 注記: 恒等演算子の係数について

フィルタ作用素の表式では説明の数学的統一性のために、レゾルベントの他に恒等演算子の項も付け加えたが、その係数 c_∞ は伝達関数 g の無限遠での値 $g(\infty) (\geq 0)$ に等しい。楕円フィルタの場合は n が偶数の場合には零だが、 n が奇数の場合は零にならない。しかし無限遠での伝達関数の値はフィルタ形状の設定から $g(\infty) \leq 1/A_{\min}$ を満たすので、通常的设计の前提ではその大きさは無視できるはずであるので、恒等演算子の項を省いても省かなくても、フィルタの性能は実質的には変わらない。

2.6 フィルタの作用を計算する際の共役対称性の利用

以下で見るように、減衰率関数 $A(t)$ が実軸上で正値形式 (1+実関数の 2 乗の形) の場合には、フィルタ伝達関数 $g(t)$ の極は全て虚数で複素共役な対の組で現われる。また、互いに複素共役である極の持つ係数も互いに複素共役になる。このことから、実数ベクトル \mathbf{v} への (恒等演算子の部分を除いた) フィルタの作用は、 $2n$ 個のレゾルベントのうちで複素共役対の片側ずつを集めた n 個のレゾルベントだけを用いて計算すれば約半分の手間で実現できる。例えばシフト量の虚部が正であるレゾルベントだけを用いる場合には $\sum_{p=1}^{2n} c_p R(\lambda_p) \mathbf{v} = \sum_{\text{Im} \lambda_q > 0} \text{Re} \{ 2c_q R(\lambda_q) \mathbf{v} \}$ となる。このように計算すると途中の演算には複素数が入るが、結果が実数ベクトルとなることが保証される。

3. 楕円フィルタ

楕円フィルタの減衰率関数は t の $2n$ 次の実有理関数で偶関数、自然数 n 以外に ϵ を実数パラメータとして $A(t) \equiv 1 + \epsilon^2 R_n^2(t)$ で表される関数形を持つ¹⁾⁵⁾。 R_n は次数 n の有理関数で、Jacobi の楕円関数 sn を用いた媒介変数 u による表示: $R_n(t) = \text{sn} [K(1/L)(nu + \delta_n), 1/L]$, $t = \text{sn} [K(1/\mu)u, 1/\mu]$ を持つ。ここで $K(k)$ は第一種の完全楕円積分、記号 δ_n は n が奇数のとき 0、偶数のとき $(-1)^{n/2}$ を表す。

この関数は、通過帯域で $A(t)$ の極大値が全て一致する等リプル条件と、阻止帯域で $A(t)$ の極小値が全て一致する等リプル条件を満たす。

3.1 必要最小次数の決定

関数の振舞を考慮すると、フィルタ特性の形状の制約条件から $A_{\max} = 1 + \epsilon^2$, $A_{\min} \leq 1 + \epsilon^2 R_n^2(\mu)$ 。これより $\epsilon^2 = A_{\max} - 1$, $R_n(\mu) \geq L_{\min}$ である。いま実数 $n_{\min} \equiv \{K'(1/L_{\min})/K(1/L_{\min})\} / \{K'(1/\mu)/K(1/\mu)\}$ と置く。但し、 $K(k)$ は第一種の完全楕円積分を表し、 $K'(k) \equiv K(\sqrt{1-k^2})$ 。すると制約条件は、 n の値を n_{\min} 以上の自然数にとれば満たせる。 μ と n の値から L の値は、例えば次式で計算できる: $1/L = \mu^{-n} \prod_{j=1}^{\text{floor}(n/2)} \text{sn}^4 [(2j-1)K(1/\mu)/n, 1/\mu]$ 。

3.2 伝達関数の極とその係数

Jacobi の楕円関数を用いた媒介変数表示により、 $g(t) = 1/A(t)$ の極 t_p とその係数 c_p は解析的式で表わされ、次の手順で計算できる。まず $b \equiv F(\tan^{-1}(1/\epsilon), \sqrt{1-L^{-2}})$ を計算する。ここで $F(\phi, k) \equiv \int_0^\phi (1-k^2 \sin^2 x)^{-1/2} dx$ は第一種の楕円積分を表す。次に $\tau \equiv (b/n)K(1/\mu)/K(1/L)$, $\theta_p \equiv (2p+1 - \text{mod}(n, 2))K(1/\mu)/n$ 。すると極の値は、 $t_p = \text{sn}(\theta_p + \sqrt{-1}\tau, 1/\mu)$, $p=1, 2, \dots, 2n$ 。(注意: この表式では正の虚部を持つ極と対応する添字は $p=1, 2, \dots, n$ にはならない。巡回的にずらせば良い。詳細省略) 極の係数は $c_p = \zeta \sqrt{-1} \cdot \text{cn}(\theta_p + \sqrt{-1}\tau, 1/\mu) \text{dn}(\theta_p + \sqrt{-1}\tau, 1/\mu)$ 。ここで $\zeta = \{-1/(2n)\}K(1/\mu)/K(1/L) \sqrt{\epsilon^2 / \{(1+\epsilon^2)(\epsilon^2+L^{-2})\}}$ 。係数 c_∞ の値は n が奇数のとき 0、

