

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

村 上 弘^{†1}

実対称定値一般固有値問題に対して、固有値が指定された区間に属する必要な固有対だけを求める解法としてフィルタ対角法がある。フィルタは指定区間の近傍に属する固有値を持つ固有ベクトルだけを選択的に通過させる性質を持つ線形作用素である。フィルタに十分に多くの線形独立なベクトルの組を入力し、その出力を特異値分解し、ある小さい閾値以下の特異値を持つ特異ベクトルを棄却する。棄却されずに残った特異ベクトルの組が張る部分空間は、指定区間の近傍に固有値を持つ固有ベクトル全体の張る不変部分空間の近似になる。棄却さ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、指定区間の近傍に固有値を持つすべての近似固有対がいっせいに得られる。フィルタ対角化法で得られたこれらの近似固有対は、Rayleigh 商反復法や Ritz 同時反復法を用いて精度を改良することもできる。フィルタ作用素をレゾルベントの線形結合で構成すると、任意の固有ベクトルに対するフィルタ作用素の伝達率はその固有値の有理関数となる。この有理関数が指定された区間の特性関数（区間で 1、区間で 0）をよく近似するように、レゾルベントのシフト量と結合係数を決める。区間の特性関数の有理関数による近似法には、アナログ電気回路の周波数フィルタの設計で使われてきたものと同じ数学的手法が使える。本論文では、アナログフィルタの古典的理論に沿いながら、レゾルベントの線形結合によるフィルタ作用素の設計法をフィルタ対角化法にただちに適用できる形で示す。

Filter Designs for the Symmetric Eigenproblems to Solve Eigenpairs Whose Eigenvalues are in the Specified Interval

HIROSHI MURAKAMI^{†1}

For a real symmetric definite generalized eigenproblem, the filter diagonalization method can solve only those required eigenpairs whose eigenvalues are in the specified interval. The filter is a linear operator which passes only those eigenvectors whose eigenvalues are in the neighbor of the specified interval. Sufficiently many linearly independent vectors are filtered as the inputs. Their

outputs are singular value decomposed, and those singular vectors are rejected whose singular values are below a small threshold. The subspace spanned by the set of singular vectors which are not rejected is an approximation of the invariant subspace spanned by all those eigenvectors whose eigenvalues are in the neighbor of the specified interval. The subspace method is applied to this set of singular vectors to obtain all those approximated eigenpairs simultaneously whose eigenvalues are in the neighbor of the specified interval. These approximated eigenpairs by the filter diagonalization method may be refined by such methods as the Rayleigh quotient iterations or the Ritz simultaneous iterations. When the filter operator is constructed as a linear combination of resolvents, the transfer rate for any eigenvector is a rational function of its eigenvalue. The shift parameters and the coefficients of the resolvents are so determined to make this rational function a good approximation of the characteristic function of the specified interval (unity inside the interval and zero outside). To approximate the characteristic function of an interval by a rational function, the same mathematical approach can be used which has been used in the design of frequency filters for analog electronic circuits. In this paper, following the classical theory of the analog filters, the design method of the filter operator as a linear combination of resolvents is shown in a manner directly applicable to the filter diagonalization method.

1. はじめに

行列の固有値問題に対して、固有ベクトルの固有値に基づいた一種のフィルタの機能を持つ作用素を用いることで、指定する区間に固有値が存在する比較的少数個の固有ベクトルの近似だけを選択的に構成できる「フィルタ対角化法」と呼ばれる手法がある（冪乗法は絶対値が最大の固有値を持つ固有ベクトルを、シフト付き逆反復法は指定したシフトに最も近い固有値を持つ固有ベクトルを、それぞれ優先的に透過させるフィルタを用いた反復による固有値解法と見なせる）。

フィルタ対角化法は近年大規模な固有値問題の比較的少数個の固有対を求める用途に用いられてきている。量子力学系のエネルギー固有状態を求める用途などで用いられているようである。フィルタ対角化法（とそれに関連する）の文献例を掲げておく¹⁾⁻¹⁴⁾。

レゾルベントの線形結合をフィルタ作用素として用いるフィルタ対角化法の特徴は、固有値分布の任意の区間の付近にある固有値を持つ固有対の近似を 1 度にまとめて求められる

^{†1} 首都大学東京数理情報科学専攻

Department of Mathematics and Information Sciences, Tokyo Metropolitan University

2 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

ことである。フィルタ作用素中のそれぞれのレゾルベントの作用の計算は、係数行列のシフトが異なる連立一次方程式を解くことに帰着する。よってフィルタ対角化法による大規模疎行列の固有対の計算は異なるシフトの大規模疎行列の連立一次方程式の解法に帰着し、異なるシフトの個数分の並列性がある。また、フィルタ作用素へ入力するベクトルの個数は、指定区間に固有値が属する固有対の数よりも多くとることが必要で、係数行列の各シフトごとに入力ベクトルの個数分の右辺だけが異なる連立一次方程式を解く必要がある。しかし、幅の狭い帯行列のように連立一次方程式が係数行列の LU 分解で容易に解ける場合には、多数の連立一次方程式を解いても求解のコストは小さい。 LU 分解を用いることができない場合に、Krylov 部分空間法系の反復法を用いて大規模な連立一次方程式を解く場合には、ブロック化算法を用いたり各シフトごとの計算を並行処理すれば、反復の際の大規模係数行列に対する記憶参照が共用できて得になる。元の問題が対称定値一般固有値問題の特別な場合である対称標準固有値問題の場合には、本手法の計算は、係数行列に対して異なる複素対角シフト（単位行列の複素数倍）を加えて得られる複数の連立一次方程式を解くことに帰着される。これらの連立一次方程式群を Krylov 部分空間法で反復的に解く場合、Krylov 部分空間が共有されるので、それを活かして計算を効率的にする方法¹⁵⁾も提案されている。

固有値問題を実対称定値一般の場合に限定したうえで、今回の論文では以前の論文¹²⁾からの進展として、フィルタ対角化法でレゾルベントの線形結合をフィルタ作用素とする場合に、パラメータである各レゾルベントのシフトと結合係数を調整して、指定区間内の固有値を持つ固有ベクトルだけをきわめて選択的に通過させる特性を持つように設計する手法を示す。これはアナログ電気回路の周波数フィルタにおいて伝達関数を有理関数で近似して設計する手法^{16),17)}の類似である。以前の論文中で「円周等分多項式」と呼んでいたフィルタはアナログ回路の分野では伝統的に Butterworth フィルタと呼ばれ、「値ずらし Chebyshev 多項式」と呼んでいたフィルタは Chebyshev フィルタと呼ばれるので、今後はその歴史的慣用を尊重することにする（以前の論文中で「Chebyshev 多項式」と呼んでいた共鳴を持つフィルタは電気回路の帯域フィルタとしては不適当なためであろう、対応する名称はなさそうである）。今回はアナログ回路のフィルタから Inverse Chebyshev および Elliptic と呼ばれる族を新たに導入した。特に Elliptic フィルタは、少ない個数のレゾルベントの線形結合で優れた遮断特性の実現ができる。

2. フィルタ対角化法の原理

行列 A が実対称で行列 B が実対称で定値である一般固有値問題 $Av = \lambda Bv$ の固有対で、

固有値が指定区間 $I = [\alpha, \beta]$ に属しているものを、フィルタ対角化法^{7),12)}を用いて求めるものとする。一般性を失わずに B が正定値の場合に限定できる。フィルタ F は線形作用素で、固有値が I に属する固有ベクトルはよく通過させるが、固有値が I から離れた固有ベクトルは強く減衰させる性質を持つように適切に構成されているとする。すると、任意の入力ベクトルをフィルタに通すとき、出力のベクトルは固有値が I の近傍以外の固有ベクトルの成分をほとんど含まなくなり、 I 近傍の固有値を持つ固有ベクトルの組で張られる次元の小さな線形部分空間へ近似的に射影される。

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 F への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 による特異値分析を適用し、ある小さな閾値よりも相対比で小さい特異値を持つ特異ベクトルを取り除く。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、固有値が区間 I の近傍にある固有対の近似対が得られる。

以上のフィルタ対角化法により得られた近似固有対は必要に応じて、たとえば数回程度の Rayleigh 逆反復もしくは Ritz 同時逆反復の適用で、精度を改良する。

注 1: 縦ベクトルの組に対し、ベクトルを横に並べた行列を W とする。 W の計量 B による特異値分解は $W \Rightarrow U\Sigma Q$ である。 U は特異ベクトルを横に並べた行列で B -正規直交性 $U^T B U = I$ を満たし、 Q は通常の直交行列で $Q^T Q = Q Q^T = I$ を満たし、 Σ は特異値を対角線上に並べた非負対角行列である。

注 2: フィルタ対角化法の計算量の考察や、より具体的な算法の中身（計量 B での正規直交化や計量 B での特異値分解の実現法の例などを含む）については、前回の論文¹²⁾に記述がある。フィルタの種類（Butterworth, Chebyshev, Inverse Chebyshev, Elliptic）が異なっても、レゾルベントの個数と各レゾルベントのシフト量の分点と線形結合の係数が異なるだけである。フィルタは使用する種類と後述のフィルタ特性の形状の仕様を決める。それにより決まる必要最小な値以上にレゾルベントの個数を設定すれば、（固有値の指定区間 $[\alpha, \beta]$ を線形に $[-1, 1]$ へ移した正規化座標での）シフトの座標と線形結合係数が複素数値として決まる。そこで、あらかじめプログラムの中でレゾルベントの個数とともにシフトの座標と線形結合係数の値を数通り表として組み込んでおけば、毎回作り直さずに済む。

3. レゾルベントの線形結合によるフィルタ

いま元の実対称定値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。 n をある自然数として、フィルタは $2n$ 個のレゾルベント $R(\lambda_p)$

3 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

の線形結合とする(後の数学的な導出との関係でレゾルベントのほかに恒等演算子 I の項を加えて定義しておく). $\mathcal{F} \equiv c_\infty I + \sum_{p=1}^{2n} \gamma_p R(\lambda_p)$. フィルタの調整可能なパラメータは, 自然数 n , 分点 $\lambda_p, p=1, 2, \dots, 2n$, 係数 $\gamma_p, p=1, 2, \dots, 2n$, それと係数 c_∞ である. 分点や係数は一般には複素数値をとる.

固有対 $(\lambda^{(\nu)}, \mathbf{v}^{(\nu)})$ に対して $\mathcal{F} \mathbf{v}^{(\nu)} = \mathbf{v}^{(\nu)} \cdot f(\lambda^{(\nu)})$ が成立する. ただし $f(\lambda) \equiv c_\infty + \sum_{p=1}^{2n} \gamma_p / (\lambda - \lambda_p)$ とおいた. つまり \mathcal{F} を任意の固有ベクトルに作用させると λ をその固有値とすれば $f(\lambda)$ 倍になる. 有理関数 $f(\lambda)$ は固有値 λ を持つ固有ベクトルに対する線形フィルタ \mathcal{F} の入力に対する出力の応答の比で, フィルタ作用素の「伝達関数」である.

3.1 パラメータを決める指針

いま, 求めたい固有対の固有値の区間を $\mathcal{I} = [\alpha, \beta]$ とする. フィルタ対角化法についての「理想的」なフィルタは, 伝達関数 $f(\lambda)$ の絶対値が区間 \mathcal{I} の特性関数, つまり区間 \mathcal{I} 内では 1 で区間外では 0 となるものである. そのような理想フィルタからの出力ベクトルに含まれる固有ベクトルは固有値が \mathcal{I} に含まれるものだけになり, 出力ベクトルの組により張られる部分空間は, 固有値が \mathcal{I} に属する固有ベクトルで張られた不変部分空間になる.

しかしそのような理想フィルタの伝達関数は有理関数では実現できない. なぜならばまず区間 \mathcal{I} の外で値が 0 ならば有理関数は恒等的に零となる, また有理関数に不連続性があればその不連続点は極であるはずだが, 極の特異性は有界変動ではないからである. そこで, 「理想的」なフィルタの伝達特性に修正を入れて, それを近似する連続関数を有理関数により構成し, 実際のフィルタの伝達関数に用いる.

実際のフィルタの伝達関数 $f(\lambda)$ の絶対値は, λ が区間 \mathcal{I} 内のときは 1 付近の値, λ が区間 \mathcal{I} から “少し” 離れた外部では非常に小さいほとんど 0 の値をとる連続関数であるとする(中間の領域では両方の間の値をとることにする). このような性質を持つ有理関数を伝達関数として採用し, レゾルベントの線形結合によるフィルタ作用素がその伝達関数を持つようにパラメータを決める.

理想フィルタの伝達関数を有理関数で近似する手法については, その数学的な類似性から, アナログ電気回路の周波数フィルタに対する参考文献 16) の記述がきわめて有用であった. 実対称定値一般固有値問題の固有値はすべて実軸上にあり, レゾルベントの線形結合を用いたフィルタ作用素の固有ベクトルに対する伝達関数は固有値の有理関数である. そのことから, アナログ回路で有理関数を用いて理想的な帯域フィルタの伝達関数の絶対値を近似したのと同じ手法が利用できる.

3.2 フィルタの設計

いま指定された区間 $\lambda \in [\alpha, \beta]$ と標準区間 $t \in [-1, 1]$ の間の線形変換 $\lambda = \mathcal{L}(t)$ を用いて, 伝達関数 $f(\lambda)$ を t の関数に変更して $g(t) = f(\lambda)$ を定義する. 線形変換は $\lambda = \mathcal{L}(t) \equiv (\beta + \alpha)/2 + (\beta - \alpha)/2 \cdot t$ で与えられる.

いま, 有理関数である $g(t)$ の複素数範囲での部分分数展開が次の形を持つとする: $g(t) = c_\infty + \sum_{p=1}^{2n} c_p / (t - t_p)$. そのとき, 対応するレゾルベントの線形結合型のフィルタ作用素の表式は $\mathcal{F} = c_\infty I + \sum_{p=1}^{2n} \gamma_p R(\lambda_p)$ となる. ただし $\lambda_p = \mathcal{L}(t_p)$, $\gamma_p = \mathcal{L}' \cdot c_p$ である. \mathcal{L}' は \mathcal{L} の微分値で定数 $(\beta - \alpha)/2$ である.

フィルタ設計に現れる減衰率関数 (attenuation function) $\mathcal{A}(t)$ は実数 t に対してつねに 1 以上の値をとる関数で, 伝達関数の逆数 $\mathcal{A}(t) \equiv 1/g(t)$ である. いま $g(t)$ が有理関数なので $\mathcal{A}(t)$ も有理関数になる.

3.3 フィルタ特性の形状

正規化された変数 t により, フィルタの通過帯域 (passband) は $|t| \leq 1$, 阻止帯域 (stopband) は比 $\mu (> 1)$ を用いて $|t| \geq \mu$ と書ける. 中間の領域 $1 < |t| < \mu$ は遷移領域 (transition region) と呼ばれる. そこで減衰率関数 $\mathcal{A}(t)$ の値が, 通過帯域では \mathcal{A}_{\max} 以下, 阻止帯域では \mathcal{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に対する概念図を, 横軸を正規化座標 t , 縦軸を減衰率 \mathcal{A} にとって描いたものを図 1 に示す.

議論の簡化のため, $\mathcal{A}(t)$ は通過帯域において実際には上限値 \mathcal{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, 後で用いる値 $L_{\min} \equiv \sqrt{(\mathcal{A}_{\min} - 1)/(\mathcal{A}_{\max} - 1)}$ を先にここで定義しておく.

注記: 電気回路の理論^{16), 17)} では, 信号の伝達率や減衰率の大きさを, 真値の代わりに常用対数値の 10 倍に [dB] (デシベル) という単位をつけて表記することがよく行われる. たとえば 3 [dB] は約 2 倍の減衰率, 10 [dB] は 10 倍の減衰率, 60 [dB] は 100 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

\mathcal{A}_{\max} の値が大きすぎると通過帯域内の減衰率の均一性が悪化し, 得られる固有対の精度が不揃いになったり, 極端な場合は必要な固有対が欠落して求まらなかつたりすることも起こりうる. 小さすぎれば次数 n が増える. そこでたとえば \mathcal{A}_{\max} を 3 [dB] 程度にとることにする.

\mathcal{A}_{\min} の値が十分に大きくないと, フィルタによる不要な固有ベクトルの除去が十分ではなくなり, 特異値の相対閾値の設定も難しくなる. ただし, この値を大きくすればそれだけ必要な次数 n が増加してしまい, レゾルベントの個数を多く必要とする. 一方で, 浮動小

4 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

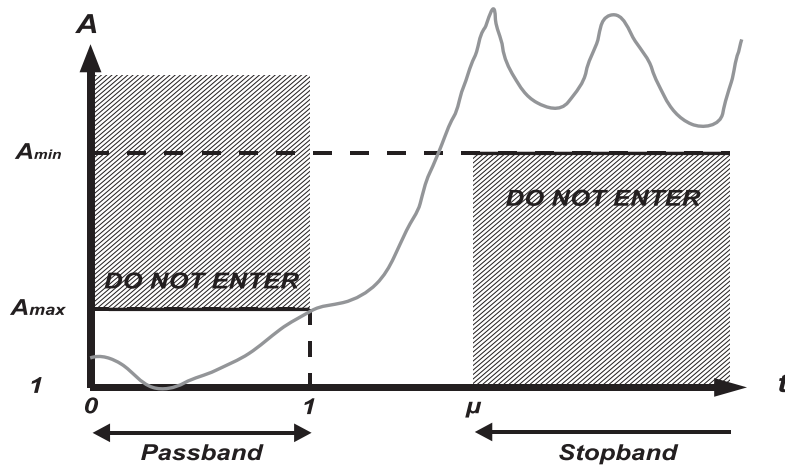


図 1 減衰率関数の形状パラメータ
Fig. 1 Shape parameters for attenuation function.