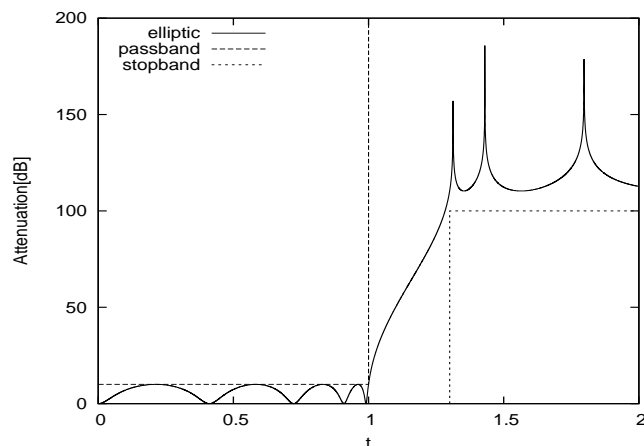


図2 楕円フィルタの減衰率関数 $\mathcal{A}(t)$ のグラフの例。
Fig.2 Sample Graphs of Attenuation Functions of Elliptic Filter.
($\mathcal{A}_{\max}=10[\text{dB}]$, $\mathcal{A}_{\min}=100[\text{dB}]$, $\mu=1.3$, $n=9$.)

偶数のとき $1/(1+\epsilon^2 L^2)$ である。

引数が複素数の楕円関数 sn , cn , dn の値は引数が実数の楕円関数 sn , cn , dn の値を組合せて計算できる。第一種の完全楕円積分 $K(k)$ や第一種の楕円積分 $F(\phi, k)$, Jacobi の楕円関数 $\text{sn}(x, k)$, $\text{cn}(x, k)$, $\text{dn}(x, k)$ の計算には、数学ライブラリ関数を用いることができる²⁾。その際に、関数定義が異なっていたり、 k の代わりに k^2 の値を引数とする実装がされている場合がある。また、ライブラリの実装によっては、引数 k の極端な値である 0, 1 の付近で関数値の精度が十分出せないものなどがあり、注意や検討が必要である。

3.3 減衰率関数のグラフの例

図2は、楕円フィルタの特性の形状パラメータとして $\mathcal{A}_{\max}=10[\text{dB}]$, $\mathcal{A}_{\min}=100[\text{dB}]$, $\mu=1.3$ を指定し、 n を必要最小値 9 に決めて、得られた減衰率関数 \mathcal{A} をプロットしたものである。(グラフの様子を理解し易いように、通過帯域の上限 \mathcal{A}_{\max} の値、 μ の値はいずれも実際の使用で想定する値よりも大きくした。) 減衰率関数は正規化座標 t の偶関数でグラフは左右対称となるから、 $t \geq 0$ の側だけをプロットした。

3.4 楕円フィルタの極の分布の例

楕円フィルタの特性の形状への要求 ($\mathcal{A}_{\max}=3[\text{dB}]$, $\mathcal{A}_{\min}=100[\text{dB}]$, $\mu=1.1$) を満たす最小次数は $n=12$ である。その伝達関数 $g(t)$ の複素平面上での極の分布をグラフにプロットした例を図3に示す。

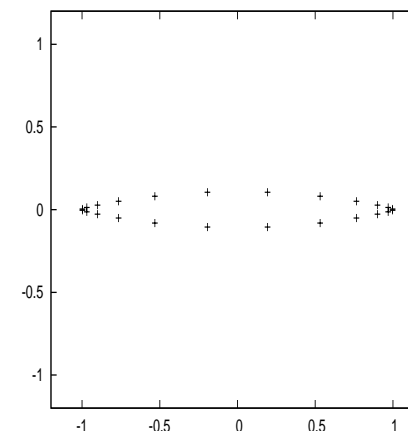


図3 複素平面上の楕円フィルタの極の分布をプロットした例。
Fig.3 Sample Plot of Complex Poles of Elliptic Filter.
($\mathcal{A}_{\max}=3[\text{dB}]$, $\mathcal{A}_{\min}=100[\text{dB}]$, $\mu=1.1$, $n=12$.)

表1 楕円フィルタの形状要求を実現する n の最小値。 $\mathcal{A}_{\max}=3[\text{dB}]$ 。
Table 1 Smallest n for Shape Requirements to the Elliptic Filter.

μ	$\mathcal{A}_{\min}=80[\text{dB}]$	$\mathcal{A}_{\min}=100[\text{dB}]$	$\mathcal{A}_{\min}=150[\text{dB}]$
1.001	20	24	35
1.003	17	21	30
1.005	16	20	28
1.01	15	18	26
1.03	13	15	22
1.05	11	14	20
1.1	10	12	17
1.2	9	10	15
1.3	8	9	13
1.5	7	8	12

3.5 特性形状の要求を実現する最小の次数の例

表1は、楕円フィルタの特性形状に対する三個の要求パラメータのうち \mathcal{A}_{\max} の値はすべて通常使用する $3[\text{dB}]$ とし、各列に \mathcal{A}_{\min} の値をそれぞれ $80[\text{dB}]$, $100[\text{dB}]$, $150[\text{dB}]$ とし、各行に μ の値をそれぞれ 1.001, 1.003, 1.005, 1.01, 1.03, 1.1, 1.2, 1.3, 1.5 とし、要求を実現する次数 n の最小値を表にしたものである。

3.6 楕円フィルタの次数の下限値 n_{\min} の漸近近似式

楕円フィルタの弁別能力を表す形状比 μ が 1 の付近 (あるいは $\ln(\mu)$ が 0 付近) での, 特性形状の要求を満たすための自然数 n の下限値を与える実数 n_{\min} の漸近近似を以下に示す. $K(k)$ の定義は $\int_0^{\pi/2} \frac{d\theta}{\sqrt{1-k^2 \sin^2 \theta}}$. いま $k \ll 1$ であるとき, 楕円積分に関する数学公式から, $K(k) \approx \pi/2 + O(k^2)$, $q(k) \equiv \exp(-\pi K'(k)/K(k)) = k^2/16 + O(k^4)$. また $k \rightarrow 1$ の時に, $\lim_{k \rightarrow 1} \{K(k) - (1/2) \ln(16/(1-k^2))\} = 0$ が成立する. 結局, $\mu \rightarrow 1$ に対する実数 n_{\min} の漸近近似式として次のものを得た: $n_{\min} = \{K'(1/L_{\min})/K(1/L_{\min})\} / \{K'(1/\mu)/K(1/\mu)\} \approx (2/\pi^2) \ln(4L_{\min}) \ln(8/\ln \mu)$. 数値的に試してみると, この近似式は正確な値に極めて近い. 但しこれは μ が 1 に近づくときの漸近近似式で正確な値を表すものではない.

3.7 楕円フィルタの極と係数の数表の例

楕円フィルタの三個の形状要求 A_{\max} , A_{\min} , μ の値を与えると, 必要な次数の最小値を出力し, それ以上の値を n として与えると伝達関数の極と係数, 無限遠での値を出力する計算プログラムを作成した.

例 1

楕円フィルタで, 形状条件が $A_{\max}=3$ [dB], $A_{\min}=150$ [dB], $\mu=1.1$ で, 次数が $n=17$ の場合の, 正規化座標 t での極とその係数の表を表 2 に示す. 上段が極の値で下段がその係数値である (この表は倍精度計算用である. 数値の Fortran 表記 1.234D-05 は倍精度の定数 1.234×10^{-5} を意味する.) 複素共役対称性から, 表には虚部が正である極とその係数だけを載せている. また極の実部の符号を変えると極係数の実部の符号も変わることを用いればさらに表のサイズを約半分に圧縮することが可能である. n が奇数なので係数 $c_{\infty} = g(\infty)$ は零である.

例 2

楕円フィルタで, 形状条件が $A_{\max}=3$ [dB], $A_{\min}=100$ [dB], $\mu=1.1$ で, 次数が $n=12$ の場合の, 正規化座標 t での極とその係数の表を表 3 に示す. 上段が極の値で下段がその係数値である (この表も倍精度計算用である.) 表には虚部が正である極とその係数だけを載せている. この表もサイズを半分に圧縮することが可能である. n が偶数なので係数 $c_{\infty} = g(\infty)$ は零ではないが, $1/A_{\min}$ 以下で小さいので省略した. このフィルタの阻止帯域での減衰率は 10 桁なので, この表の数値の有効数字 16 桁は実際は過剰だがそのまま載せている.

4. 特異値の切断に用いる相対閾値の考察

現在用いている対称定値一般固有値問題用のフィルタ対角化法では, 十分多くのランダムなベクトルの組を計量 B で正規直交化して, フィルタへの入力とする. 出力として得られ

表 2 楕円フィルタ: 正規化座標での極とその係数
 $A_{\max}=3$ [dB], $A_{\min}=150$ [dB], $\mu=1.1$, $n=17$.

	実部	虚部
1	-0.9988990674636575D+00 0.4208879041201686D-03	0.1832241780825514D-02 -0.7299481769619925D-03
2	-0.9854347625608995D+00 0.5186819992362039D-03	0.5927158108762483D-02 -0.2361947294880509D-02
3	-0.9554033152120891D+00 0.7216244236726592D-03	0.1135056762543125D-01 -0.4525751747884339D-02
4	-0.9023908876059291D+00 0.1030426201848702D-02	0.1905329428422484D-01 -0.7604384719530673D-02
5	-0.8167125542626837D+00 0.1404195176021395D-02	0.2984298449862535D-01 -0.1192790439816883D-01
6	-0.6868077838555658D+00 0.1705866291400565D-02	0.4377872029563018D-01 -0.1753132683591053D-01
7	-0.5036279126962596D+00 0.1679857612187386D-02	0.5915500708890110D-01 -0.2373919487772945D-01
8	-0.2684756930096834D+00 0.1085531284204580D-02	0.7188213075555887D-01 -0.2889753026849765D-01
9	0.0000000000000000D+00 0.0000000000000000D+00	0.7692130784434680D-01 -0.3094493130792976D-01
10	0.2684756930096810D+00 -0.1085531284204572D-02	0.7188213075555897D-01 -0.2889753026849769D-01
11	0.5036279126962588D+00 -0.1679857612187384D-02	0.5915500708890115D-01 -0.2373919487772946D-01
12	0.6868077838555658D+00 -0.1705866291400566D-02	0.4377872029563019D-01 -0.1753132683591053D-01
13	0.8167125542626832D+00 -0.1404195176021397D-02	0.2984298449862541D-01 -0.1192790439816885D-01
14	0.9023908876059290D+00 -0.1030426201848704D-02	0.1905329428422488D-01 -0.7604384719530687D-02
15	0.9554033152120891D+00 -0.7216244236726585D-03	0.1135056762543125D-01 -0.4525751747884337D-02
16	0.9854347625608992D+00 -0.5186819992362051D-03	0.5927158108762525D-02 -0.2361947294880525D-02
17	0.9988990674636575D+00 -0.4208879041201745D-03	0.1832241780825549D-02 -0.7299481769620064D-03

表 3 楕円フィルタ：正規化座標での極とその係数
 $A_{\max}=3[\text{dB}]$, $A_{\min}=100[\text{dB}]$, $\mu=1.1$, $n=12$.