数点数の精度が 10 進で D 桁のとき、固有値が阻止帯域にある固有ベクトルの強度を減衰させる効果は A_{\min} の値が $10D$ [dB] の程度で飽和し、それ以上値を増やしてもフィルタの必要な次数 n が増すだけで非効率になる。なぜならば、入力ベクトルにフィルタを作用させる際の計算が仮にまったく誤差なしで行えたとして、入力ベクトルに含まれる固有ベクトルに対して、固有値が通過帯域にある方の強度はあまり減衰させず、他方で阻止帯域にある方の強度を D 桁以上減衰させて出力ベクトルを得たとして、その演算結果を最終的に D 桁の精度の浮動小数点数に丸めるとするならば、相対的に強度を D 桁以上に減衰させられた固有ベクトルの方は丸め誤差に覆われて、実質的に出力の値には反映されなくなるからである（これを確認する実験の例を「付記 2」に後述）（注：ただし、 A_{\min} の値を $10D$ [dB] を超えて増大させても、固有値が遷移領域に属する固有ベクトルに対する減衰効果は飽和せず増加する可能性は残る）。そこで、たとえば倍精度演算では 150 [dB] 程度とする（あるいはそれ以下の値 100 [dB] などとする）。

値 $\mu (> 1)$ を大きくすると遷移領域 $1 < |t| < \mu$ の幅が広がり、阻止帯域にはない固有値を持つ固有ベクトルの個数が増加する傾向が一般には生じる、その結果フィルタの出力が張る部分空間の次元も増える傾向にある。するとそれに比例して入力ベクトルの個数を増す必要性も出てくる。また、固有値が遷移領域にある固有ベクトルに対しては、フィルタの

減衰率は A_{\max} と A_{\min} の中間の領域の値を持つため、そのような固有ベクトルが多く存在するとフィルタの出力ベクトルの組の特異値の分布からは明瞭な境界が失われることになる。 $\mu=1.1$ かそれより小さな値が適当であろう。他方で μ を 1 に近づけると次数 n は増加する（Butterworth フィルタでは増加は急激で、Chebyshev や Inverse Chebyshev フィルタでは増加は緩やかで、Elliptic フィルタでは μ を 1 に非常に近くしても n の増加はきわめて緩やかである）。 μ を 1 に近づけた場合の利点は、フィルタの遷移領域に入る（算法には都合が良くない）固有値の存在確率を減らせることである。もしも結果的に遷移領域に入る固有値が存在しなければ、特異値分布の分離は容易で、またフィルタ対角化法の結果の固有対の精度も高くなる。また、 μ を 1 に近づけると、フィルタにより生成される部分空間の次元数が指定区間に固有値を持つ固有ベクトルの真の個数に近づくので、計算に必要な入力ベクトルと出力ベクトルの個数が減る。ただし、必要となる固有対の固有値の存在範囲を最初から正確に知ったうえで計算を行うことは稀であろうし、 μ を著しく 1 に近づけて遷移領域の幅を狭くすると、指定区間の端付近の固有値を持つ解は摂動により固有値が少しずれると属する帯域が変わり、求まったり求まらなかったりすることになる。

3.4 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) フィルタ特性への要求事項として、3 個の形状パラメータ μ, A_{\max}, A_{\min} を与える。
- (2) 使用するフィルタの種類を後述の Butterworth, Chebyshev, Inverse Chebyshev, Elliptic の中から選ぶ（上記フィルタの種類に対応して、減衰率関数 $\mathcal{A}(t)$ が具体的な関数形としてパラメータ n, ϵ を残して決まる）。
- (3) フィルタ特性の 3 個の形状パラメータから、 $|t| \leq 1$ のときには $\mathcal{A}(t) \leq A_{\max}$ （この制約条件は tight であると仮定した）、 $\mu \leq |t|$ のときには $A_{\min} \leq \mathcal{A}(t)$ を満たすように $\mathcal{A}(t)$ の持つパラメータ（代表的な四種類のフィルタの場合は n, ϵ ）を決める。
- (4) 複素平面内で $g(t) \equiv 1/\mathcal{A}(t)$ の部分分数展開の極の位置と極の係数を求める（代表的な 4 種類のフィルタの伝達関数については、すべての極は重複がなく、極の位置は解析的表式で計算できる¹⁷⁾）。さらに、 $g(t)$ の部分分数展開の極 $t = t_p$ に対する係数 c_p は、ロピタルの定理から導関数を用いて $c_p = 1/\mathcal{A}'(t_p)$ により求められる。
- (5) 得られた部分分数展開の極とその係数に対応して、フィルタ作用素のレゾルベントの分点（シフト量）と結合係数が決まり、フィルタ作用素が決定される。

3.5 注記：恒等演算子の項の係数について

フィルタ作用素の表式では数学的説明の統一性のため、レゾルベントの線形結合以外にも

5 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

恒等演算子の定数倍も付け加えておいたが、その係数 c_∞ は伝達関数 $g(t)$ の無限遠での値 $g(\infty)$ (≥ 0) に等しく、減衰率関数 $\mathcal{A}(t)$ が多項式である Butterworth や Chebyshev では値はつねに零となり、減衰率関数が多項式ではない Inverse Chebyshev や Elliptic では n が奇数の場合は零だが、 n が偶数の場合は零にはならない。しかしもともとフィルタの形状は阻止帯域で伝達関数の値がきわめて小さくなるように設定しており、 $g(\infty) \leq 1/\mathcal{A}_{\min}$ を満たすので、通常の設計においては無視しうる大きさとなり、フィルタの機能は恒等演算子の定数倍の項を省いても実質的には変化しない。

3.6 フィルタを作用する際の共役対称性の利用

以降で見るように減衰率関数 $\mathcal{A}(t)$ が正值 ($1 + \text{実関数の二乗の形}$) の場合にはフィルタ伝達関数 $g(t)$ の極はすべて虚数で、必ず複素共役の対の組で現れる。さらに互いに複素共役である極の対応する係数は互いに複素共役になる。このことから、実数ベクトル \mathbf{v} に対して (恒等演算子の部分を除いた) フィルタを作用させる計算は、 $2n$ 個のレゾルベントすべてではなくてその半分の、複素共役の関係にある対の片側ずつを集めた合計 n 個のレゾルベントだけを用いて実現でき、演算量と作業記憶量を半減させることができる。たと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、 $\sum_{p=1}^{2n} \gamma_p R(\lambda_p) \mathbf{v} = \sum_{\text{Im} \lambda_q > 0} \text{Re}\{2\gamma_q R(\lambda_q) \mathbf{v}\}$ となる。このようにすると途中には複素数演算が入るが、計算結果が実数ベクトルになることが自然に保証される。

4. 代表的な 4 種類のフィルタの実現

4.1 Butterworth フィルタ

Butterworth フィルタの減衰率関数は t の $2n$ 次の多項式で偶関数、 n を自然数、 ϵ を実数のパラメータとして $\mathcal{A}(t) \equiv 1 + \epsilon^2 t^{2n}$ で表される関数形を持つ。

4.1.1 制約条件を満たす次数の決定

冪乗関数値の性質から、フィルタ特性の形状の制約条件は $\mathcal{A}_{\max} = 1 + \epsilon^2$, $\mathcal{A}_{\min} \leq 1 + \epsilon^2 \mu^{2n}$ 。すると $\epsilon^2 = \mathcal{A}_{\max} - 1$ で、 $\mu^n \geq L_{\min}$ から、実数 $n_{\min} \equiv \ln(L_{\min}) / \ln(\mu)$ とおくと、制約条件は、自然数 n を実数 n_{\min} 以上の値にとれば満たされる。

4.1.2 極とその係数

伝達関数 $g(t)$ の極は $t_p = 1/\epsilon^{1/n} (\cos \theta_p + \sqrt{-1} \sin \theta_p)$ 、ここで、 $\theta_p \equiv (2p-1)\pi/(2n)$ 、 $p = 1, 2, \dots, 2n$ 。極は複素平面上で円周上の $2n$ 等分点で、 n 組の複素共役対になっている。添字 $p=1, 2, \dots, n$ の極は虚部がすべて正である。極 t_p の係数は $c_p = -t_p/(2n)$ で、係数 c_∞ はつねに零である。

4.2 Chebyshev フィルタ

Chebyshev フィルタの減衰率関数は t の $2n$ 次の多項式で偶関数、自然数 n と ϵ を実数パラメータとして $\mathcal{A}(t) \equiv 1 + \epsilon^2 T_n^2(t)$ で表される関数形を持つ。ここで $T_k(x)$ は次数 k の Chebyshev 多項式を表す。この関数は通過帯域内で極大値極小値がそれぞれすべて一致する等リプル性を持つ。Chebyshev フィルタは (通過帯域にリプルを許したことにより) 要求されたフィルタ特性の形状の制約を Butterworth フィルタより低い次数で実現できる。

4.2.1 制約条件を満たす次数の決定

Chebyshev 多項式の関数値の性質から、フィルタ特性の形状の制約は $\mathcal{A}_{\max} = 1 + \epsilon^2$, $\mathcal{A}_{\min} \leq 1 + \epsilon^2 T_n^2(\mu)$ 。すると $\epsilon^2 = \mathcal{A}_{\max} - 1$, $T_n(\mu) \geq L_{\min}$ 。これより実数 $n_{\min} \equiv \cosh^{-1}(L_{\min}) / \cosh^{-1}(\mu)$ とおくと制約条件は、自然数 n を実数 n_{\min} 以上の値にとれば満たされる。

4.2.2 極とその係数

フィルタの極と係数は、Chebyshev 多項式の (複素) 媒介変数 ϕ を用いた表示 $T_n(t) = \cos(n\phi)$, $t = \cos \phi$ を利用すると、解析的な式で表せる。 $g(t)$ の極 t_p の計算式は: $t_p = \cosh \tau \cdot \cos \theta_p + \sqrt{-1} \sinh \tau \cdot \sin \theta_p$ 、ただし $\tau = (1/n) \sinh^{-1}(1/\epsilon)$, $\theta_p = (2p-1)\pi/(2n)$, $p=1, 2, \dots, 2n$ 。極は複素平面内の -1 と 1 を焦点とする楕円の周上にあり、 n 組の複素共役対になっている。虚部が正の極は添字 $p=1, 2, \dots, n$ に対応する。極の係数 c_p は: $c_p = (-1/2)T_n(t_p)/T_n'(t_p) = (-1/2)T_n(t_p)/(nU_{n-1}(t_p))$ 。ここで $U_k(x)$ は次数 k の第二種 Chebyshev 多項式を表す。係数 c_∞ はつねに零である。

4.3 Inverse Chebyshev フィルタ

Inverse Chebyshev フィルタの減衰率関数は t の $2n$ 次の有理関数で偶関数、自然数 n 以外に ϵ を実数パラメータとして $\mathcal{A}(t) \equiv 1 + \epsilon^2 \{T_n(\mu)/T_n(\mu/t)\}^2$ で表される関数形を持つ。