	実部	虚部
1	-0.9978032747225308D+00 0.8664302031627370D-03	0.3701716028012426D-02 -0.1465613329228266D-02
2	-0.9697327336698406D+00 0.1266917017646285D-02	0.1285709956095331D-01 -0.5096119758585352D-02
3	-0.9008627058055854D+00 0.2095891409389544D-02	0.2747634325702563D-01 -0.1091872191677904D-01
4	-0.7653097823348445D+00 0.3144348312333809D-02	0.5078939655526350D-01 -0.2027351913073286D-01
5	-0.5316970895763372D+00 0.3449070765877447D-02	0.8117259779721395D-01 -0.3259458505193008D-01
6	-0.1933304451465211D+00 0.1624414004965714D-02	0.1054819597085268D+00 -0.4255796367297806D-01
7	0.1933304451465205D+00 -0.1624414004965710D-02	0.1054819597085268D+00 -0.4255796367297807D-01
8	0.5316970895763370D+00 -0.3449070765877447D-02	0.8117259779721399D-01 -0.3259458505193009D-01
9	0.7653097823348451D+00 -0.3144348312333808D-02	0.5078939655526342D-01 -0.2027351913073283D-01
10	0.9008627058055850D+00 -0.2095891409389546D-02	0.2747634325702568D-01 -0.1091872191677905D-01
11	0.9697327336698404D+00 -0.1266917017646290D-02	0.1285709956095339D-01 -0.5096119758585382D-02
12	0.9978032747225308D+00 -0.8664302031627269D-03	0.3701716028012395D-02 -0.1465613329228254D-02

たベクトルの組に対して、計量 B による特異値分解を行い、丸め誤差の拡大を抑えるために、相対的に非常に小さい特異値を持つ特異ベクトルを棄却する処理「切断」を加える

切断後に残った特異ベクトルが含む丸め誤差の拡大率は、相対閾値を ϵ_{SVD} に設定した場合、 $\epsilon_{\text{SVD}}^{-1}$ 程度になる。計量 B で正規直交な乱数ベクトルを入力とした場合には、入力ベクトルの組に含まれる各固有ベクトルの含有量は統計的な確率分布を伴う変数で、フィルタの出力ベクトルの組に含まれる固有ベクトルの含有量も確率分布を持った量になる。

不要な固有ベクトルの除去にフィルタを用いる趣旨からは、特異値の相対比較による切断に用いる小さい値を持つ閾値を $r \equiv A_{\max}/A_{\min}$ より小さい値に設定しても意味がない。閾値として適切な値は例えば $r^{2/3}$ と $r^{1/3}$ の間の辺りであろう。

丸め誤差を伴う数値計算の場合には、 A_{\min} として実現可能な値は計算精度から制限を受ける。計算に用いる浮動小数点数のマシンイプシロンを ϵ_M とするとき、実際の計算でフィルタを適用した場合の減衰率が ϵ_M^{-1} より大きいことは期待できない。このことから切断の閾値として適切な値は計算精度にも関係することがわかる。

フィルタ対角化の算法にとって不都合な状況は、乱数で生成された入力ベクトルの組に求めたい固有ベクトルがごく僅かしか含まれない場合である。乱数ベクトルの個数を増せば、そのような状況が生じる確率は小さくなる。このような統計的分布を持つデータに対する算法の振舞いの理論的な考察が必要であろう。

5. 楕円フィルタを用いたフィルタ対角化法の実験例

楕円フィルタを用いたフィルタ対角化法の実験を数例示す。一般固有値方程式の係数の実対称行列 A , B を幅の狭い帯行列にとり、 N 次行列 A , B の半帯幅を h (行列の (i, j) 要素が $|i - j| \leq h$ 以外はすべて零) として、行列 A の帯内の非零要素を $a_{i,j} \equiv \max(i, j) - 1$ 、行列 B の帯内の非零要素を $b_{i,j} \equiv 1/(i + j - 1) + \delta_{i,j}$ にとった ($\delta_{i,j}$ は Kronecker 記号)。

レゾルベント $R(\lambda_p) = (A - \lambda_p B)^{-1} B$ をベクトルへ作用させる計算は、ベクトルに B を乗じたものを右辺とし、帯行列 $(A - \lambda_p B)$ を係数とする線形方程式を、片側枢軸選択つきの複素帯行列の LU -分解を利用して解いた。

使用した例題の固有値には縮重や極端な近接がなかったため、フィルタ対角化法により得られたすでに相当良質の近似固有対の改良には、普通の Rayleigh 商付き逆反復法を用いた。

計算で得られた近似固有対 (λ, \mathbf{v}) の精度の見積りには、 \mathbf{v} がすでに B -正規化されているとき、固有値の誤差の限界すなわち区間 $[\lambda - \Delta, \lambda + \Delta]$ が真の固有値を含むような距離 Δ の値が残差ベクトル $\mathbf{r} \equiv (A - \lambda B)\mathbf{v}$ のノルムにより $\Delta \equiv \sqrt{\mathbf{r}^T B^{-1} \mathbf{r}}$ として与えられることを利用した (これは実対称な標準固有値問題の Wilkinson 限界として知られている残差の 2-ノルム $\|\mathbf{r}\|_2$ を、実対称定値一般固有値問題へ簡単に拡張したものである)。そのほか逆反復法による改良による変動の大きさが参考となる。

表 4 実験に用いた計算機システム .

Table 4 Specifications of the Computer System for Experiments.

CPU	Intel Core i7 920 (2.66GHz, 8MB L3) (コアを 1 個のみ使用) .
メモリ	DDR3-1333 PC3-10600 2GB × 6=12GB (triple channel) .
コンパイラ	Intel Fortran v11.1 for intel64 (オプション-fast) .
浮動小数点数	IEEE 64-bit 倍精度 .
OS	Fedora10 for intel64 .

数値実験を行った計算機システムの仕様を表 4 に掲げる . この CPU は 4 個の計算コアを持つが , 今回の実験では並列化をせずにコアを 1 個だけ用いて計算した .

5.1 実験の各グラフの説明

各近似固有対について , 固有値を横軸 , 残差のノルムの対数を縦軸にとり , 以下の各段階ごとに , グラフ中に折れ線でプロットした . ITER0 はフィルタ対角化法の計算結果であり , ITER1 はフィルタ対角化法の後に逆反復を 1 回適用して近似解を改良した結果 , ITER2 はフィルタ対角化法の後に逆反復を 2 回適用して近似解を改良した結果である .

グラフは , 閾値で切断されずに残った特異ベクトルに対する Rayleigh-Ritz 法で得た近似対全部のものと , 固有値が区間 $I = [\alpha, \beta]$ 内にある近似対だけのものとを別にしてプロットしている . 経過時間は「フィルタ対角化法」と「逆反復 2 回」に分けて示した .

以下の三個の例題 1 から 3 は , いずれも行列の次数 $N=10^6$, 半帯幅 $h=10$, 区間 $I=[-10, 10]$ で , 固有対の個数は $r=52$ 個の問題である . またランダムに与える (B -正規直交の) 初期ベクトルの個数は 100 とした . フィルタ対角化法は Rayleigh-Ritz 法を用いるため , 初期ベクトルの個数は真の固有対の個数以上であることが必要である (初期ベクトルの個数が不足すると特異値分解での階数低下が十分に起きない) . 計算は全て IEEE 64bit 倍精度演算で行い , 特異値を切断する相対閾値として 10^{-7} を用いた .

例題 1

楕円フィルタの形状の要求 $A_{\max}=3$ [dB] , $A_{\min}=150$ [dB] , $\mu=1.1$ に対して , 次数 n とし形状の要求を満たす最小の値 17 を用いた .

フィルタの出力を計量 B で特異値分解して得られた特異値の分布のグラフを図 4 に示す . 中間の値を持つ特異値が 4 個あるのが分かる .

特異値を相対閾値で切断して得られた特異ベクトルを基底とする部分空間の階数は 54 となった . この 54 次元の部分空間に Rayleigh-Ritz 法を適用して得られた近似固有対のうち , 固有値が I 内にある近似固有対は 52 個となった .

図 5 は部分空間法で得られた近似固有対の全てについて , 図 6 は固有値が区間 I 内にあ

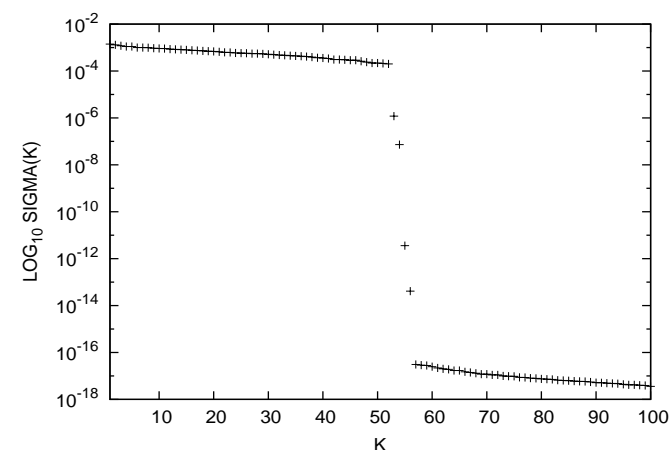


図 4 例題 1 : 特異値の分布 .

Fig.4 Example1: Distribution of Singular Values.

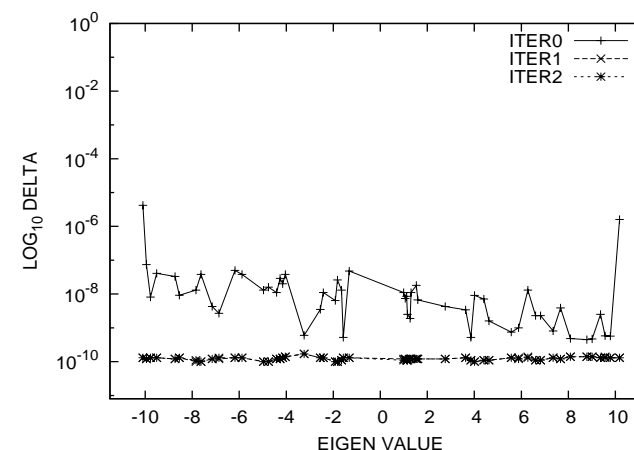


図 5 例題 1 : 近似固有対の品質 (部分空間法による全対) .

Fig.5 Example 1: Qualities of Approximated Eigenpairs (all pairs by subspace method).