この関数型は、阻止帯域で $\mathcal{A}(t)$ の極小値がすべて一致する等リプル性を持つ。Inverse Chebyshev フィルタは (阻止帯域にリプルを許したこと) 要求されたフィルタ特性の形状の制約を Butterworth フィルタより低い次数で実現できる。

4.3.1 制約条件を満たす次数の決定

関数値の振舞いを考慮すれば、フィルタ特性の形状の制約は $\mathcal{A}_{\max} = 1 + \epsilon^2$, $\mathcal{A}_{\min} \leq 1 + \epsilon^2 T_n^2(\mu)$ から、 $\epsilon^2 = \mathcal{A}_{\max} - 1$, $T_n(\mu) \geq L_{\min}$ となる。実数 $n_{\min} \equiv \cosh^{-1}(L_{\min}) / \cosh^{-1}(\mu)$ とおくと、制約条件は、自然数 n を実数 n_{\min} 以上の値にとれば満たされる。この n_{\min} の式は Chebyshev フィルタに対するものと同一である。

4.3.2 極とその係数

この場合にも Chebyshev の場合と同様に極とその係数は解析的な式で表せる。いま、 $1/c = \epsilon T_n(\mu)$, $\tau = (1/n) \sinh^{-1}(1/c)$, $\theta_p = (2p-1)\pi/(2n)$, $p=1, 2, \dots, 2n$ と定義する。すると $g(t) = 1/A(t)$ の極は $t_p = \mu/x_p$ と表される。ここで $x_p = \cosh \tau \cdot \cos \theta_p - \sqrt{-1} \sinh \tau \cdot \sin \theta_p$ ($g(t)$ の零点は $z_j = \mu/\cos \theta_j$, $j=1, 2, \dots, 2n$ である)。極の係数は： $c_p = -\mu/(2x_p^2) T_n(x_p)/T'_n(x_p) = -\mu/(2x_p^2) T_n(x_p)/(nU_{n-1}(x_p))$ である。係数 c_∞ は、 n が奇数のとき零で、偶数のとき $1/(1+(1/c)^2)$ である。

4.4 Elliptic フィルタ

Elliptic フィルタの減衰率関数は t の $2n$ 次の有理関数で偶関数、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。 R_n は次数 n の有理関数で、楕円有理関数あるいは Chebyshev 有理関数とも呼ばれ、Jacobi の楕円関数 sn を用いた (複素) 媒介変数 ϕ による表示 $R_n(t) = \text{sn}[K(1/L)(n\phi + \Delta_n), 1/L]$, $t = \text{sn}[K(1/\mu)\phi, 1/\mu]$ を持つ。 $K(k)$ は第一種完全楕円積分であり、記号 Δ_n は n が奇数のとき 0、偶数のとき $(-1)^{n/2}$ を表す。

この減衰率関数 $A(t)$ は、通過帯域での極大値と極小値がそれぞれすべて一致する等リプル性と、阻止帯域での極小値がすべて一致する等リプル性を持つ。Elliptic フィルタは (通過帯域と阻止帯域の両方にリプルを許したことで) 要求されたフィルタ特性の形状の制約を Chebyshev フィルタや Inverse Chebyshev フィルタより低い次数で実現できる。

4.4.1 制約条件を満たす次数の決定

関数値の振舞いを考慮すると、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 $n_{\min} \equiv \{K'(1/L_{\min})/K(1/L_{\min})\} / \{K'(1/\mu)/K(1/\mu)\}$ とおく。 $K(k)$ は第一種完全楕円積分で、 $K'(k) \equiv K(\sqrt{1-k^2})$ 。制約条件は、自然数 n を実数 n_{\min} 以上の値にとれば満たされる。 μ と n の値から L の値は： $1/L = (1/\mu)^n \prod_{j=1}^{\text{floor}(n/2)} \text{sn}^4[(2j-1)K(1/\mu)/n, 1/\mu]$ により計算できる。

4.4.2 極とその係数

楕円関数による媒介変数表示を用いると、 $g(t) = 1/A(t)$ の極 t_p とその係数 c_p は解析的な式で表せて、以下の手順で求められる。まず $b \equiv F(\tan^{-1}(1/\epsilon), \sqrt{1-L^{-2}})$ を計算する。ここで $F(\phi, k) \equiv \int_0^\phi (1-k^2 \sin^2 x)^{-1/2} dx$ は第一種楕円積分である。次に $\tau \equiv (b/n)K(1/\mu)/K(1/L)$, $\theta_p = (2p+1 - \text{mod}(n, 2))K(1/\mu)/n$ 。すると極の値は、 $t_p = \text{sn}(\theta_p + \sqrt{-1}\tau, \mu^{-1})$, $p=1, 2, \dots, 2n$ (注意：この式では虚部が正の極に対応する添字は

$p=1, 2, \dots, n$ ではない)。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 が奇数のとき零で、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 の値を組み合わせた式で計算できる。第一種完全楕円積分 $K(k)$ や第一種楕円積分 $F(\phi, k)$, 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¹⁸⁾を用いることができる。その際に、関数定義が異なっていたり、引数が母数 k の値ではなくて k^2 の値とする実装がされていたりする場合がある。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 や 1 の付近で関数値の精度が十分出ないものがあり、注意や検討が必要である。

文献 17) にも、アナログ回路の場合に則した形で楕円フィルタの伝達関数の極を解析的方式を用いて求める手順が詳しく書かれている。

注記：Inverse elliptic は Elliptic と一致する

Chebyshev に対する Inverse Chebyshev と同様の方法で Elliptic に対する Inverse elliptic に相当するものを作れば、それは Elliptic と一致する。いま自然数 n , 正実数 ϵ をパラメータとして Inverse elliptic の減衰率関数を Inverse Chebyshev と同様に $\tilde{A}(t) \equiv 1 + \epsilon^2 \{R_n(\mu)/R_n(\mu/t)\}^2 = 1 + \epsilon^2 L^2 / R_n^2(\mu/t)$ とおくと、有理関数 $R_n(t)$ の持つ性質 $R_n(\mu)/R_n(\mu/t) = R_n(t)$ により $1 + \epsilon^2 R_n^2(t)$ と等しくなり、Elliptic の減衰率関数と一致する。同様に、Butterworth に対応する Inverse Butterworth も Butterworth に一致する。

5. 代表的な 4 種類のフィルタの比較

5.1 減衰率関数のグラフの例

代表的な 4 種類のフィルタに対して、フィルタ特性の形状パラメータとして $\mu=1.3$, $A_{\max}=10$ [dB], $A_{\min}=100$ [dB] を指定して必要最小の n の値を決め、それに対する減衰率関数 $A(t)$ をプロットしたものが図 2 (Butterworth, $n=40$), 図 3 (Chebyshev, $n=15$), 図 4 (Inverse Chebyshev, $n=15$), 図 5 (Elliptic, $n=9$) である。 μ の値を 1.3, 透過帯域での上限値 A_{\max} の値を 10 [dB] と、いずれも実際の使用で想定する値よりも大きくしたのは、グラフの様子を理解しやすいようにするためである。4 種類のフィルタの減衰率関数は正規化座標 t の偶関数で、グラフは左右対称なので $t \geq 0$ の側だけを示した。

5.2 フィルタ形状を実現する n の最小値の比較

2 個の表 (表 1, 表 2) はそれぞれ、フィルタの特性の形状への 3 個の要求パラメータの

7 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

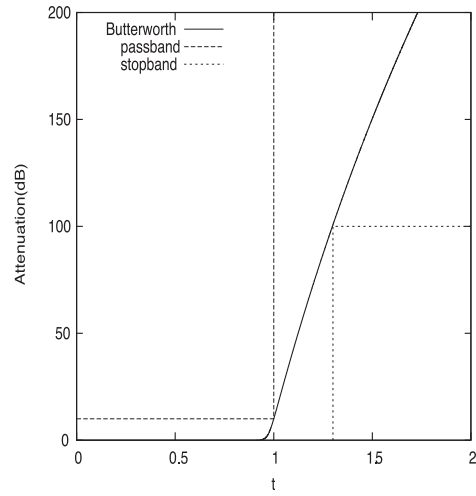


図 2 Butterworth フィルタの減衰率のグラフの例

Fig. 2 Sample graphs of attenuation of Butterworth. ($\mu=1.3$, $\mathcal{A}_{\max}=10$ [dB], $\mathcal{A}_{\min}=100$ [dB]).

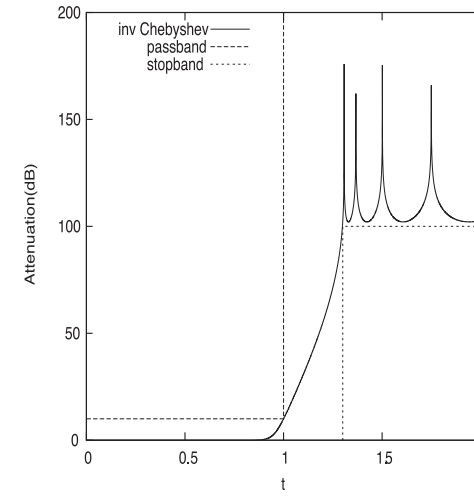


図 4 Inverse Chebyshev フィルタの減衰率のグラフの例

Fig. 4 Sample graphs of attenuation of Inverse Chebyshev. ($\mu=1.3$, $\mathcal{A}_{\max}=10$ [dB], $\mathcal{A}_{\min}=100$ [dB]).