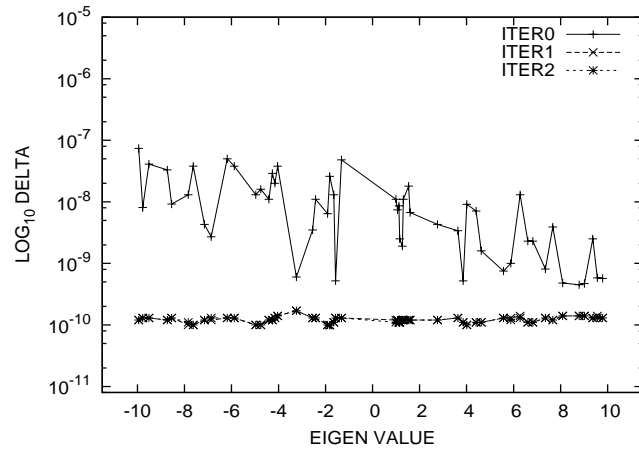


図 6 例題 1 : 近似固有対の品質 (固有値が区間内のもの) .

Fig. 6 Example 1: Qualities of Approximated Eigenpairs (whose eigenvalues are in the interval).

る近似固有対だけについて、それぞれ固有値を横軸に、 Δ の値を対数目盛で縦軸にとり、上側からそれぞれフィルタ対角化法 (ITER0)、逆反復を 1 回、2 回適用 (ITER1, ITER2) した結果をプロットした折線グラフである。

逆反復は 1 回で十分に収束したので、グラフ ITER1 と ITER2 はほとんど重なっている。固有値が区間からはみ出している固有対については、ITER0 では Δ の値が 10^{-5} 程度のものであるが、ITER1 では修正されて 10^{-10} 程度となっている。固有値が区間内にある対については、 Δ の値は ITER0 では 10^{-7} 以下、ITER1 では 10^{-10} 程度となっていて、区間が $[-10, 10]$ であることを考慮すると、フィルタ対角化法で得られた固有値の相対精度は 8 桁程度、逆反復を加えて得られた固有値の相対精度は 11 桁程度であることが分かる。

「フィルタ対角化法」と「逆反復 2 回」の計算の経過時間は、それぞれ 456 秒と 85 秒であった。

例題 2

楕円フィルタの形状を $A_{\max}=3$ [dB], $A_{\min}=150$ [dB], は例題 1 と同じだが、 μ の値を 1.01 と小さく設定し、形状を満たす最小の次数 26 を n に用いた。

フィルタの出力を計量 B で特異値分解して得られた特異値のグラフを図 7 に示す。特異値の分布は極めて良く分離しており、中間の値を持つ特異値が無いことが分かる。これは μ を小さな値 1.01 にとったので僅かに区間から外れた固有値を持つ固有ベクトルも強く減衰を受けたためである。

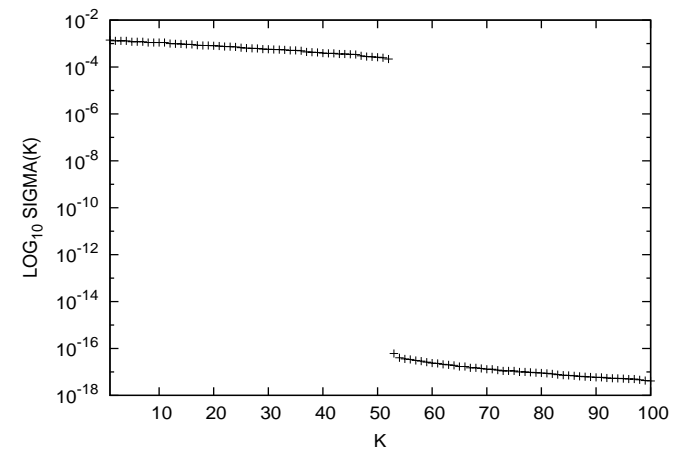


図 7 例題 2 : 特異値の分布 .

Fig. 7 Example 2: Distribution of Singular Values.

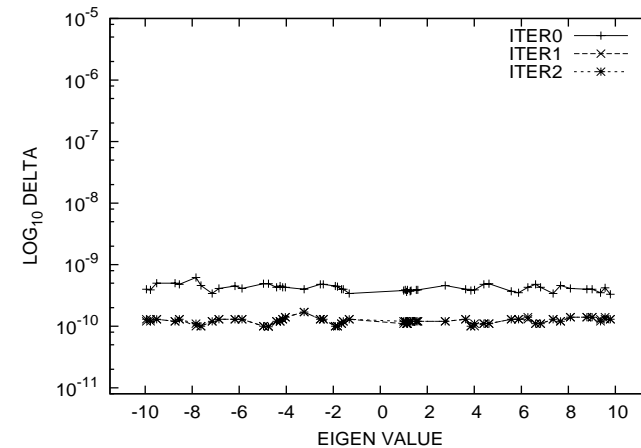


図 8 例題 2 : 近似固有対の品質 (固有値が区間内のもの) .

Fig. 8 Example 2: Qualities of Approximated Eigenpairs (whose eigenvalues are in the interval).

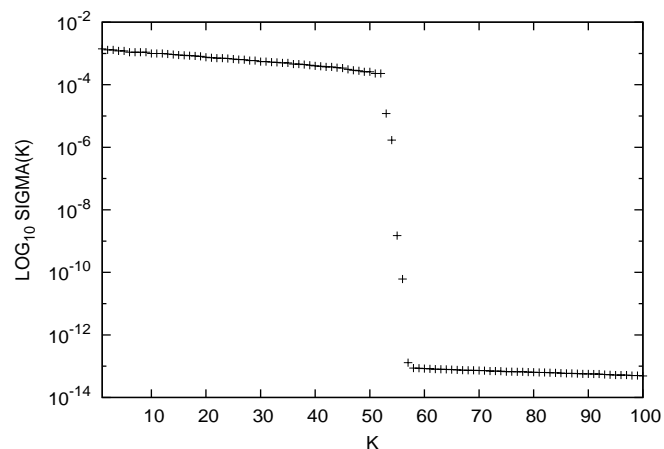


図 9 例題 3 : 特異値の分布 .