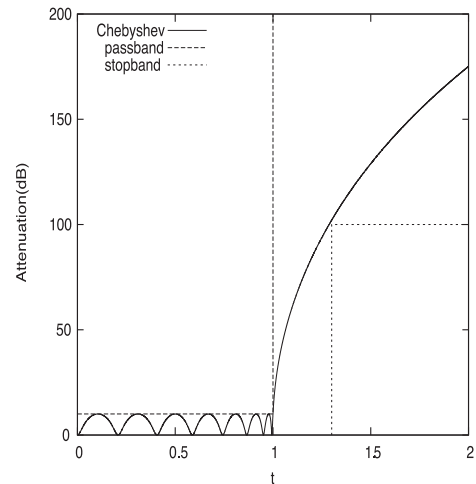


図 3 Chebyshev フィルタの減衰率のグラフの例

Fig. 3 Sample graphs of attenuation of Chebyshev. ($\mu=1.3$, $\mathcal{A}_{\max}=10$ [dB], $\mathcal{A}_{\min}=100$ [dB]).

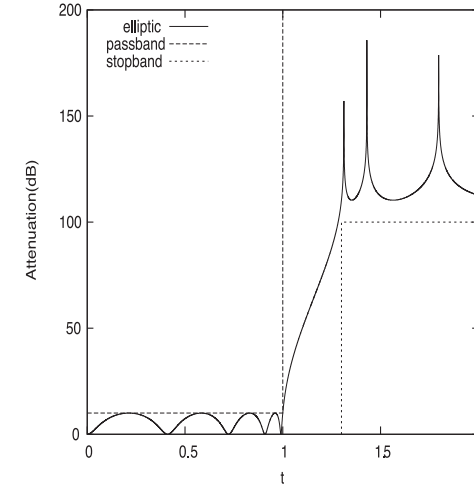


図 5 Elliptic フィルタの減衰率のグラフの例

Fig. 5 Sample graphs of attenuation of Elliptic. ($\mu=1.3$, $\mathcal{A}_{\max}=10$ [dB], $\mathcal{A}_{\min}=100$ [dB]).

8 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

表 1 フィルタ形状の要求を実現可能な n の最小値
 Table 1 Minimum n for the filter shape requirements.
 $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=150$ [dB].

μ	Butterworth	Chebyshev	Elliptic
1.001	17,281	402	35
1.003	5,766	232	30
1.005	3,463	180	28
1.01	1,736	128	26
1.03	585	74	22
1.05	355	58	20
1.1	182	41	17
1.2	95	29	15
1.3	66	24	13
1.5	43	19	12

表 2 フィルタ形状の要求を実現可能な n の最小値
 Table 2 Minimum n for the filter shape requirements.
 $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB].

μ	Butterworth	Chebyshev	Elliptic
1.001	11,522	274	24
1.003	3,845	158	21
1.005	2,309	123	20
1.01	1,158	87	18
1.03	390	50	15
1.05	237	39	14
1.1	121	28	12
1.2	64	20	10
1.3	44	17	9
1.5	29	13	8

うち \mathcal{A}_{\max} , \mathcal{A}_{\min} の 2 個を固定している．表 1 は $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=150$ [dB] の場合，表 2 は $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB] の場合である．表には残り 1 個の要求パラメータ μ の値と，3 個の要求パラメータから決まる自然数 n の最小値をフィルタの種類ごとに並べて記入した．Inverse Chebyshev の場合は n の最小値が Chebyshev と同一なので省略した．表の数値から同じ要求を実現するための n の最小値が Butterworth よりも Chebyshev (あるいは Inverse Chebyshev) では小さく，さらに Chebyshev よりも Elliptic では小さいことが分かる．次数の違いは μ が 1 に近いほど顕著になる (後で漸近挙動について述べる)．

図 6, 図 7 の 2 枚の図は，先ほどの 2 個の表にそれぞれ対応していて，フィルタの特性の形状への 3 個の要求パラメータのうち \mathcal{A}_{\max} , \mathcal{A}_{\min} の 2 個を固定している．図 6 が

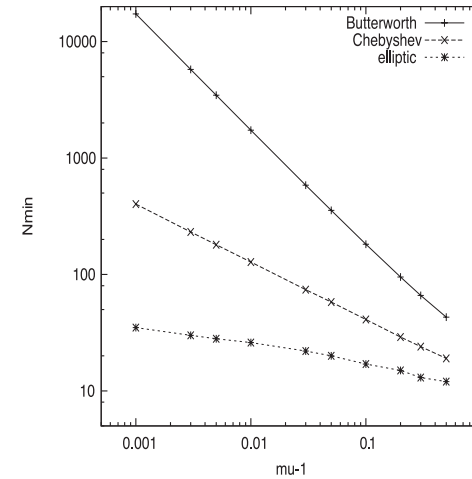


図 6 フィルタ形状の要求を実現可能な n の最小値のグラフ
 Fig. 6 Graphs of minimum n for the filter shape requirements. $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=150$ [dB].

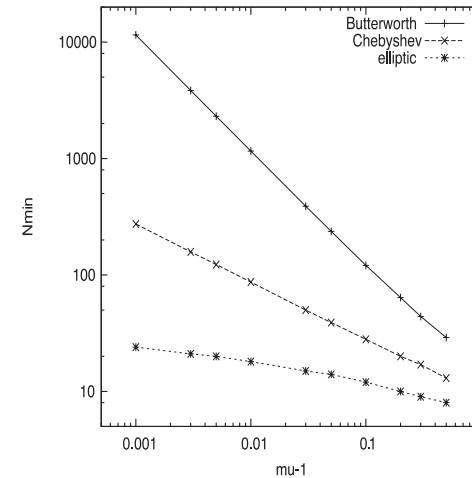


図 7 フィルタ形状の要求を実現可能な n の最小値のグラフ
 Fig. 7 Graphs of minimum n for the filter shape requirements. $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB].

9 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

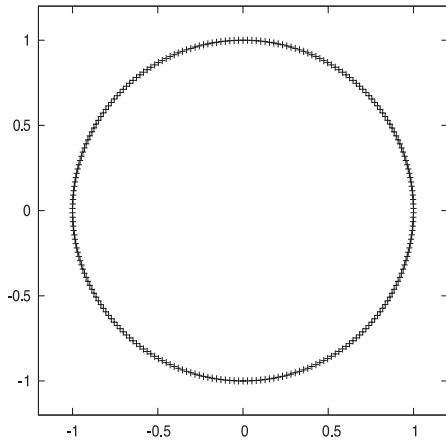


図 8 Butterworth フィルタの極の位置のプロットの例
Fig. 8 Sample plots of poles of Butterworth Filters.
 $\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB].

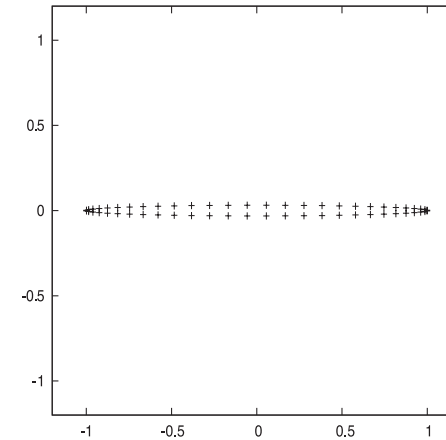


図 9 Chebyshev フィルタの極の位置のプロットの例
Fig. 9 Sample plots of poles of Chebyshev Filters. $\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB].

$\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=150$ [dB] の場合で, 図 7 が $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB] の場合である. グラフの横軸は残り 1 個の要求パラメータ μ から 1 を引いた値 ($\mu - 1$) で, グラフの縦軸は 3 個の要求パラメータから決まる自然数 n の最小値である. 両軸とも対数プロットにしてある. Inverse Chebyshev の場合は n の最小値が Chebyshev と同一なので省略した. 各グラフの曲線は上から Butterworth, Chebyshev, Elliptic である. 同じ要求を実現するための n の最小値が Butterworth よりも Chebyshev (あるいは Inverse Chebyshev) の方が小さく, またさらに Chebyshev よりも Elliptic の方が小さいことがよく分かる. 最小値の大きさの違いは μ が 1 に近づくほど顕著になる.

5.3 フィルタの極の分布の例

フィルタ特性の形状への同一の要求 $\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB] を満たして次数が最小の伝達関数の極の分布を, 複素平面上にプロットした例を図 8 (Butterworth, $n=121$), 図 9 (Chebyshev, $n=28$), 図 10 (Inverse Chebyshev, $n=28$), 図 11 (Elliptic, $n=12$) に示す.

5.4 フィルタ次数の下限値の漸近近似式

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) において, フィルタ特性の形状の要求条件を満たすために必要な次数 n の下限値である実数 n_{\min} の漸

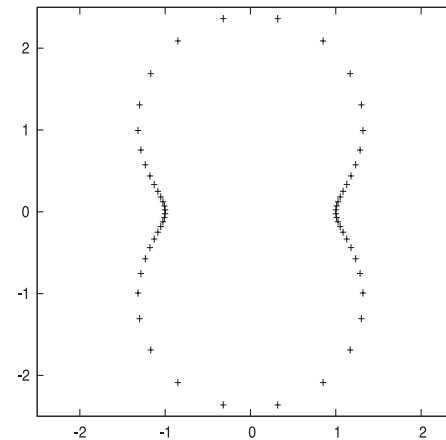


図 10 Inverse Chebyshev フィルタの極の位置のプロットの例
Fig. 10 Sample plots of poles of Inverse Chebyshev Filters. $\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB].

近近似式を以下に示す. これにより, 同等のフィルタ特性を実現するための n の値 (それはフィルタ作用を実現するために必要なレゾルベントの個数になる) が Butterworth フィルタよりも Chebyshev フィルタ (あるいは Inverse Chebyshev フィルタ) では少なく済む

こと、また Elliptic フィルタではさらに少ないことがよく理解できる。

5.4.1 Butterworth フィルタ

$\mu \rightarrow 1$ のとき $\ln \mu \approx \mu - 1$ なので、 $n_{\min} = \ln(L_{\min}) / \ln(\mu) \approx \ln(L_{\min}) / (\mu - 1)$ 。

5.4.2 Chebyshev フィルタ

$\mu \rightarrow 1$ のとき $\cosh^{-1}(\mu) \approx \sqrt{2(\mu-1)}$ である。 L_{\min} の値は非常に大きいので $\cosh^{-1}(L_{\min}) \approx \ln(2L_{\min})$ である。すると $n_{\min} = \cosh^{-1}(L_{\min}) / \cosh^{-1}(\mu) \approx \ln(2L_{\min}) / \sqrt{2(\mu-1)}$ 。

Inverse Chebyshev の下限値 n_{\min} の式は Chebyshev のものと同一である。

5.4.3 Elliptic フィルタ

$K(k)$ 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¹⁹⁾ 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 $k \rightarrow 1$ のときに、 $\lim_{k \rightarrow 1} \{K(k) - (1/2) \ln(16/(1-k^2))\} = 0$ が成立する。結局、 $\mu \rightarrow 1$ に対する n_{\min} の漸近近似式として次のものを得た： $n_{\min} \equiv \{K'(1/L_{\min})/K(1/L_{\min})\} / \{K'(1/\mu)/K(1/\mu)\} \approx (2/\pi^2) \ln(4L_{\min}) \ln(8/\ln \mu)$ 。数値的に試してみると、この近似式は正確な値ときわめて近い。

5.4.4 漸近近似式のまとめ

いま、 $x \equiv 1/(\mu-1)$ とおくと、 $\mu \rightarrow 1$ のとき $x \rightarrow \infty$ であり、そのとき実数 n_{\min} の漸近的な

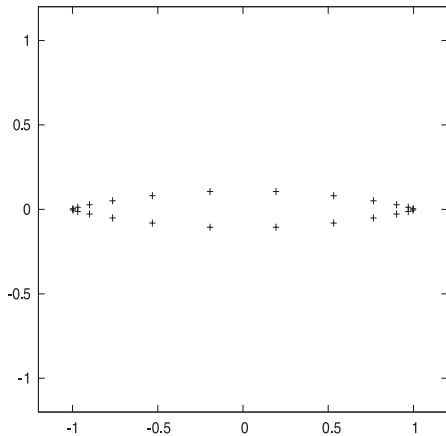


図 11 Elliptic フィルタの極の位置のプロットの例
Fig. 11 Sample plots of poles of Elliptic Filters.
 $\mu=1.1, A_{\max}=3$ [dB], $A_{\min}=100$ [dB].

振舞いは、Butterworth の場合は $n_{\min} \approx \ln(L_{\min})x$ 、Chebyshev あるいは Inverse Chebyshev の場合は $n_{\min} \approx \ln(2L_{\min})\sqrt{x/2}$ 、Elliptic の場合は $n_{\min} \approx (2/\pi^2) \ln(4L_{\min}) \ln(8x)$ となり、 μ が 1 に近づくときの n_{\min} の振舞いの違いがよく分かる。ただし、これらは μ が 1 に近いときの漸近近似式で、正確な値を表すものではない。

6. Elliptic フィルタの数表の例

典型的な 4 種類のフィルタに対しては、特性の形状に対する 3 個の要求条件を与えると、形状の条件を満たすための必要最小な次数が決まる。次数をそれ以上の値に指定すると、条件を満たす伝達関数が決まる。

よく用いる場合の伝達関数の極と係数をあらかじめ数表にしておけば、任意の区間を指定したときにそれを透過帯域とするフィルタ作用素のレゾルベントの線形結合のパラメータがただちに構成できる（実際、標準区間 $t \in [-1, 1]$ を任意の指定区間 $\lambda \in [\alpha, \beta]$ に移す線形一次変換を $\lambda = \mathcal{L}(t)$ とするとき、正規化座標 t での伝達関数の極の座標 t_p からレゾルベントのシフト量は $\lambda_p = \mathcal{L}(t_p)$ となり、伝達関数の極 t_p の係数 c_p を用いてレゾルベントの結

表 3 例 1: Elliptic フィルタ: 正規化座標での極とその係数の値
 $\mu=1.1, A_{\max}=3$ [dB], $A_{\min}=150$ [dB], $n=17$

Table 3 Sample 1: Poles and their coefficients in the normalized coordinate (Elliptic Filter).

番号	実部 (複号同順)	虚部
1, 17	$\pm 0.9988990674636575D+00$ $\mp 0.4208879041201686D-03$	$0.1832241780825514D-02$ $-0.7299481769619925D-03$
2, 16	$\pm 0.9854347625608995D+00$ $\mp 0.5186819992362039D-03$	$0.5927158108762483D-02$ $-0.2361947294880509D-02$
3, 15	$\pm 0.9554033152120891D+00$ $\mp 0.7216244236726592D-03$	$0.1135056762543125D-01$ $-0.4525751747884339D-02$
4, 14	$\pm 0.9023908876059291D+00$ $\mp 0.1030426201848702D-02$	$0.1905329428422484D-01$ $-0.7604384719530673D-02$
5, 13	$\pm 0.8167125542626837D+00$ $\mp 0.1404195176021395D-02$	$0.2984298449862535D-01$ $-0.1192790439816883D-01$
6, 12	$\pm 0.6868077838555658D+00$ $\mp 0.1705866291400565D-02$	$0.4377872029563018D-01$ $-0.1753132683591053D-01$
7, 11	$\pm 0.5036279126962596D+00$ $\mp 0.1679857612187386D-02$	$0.5915500708890110D-01$ $-0.237391948772945D-01$
8, 10	$\pm 0.2684756930096834D+00$ $\mp 0.1085531284204580D-02$	$0.7188213075555887D-01$ $-0.2889753026849765D-01$
9	$0.0000000000000000D+00$ $0.0000000000000000D+00$	$0.7692130784434680D-01$ $-0.3094493130792976D-01$

11 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

表 4 例 2: Elliptic フィルタ: 正規化座標での極とその係数の値
 $\mu=1.01$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=150$ [dB], $n=26$

Table 4 Sample 2: Poles and their coefficients in the normalized coordinate (Elliptic Filter).

番号	実部 (複号同順)	虚部
1, 26	$\pm 0.9998880837186579D+00$ $\mp 0.4294324655540097D-04$	$0.1865290451411704D-03$ $-0.74256618387568955D-04$
2, 25	$\pm 0.9985119810649964D+00$ $\mp 0.5436278294858447D-04$	$0.6092418502130730D-03$ $-0.2425428233504145D-03$
3, 24	$\pm 0.9953941328647955D+00$ $\mp 0.8005956060017231D-04$	$0.1193404600046241D-02$ $-0.4751325189541179D-03$
4, 23	$\pm 0.9897120836393259D+00$ $\mp 0.1262621596046400D-03$	$0.2090580899787998D-02$ $-0.8324256421510467D-03$
5, 22	$\pm 0.9799882313062193D+00$ $\mp 0.2034608987672980D-03$	$0.3525808071597722D-02$ $-0.1404185329401036D-02$
6, 21	$\pm 0.9637570985418453D+00$ $\mp 0.3271234236003145D-03$	$0.5839127143624647D-02$ $-0.2326255471716706D-02$
7, 20	$\pm 0.9370785079112591D+00$ $\mp 0.5165197017354089D-03$	$0.9527056259846114D-02$ $-0.3797510706166005D-02$
8, 19	$\pm 0.8939411924457809D+00$ $\mp 0.7874093099285791D-03$	$0.1525085222093224D-01$ $-0.6084058535425789D-02$
9, 18	$\pm 0.8258021687698470D+00$ $\mp 0.1130819077704549D-02$	$0.2372486318849914D-01$ $-0.9476206058496224D-02$
10, 17	$\pm 0.7219462647702741D+00$ $\mp 0.1471704458993376D-02$	$0.3532744860413157D-01$ $-0.1413416099463384D-01$
11, 16	$\pm 0.5719424388805603D+00$ $\mp 0.1627185288138456D-02$	$0.4930888668608999D-01$ $-0.1976776261739063D-01$
12, 15	$\pm 0.3713615401674180D+00$ $\mp 0.1347921209490644D-02$	$0.6291091202075054D-01$ $-0.2527013120546175D-01$
13, 14	$\pm 0.1292420678305036D+00$ $\mp 0.5339656571804936D-03$	$0.7161010797562235D-01$ $-0.2880037172946217D-01$

合係数は $\gamma_p = \mathcal{L}' \cdot c_p$ となる。フィルタ作用素は $\mathcal{F} = c_\infty + \sum_p \gamma_p R(\lambda_p)$ で与えられる。

以下に、正規化座標での伝達関数の極と極の係数の数表の例を、Elliptic の場合について 3 通り示す。

例 1: フィルタの特性形状の 3 個の条件は $\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=150$ [dB] で、次数には最小値を採用して $n=17$ とした。その場合の正規化座標 t での Elliptic フィルタの極とその係数を表 3 に示す。上段が極の値、下段がその係数値である (この表の数値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)。複素共役対称性により、表には虚部が正である極とその係数だけを、極の実部が大

表 5 例 3: Elliptic フィルタ: 正規化座標での極とその係数の値
 $\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB], $n=12$

Table 5 Sample 3: Poles and their coefficients in the normalized coordinate (Elliptic Filter).