Fig.9 Example 3: Distribution of Singular Values.

特異値を相対閾値で切断して得られた部分空間の階数はちょうど 52 となった . 得られた部分空間に Rayleigh-Ritz 法を適用して得られた近似固有対のうち, 固有値が \mathcal{I} 内にあるものの個数も 52 に一致した .

部分空間法で求めた近似固有対は固有値が全て区間 \mathcal{I} にあったので, それら全てについて近似固有値を横軸に, 近似固有対に対応する Δ の値を対数目盛で縦軸にとり, 上側からそれぞれフィルタ対角化法 (ITER0), 逆反復適用 1 回 (ITER1), 逆反復適用 2 回 (ITER2) の結果を折線グラフでプロットしたグラフを図 8 に示す .

逆反復は 1 回で十分に収束したため, グラフ中の ITER1 と ITER2 の折線はほとんど重なっている . 得られた近似固有対は, ITER0 では Δ の値は 10^{-9} 以下, ITER1 では Δ の値は 10^{-10} 程度となっている . フィルタ対角化法の近似対が既に極めて高い精度も持っている . 区間が $[-10, 10]$ であることを考慮すると, フィルタ対角化法で得られた近似固有値の相対精度は 10 桁程度で, 逆反復を加えて得られた固有値の相対精度は 11 桁程度あることがグラフから分かる .

「フィルタ対角化法」と「逆反復 2 回」の計算の経過時間は, それぞれ 644 秒と 82 秒であった . 例題 1 に比べてフィルタ対角化法の時間が増えた主な理由は, 次数 n の値が例題 1 の場合の 17 から 26 に増えたためである .

例題 3

楕円フィルタの形状を $A_{\max}=3[\text{dB}]$, $A_{\min}=100[\text{dB}]$, $\mu=1.1$ と, 例題 1 に比べて減衰

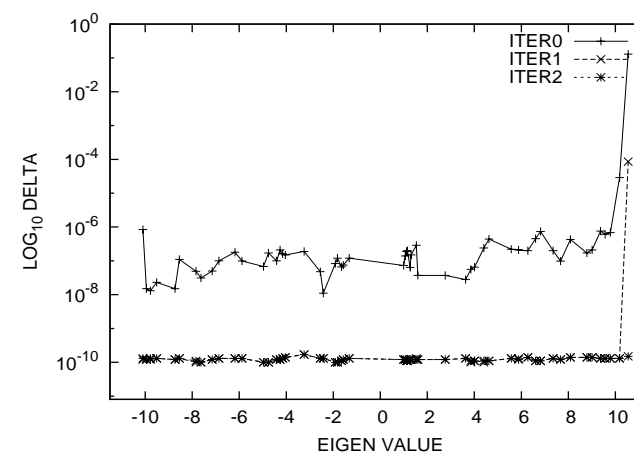


図 10 例題 3 : 近似固有対の品質 (部分空間法による全対) .

Fig.10 Example 3: Qualities of Approximated Eigenpairs (all pairs by subspace method).

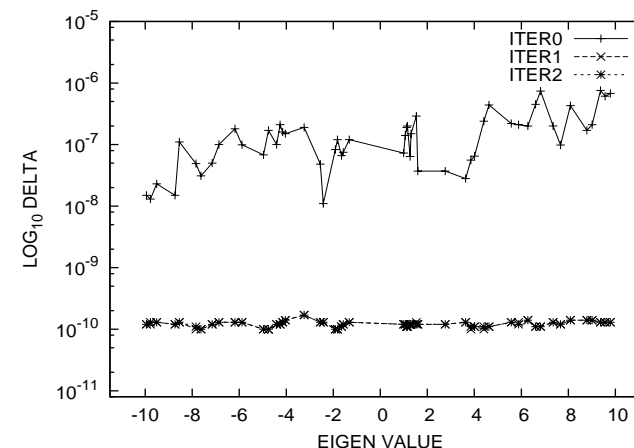


図 11 例題 3 : 近似固有対の品質 (固有値が区間内のもの) .

Fig.11 Example 3: Qualities of Approximated Eigenpairs (whose eigenvalues are in the interval).

量の下限 A_{\min} を小さくとり、形状を満たす最小の次数 12 を n とした。

フィルタの出力を計量 B で特異値分解して得られた特異値のグラフを図 9 に示す。中間の値を持つ特異値が 4 個あることが分かる。

特異値を相対閾値で切断して得られた部分空間の階数は 55 となった。Rayleigh-Ritz 法をフィルタ後の階数 55 の部分空間に適用し、図 10 は得られた近似固有対の全てについて、図 11 は得られた近似固有対のうちで固有値が区間 I 内のものだけについて、それぞれ近似固有値を横軸に、固有対の Δ の値を対数目盛で縦軸にとり、上側からそれぞれフィルタ対角化法 (ITER0)、逆反復をそれぞれ 1 回、2 回 (ITER1, ITER2) 適用した結果をプロットしたグラフをそれぞれ示す。

固有値が区間中の固有対に対しては、逆反復は 1 回で十分に収束したので、グラフ ITER1 と ITER2 は固有値が区間内の部分についてはほとんど重なっている。区間からはみ出した固有値を持つ固有対については ITER0 で Δ の値が 10^{-1} 程度のものであり、ITER1 でも Δ の値は 10^{-4} 程度と大きい誤差を持つことがわかる。区間内の近似固有対については、ITER0 では Δ の値は 10^{-6} 程度、ITER1 では Δ の値は 10^{-10} 程度となっていて、区間が $[-10, 10]$ であることを考慮すると、フィルタ対角化法で得られた固有値の相対精度は 7 桁程度、逆反復を加えて得られた固有値の相対精度は 11 桁程度であることが分かる。

「フィルタ対角化法」と「逆反復 2 回」の計算の経過時間は、それぞれ 352 秒と 87 秒であった。例題 1 や例題 2 に比べてフィルタ対角化法の経過時間が短い主な理由は、次数が $n = 12$ で、例題 1 の $n = 17$ や例題 2 の $n = 26$ に比べて小さいためである。

6. まとめ

フィルタ対角化法では、使用するフィルタの特性形状への要求をまず適切に決める。フィルタの形状要求を満たすために必要な n の値 (フィルタ作用に必要なレゾルベントの個数) は、楕円フィルタを用いる場合が最も小さくなり、フィルタ対角化法の全体の演算量や並列処理を行う場合の記憶量が小さくできるので有利である。

フィルタの種類 (パワーース, チェビシェフ, 逆チェビシェフ, 楕円) と、形状要求の三個のパラメタを与えると、要求の実現に必要な次数の最小値を求め、さらにその最小値以上の値を次数 n として与えると、正規化座標に於けるフィルタの伝達関数の極の位置と極の係数を計算するプログラムを作成した。任意の指定区間に対し、フィルタ作用素を構成するレゾルベントのシフト量と係数は、伝達関数の極とその係数の値から容易に導びける。

今回、従来のフィルタ対角化法のプログラムに、楕円フィルタの特性形状を指定してそれを満たすフィルタ作用素の構成パラメタを生成する機能を組み込み、数値実験を行った。実験の例は、フィルタを特性の形状で指定することが適切であることを示している。

参考文献

- 1) Daniels, R.W.: *Approximation Methods for Electronic Filter Design*, McGraw-Hill, 1974.
- 2) Thompson, W.J.: *Atlas for Computing Mathematical Functions*, Chap.17: Elliptic Integrals and Elliptic functions, Wiley-Interscience, 1997.
- 3) Chen, R. and Gui, H.: A general and efficient filter-diagonalization method without time propagation, *J.Chem.Phys.*, Vol.105, pp.1311-1317(1996).
- 4) Mandelshtam, V.A. and Taylor, H.S.: A low-storage filter diagonalization method for quantum eigenenergy calculation or for spectral analysis of time signals, *J.Chem.Phys.*, Vol.106, pp.5085-5090(1996).
- 5) Lutovac, M.D., Tošić, D.Y. and Evans, B.L.: *Filter Design for Signal Processing*, Prentice Hall, 2001.
- 6) Toledo, S. and Rabani, E.: Very Large Electronic Structure Calculations Using an Out-of-Core Filter-Diagonalization Method, *J. Comput. Phys.*, Vol.180, No.1, pp.256-269(2002).
- 7) Sakurai, T. and Sugiura, H.: A projection method for generalized eigenvalue problems using numerical integration, *J.Comp.Appl.Math.*, Vol.159, pp.119-128 (2003).
- 8) Zhou, Y., Saad, Y., Tiago, M.L. and Chelikowsky, J.R.: Self-Consistent-Field Calculations using Chebyshev Filtered Subspace Iteration, *J.Comput.Phys.* Vol.219, No.1, pp.172-184(2006).
- 9) 村上 弘: 行列の対称定値一般固有値問題の固有値フィルタと部分空間法による解法, HPCS2007 シンポジウム論文集 IPSJ Symposium Series, Vol.2007, No.1, p.61 (2007).
- 10) Murakami, H.: The Filter Diagonalization Method by the Shifted Inverses, ICCM2007, Conference Abstracts, p.126 (2007). Proceeding paper in CD-ROM (28pages, file name p126_G7-8_proc.pdf).
- 11) 村上 弘: 帯対称定値一般固有値問題のフィルタ対角化法の実験, 情報処理学会研究報告 2007-HPC-110(6), pp.31-36 (2007).
- 12) 村上 弘: レゾルベントの線形結合によるフィルタ対角化法, 情報処理学会論文誌コンピュータシステム, Vol.49, No.SIG2(ACS21), pp.66-87(2008).
- 13) 村上 弘: 非対称な固有値問題へのフィルタ対角化法, 情報処理学会研究報告 2008-HPC-115(1), pp.1-6(2008).
- 14) Murakami, H.: Application of Filter Diagonalization Method to Numerical Solution of Algebraic Equations, Proceedings of SNC2009, pp.95-104 (Aug, 2009).
- 15) Polizzi, E.: Density-matrix-based algorithm for solving eigenvalue problems, *Phys.Rev.B*, Vol.79, No.11, p.115112, 6pages (2009).
- 16) Ikegami, T., Tadano, H., Umeda, H. and Sakurai, T.: Hierarchical parallel algorithm to solve large generalized eigenproblems, HPCS2010 論文集, pp.107-114 (2010).
- 17) Murakami, H.: Filter Diagonalization Method by Resolvents for Symmetric Eigenproblems, HPCS2010 論文集, p.54 (2010).
- 18) 村上 弘: フィルタ対角化法の帯域通過フィルタの最適化, 情報処理学会研究報告, Vol. 2008-HPC-124, No.3 (全 8 頁), (2010) .