番号	実部 (複号同順)	虚部
1, 12	$\pm 0.9978032747225308D+00$ $\mp 0.8664302031627370D-03$	$0.3701716028012426D-02$ $-0.1465613329228266D-02$
2, 11	$\pm 0.9697327336698406D+00$ $\mp 0.1266917017646285D-02$	$0.1285709956095331D-01$ $-0.5096119758585352D-02$
3, 10	$\pm 0.9008627058055854D+00$ $\mp 0.2095891409389544D-02$	$0.2747634325702563D-01$ $-0.1091872191677904D-01$
4, 9	$\pm 0.7653097823348445D+00$ $\mp 0.3144348312333809D-02$	$0.507893965526350D-01$ $-0.2027351913073286D-01$
5, 8	$\pm 0.5316970895763372D+00$ $\mp 0.3449070765877447D-02$	$0.8117259779721395D-01$ $-0.3259458505193008D-01$
6, 7	$\pm 0.1933304451465211D+00$ $\mp 0.1624414004965714D-02$	$0.1054819597085268D+00$ $-0.4255796367297806D-01$

きいものから順に番号を付けて載せている。また $g(t)$ が実数の偶関数なので、極とその係数の実部の符号を同時に反転できる対称性を持つことを利用して、実部の符号を複号同順とすることで表を小さくまとめている。 n が奇数なので、定数項の係数 c_∞ は零である。

例 2: フィルタの特性形状の 3 個の条件は $\mu=1.01$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=150$ [dB] で、次数には最小値 $n=26$ を採用した。その場合の正規化座標 t での Elliptic フィルタの極とその係数を表 4 に示す。上段が極の値、下段がその係数値である (この表の数値は倍精度計算用である)。この表は虚部が正である極とその係数だけを、極の実部が大きいものから順に番号を付けて載せている。また、極とその係数の実部の符号を複号同順として表を小さくまとめている。 n が偶数なので定数項の係数 $c_\infty = g(\infty)$ は非零だが、その値 ($0.3522457843321021D-15$) は $1/\mathcal{A}_{\min}$ 以下で省略できる。

μ の値を例 1 の場合の 1.1 から 1.01 へと小さく変えており、例 1 と比べて遷移領域が狭く、固有値に基づく固有ベクトルの選別能力が鋭くなっている。その代わりに次数は例 1 の $n=17$ から $n=26$ に増えている。

例 3: フィルタの特性形状の 3 個の条件は $\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=100$ [dB] で、次数には最小値 $n=12$ を採用した。その場合の正規化座標 t での Elliptic フィルタの極とその係数を表 5 に示す。上段が極の値、下段がその係数値である (この表の数値は倍精度計算用である)。この表は虚部が正である極とその係数だけを、極の実部が大きいものから順に番号を付けて載せている。また、極とその係数の実部の符号を複号同順として表を小さくまとめている。

n が偶数なので、定数項の係数 $c_\infty = g(\infty)$ は非零だが、その値 (0.3945037918959866D-10) は $1/A_{\min}$ 以下で省略できる。このフィルタの阻止帯域での減衰率は 10 桁なので、表の数値の有効数字 16 桁は過剰であるが、そのまま載せている。

阻止帯域での減衰率を例 1 の場合の 15 桁から 10 桁に下げているので、例 1 と比べると不要な固有ベクトルの除去の程度が悪くなっている。その代わりに次数は例 1 の $n=17$ から $n=12$ に減っている。

7. 特異値の切断に用いる相対閾値の考察

今現在の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、それを入力ベクトルの組としてフィルタを作用させる。そうして得られた出力のベクトルの組に対して、計量 B による特異値分解を行う。そうして丸め誤差の影響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に含まれている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ともなう変数で、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。

不要な固有成分の除去にフィルタを用いるという趣旨からは、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 $r \equiv A_{\max}/A_{\min}$ よりも小さく設定することには意味がない。閾値として適切な値は、たとえば $r^{2/3}$ と $r^{1/3}$ の間のあたりであろう。

丸め誤差をともな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は ϵ_M^{-1} より大きいことは期待できない。これから閾値として適切な値は計算精度にも関係することが分かる。

フィルタ対角化の算法にとってきわめて不都合な状況は、乱数で生成した入力ベクトルの組が必要な固有ベクトルを含まないかごくわずかしか含んでいない場合である。通常は、そのような状況が生じる確率は小さく、独立と見なせる複数の乱数ベクトルを B -正規直交化したベクトルの組が、ある必要な固有ベクトルを他の固有ベクトルと比べてきわめて微小にしか含んでいないような状況は、乱数ベクトルの個数を増やせばそれだけ稀にしか生じなくなるはずである。必要な固有ベクトルどうしの間で、乱数ベクトルの組への含有率のバラツキが大きいと、閾値をそのバラツキの幅だけさらに下げて設定しなければ、特異値を

切断する影響で近似固有対の相対精度が低下したり、欠落したりすることが起きる。しかし閾値は演算精度に比べて十分な余裕を持って大きい値に設定しないと、切断を行う意味が失われて丸め誤差を拡大してしまうので、下げる場合も限界がある。演算精度の高い計算（たとえば単精度よりも倍精度、倍精度よりも 4 倍精度）では演算精度が低い計算の場合に比べて相対閾値を相当小さい値に設定できるので、必要な固有対が特異値の切断の影響で大きく精度を失ったり欠落したりすることが起きる確率をきわめて小さくできることになる。このような統計分布するデータに対する算法の振舞いに対する理論的考察が必要となる。

8. フィルタ対角化法の実験例

実対称行列 A, B を幅の狭い帯行列にとって、数値実験をいくつか実施した。 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)。

フィルタの作用の実現には、片側軸選択つきの複素帯行列の LU -分解²⁰⁾ を利用し、帯行列 $(A - \lambda_p B)$ を係数とする線形方程式を解いた。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、普通の逆反復法^{21), 22)} 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 $\mathbf{r} \equiv (A - \lambda B)\mathbf{v}$ のノルムを用いて $\Delta \equiv \sqrt{\mathbf{r}^T B^{-1} \mathbf{r}}$ で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場合に Wilkinson 限界として知られている残差の 2-ノルム $\|\mathbf{r}\|_2$ の、実対称定値一般固有値問題への簡単な拡張である)。そのほか逆反復法による改良の際の固有値の変動の大きさが参考となる。

数値実験を行った計算機システムの仕様を表 6 に掲げる。この CPU は計算コアを 4 個持つが、今回の実験では並列化はせずにコアを 1 個だけ用いて計算を行った。

実験の各グラフの説明

例題の固有値はよく分離していたので、フィルタ対角化法で得た近似固有対は素朴に固有対ごとの Rayleigh 商を用いる逆反復法で改良した。フィルタ対角化法の与えた近似対は、すでに精度が良かった。

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、

13 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

表 6 実験に用いた計算機システムの仕様

Table 6 Specifications of the computer system for experiments.

CPU	Intel Core i7 920 (2.66 GHz, 8 MB L3) (コアを 1 個のみ使用).
メモリ	DDR3-1333 PC3-10600 2 GB × 6=12 GB (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.

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.

残差のノルムのグラフには, フィルタ対角化法の中で使用している Rayleigh-Ritz 法で得た近似固有対をすべてプロットしたものと, 近似固有対で固有値が区間 $I = [\alpha, \beta]$ 内のものだけをプロットしたものの 2 種類を示した. 経過時間は「フィルタ対角化法」と「逆反復 2 回」に分けた形で示している.

以下の 4 つの例題 1 から 4 までは, 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). 計算はすべて倍精度演算で行い, 特異値を切断する相対閾値を 10^{-7} として処理した.

例題 1

Elliptic フィルタを用いて, フィルタの形状要求を $\mu=1.1, A_{\max}=3$ [dB], $A_{\min}=150$ [dB] に設定し, 次数 n には形状要求を満たせる最小値 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 12 に示す. 中間の値を持つ特異値が 4 個あるのが分かる. グラフ中の水平線は, 特異値の相対値 10^{-7} での切断位置を表す.

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 13 は部分空間法で得られた近似固有対のすべてについて, 図 14 は固有値が区間 I 内

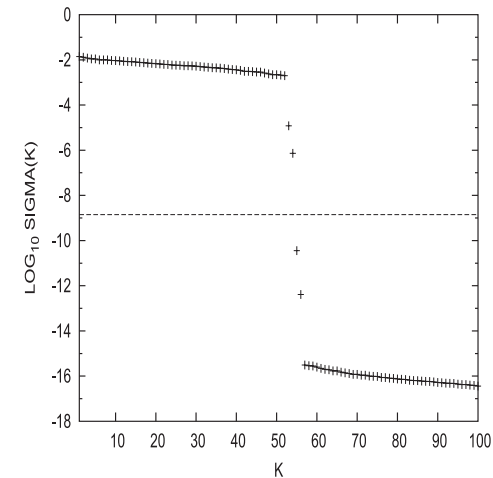


図 12 例題 1: 特異値の分布

Fig. 12 Example1: Distribution of singular values.

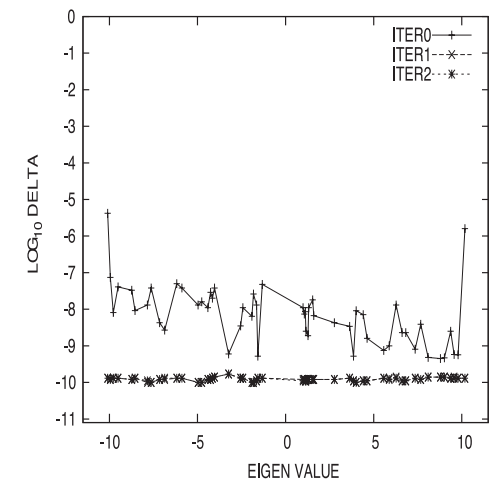


図 13 例題 1: 近似固有対の品質 (部分空間法による全対)

Fig. 13 Example 1: Qualities of approximated eigenpairs (all pairs by subspace method).

14 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

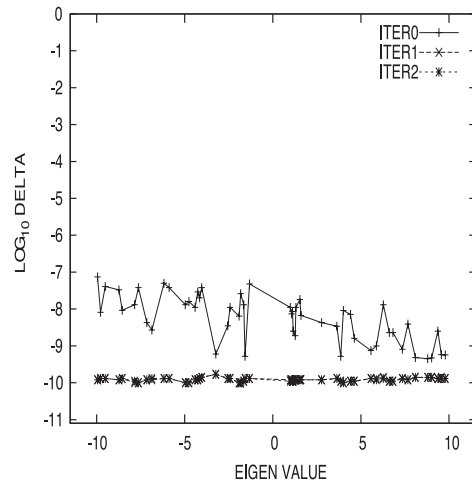


図 14 例題 1：近似固有対の品質（固有値が区間内のもの）

Fig. 14 Example 1: Qualities of approximated eigenpairs (whose eigenvalues are in the interval).

にあ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) した結果を折線プロットしたグラフである。

逆反復は 1 回で十分に収束したので、ITER1 と ITER2 の折線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、 Δ の値が ITER0 では 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} 程度で、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

Elliptic フィルタを用いて、 μ の値を例題 1 の 1.1 より小さく 1.01 とし、フィルタの形状要求を $\mu=1.01$, $A_{\max}=3$ [dB], $A_{\min}=150$ [dB] と設定した。次数 n には形状要求を満たせる最小値 26 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グラフを図 15 に示す。特

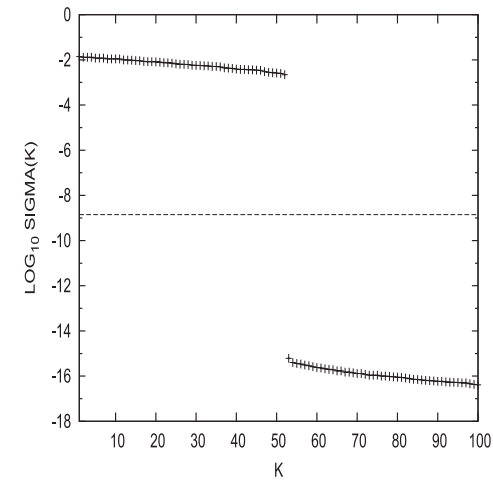


図 15 例題 2：特異値の分布

Fig. 15 Example 2: Distribution of singular values.

異値の分布がはっきり分離していて、中間の値を持つ特異値がないことが分かる。これは μ を小さな値 1.01 にとった効果であるといえる。グラフ中の水平線は、特異値の相対値 10^{-7} での切断位置を表す。

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すべて区間 \mathcal{I} 内にあったので、それらすべ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 16 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} 以下で、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段階で近似固有対の精度がすでにきわめて高いことが分かる。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 桁程度であることが分かる。

15 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

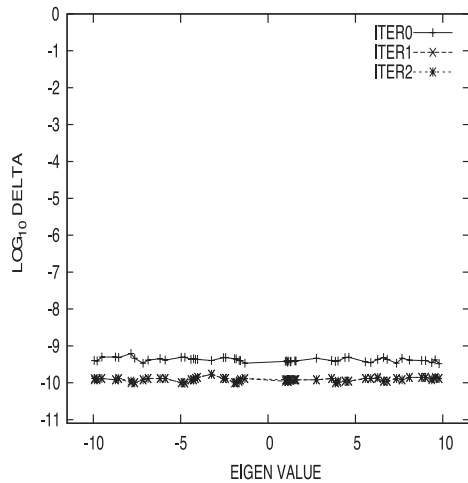


図 16 例題 2 : 近似固有対の品質 (固有値が区間内のもの)

Fig. 16 Example 2: Qualities of approximated eigenpairs (whose eigenvalues are in the interval).

「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、例題 1 の場合の次数 $n=17$ に比べて次数が $n=26$ に増えたためである。

例題 3

Elliptic フィルタを用いて、例題 1 と比べて阻止帯域での減衰量の下限の要求値 A_{\min} を 150 [dB] から 100 [dB] に下げて、フィルタの形状要求を $\mu=1.1$, $A_{\max}=3$ [dB], $A_{\min}=100$ [dB] と設定した。次数 n は形状要求を満たせる最小値 12 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グラフを図 17 に示す。中間の値を持つ特異値が 4 個あるのが分かる。グラフ中の水平線は、特異値の相対値 10^{-7} での切断位置を表す。

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て得られた近似固有対のうち、固有値が \mathcal{I} 内にある近似固有対は 52 個となった。

図 18 は得られた近似固有対のすべてについて、図 19 は得られた近似固有対のうちで固有値が区間 \mathcal{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

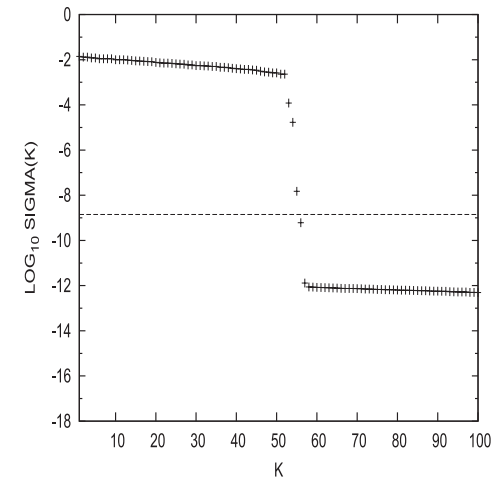


図 17 例題 3 : 特異値の分布

Fig. 17 Example 3: Distribution of singular values.

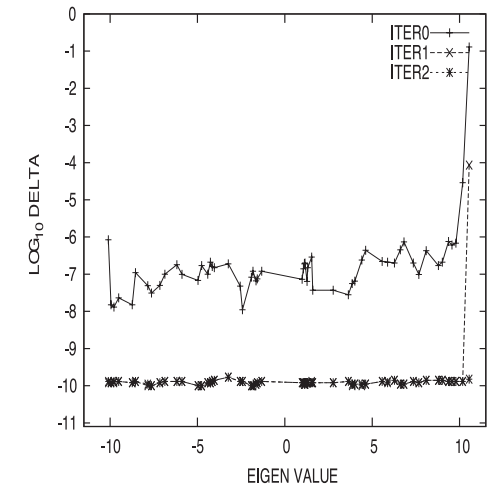


図 18 例題 3 : 近似固有対の品質 (部分空間法による全対)

Fig. 18 Example 3: Qualities of approximated eigenpairs (all pairs by subspace method).

16 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

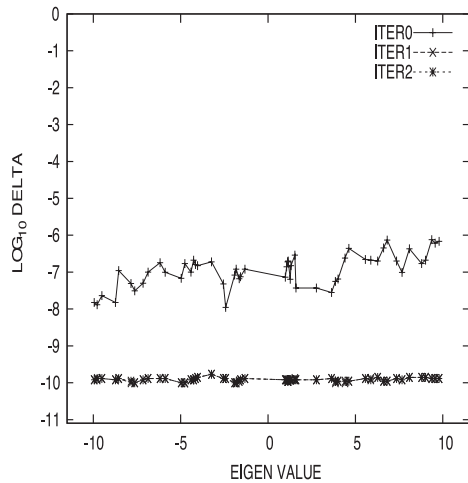


図 19 例題 3 : 近似固有対の品質 (固有値が区間内のもの)

Fig. 19 Example 3: Qualities of approximated eigenpairs (whose eigenvalues are in the interval).

回, 2 回 (ITER1, ITER2) 適用した結果をプロットしたグラフである。

区間からはみ出した固有値を持つ近似固有対については ITER0 では 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分かる。固有値が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} 程度で, 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。フィルタ対角化法の経過時間が例題 1 や例題 2 に比べて短い主な理由は, 例題 1 の $n=17$ や例題 2 の $n=26$ と比べて次数が $n=12$ で小さいためである。

例題 4

フィルタの形状要求は例題 1 とまったく同じ $\mu=1.1$, $A_{\max}=3$ [dB], $A_{\min}=150$ [dB] に設定して, フィルタの種類だけを Elliptic から Inverse Chebyshev に換えた例である。次数 n は形状要求を満たせる最小値 41 を用いた。

得られた特異値のグラフを図 20 に示す。中間の値を持つ特異値が 2 個あるのが分かる。グラフ中の水平線は, 特異値の相対値 10^{-7} での切断位置を表す。

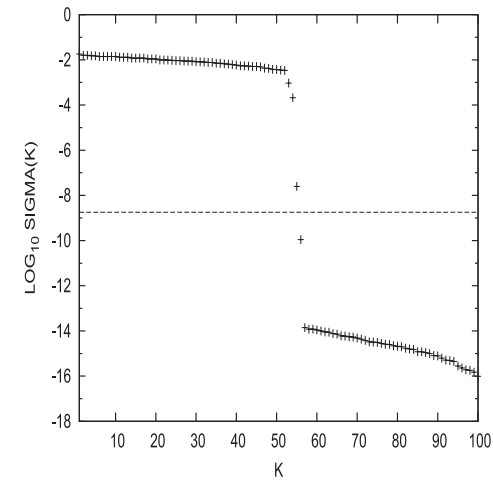


図 20 例題 4 : 特異値の分布

Fig. 20 Example 4: 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 55 となった。フィルタ後の階数 55 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で, 図 21 では全対を, 図 22 では固有値が区間 \mathcal{I} 内の対だけを選び, それぞれ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の結果 (ITER0), 逆反復を 1 回, 2 回適用の結果 (ITER1, ITER2) を折線グラフでプロットしたグラフをそれぞれ示す。

区間からはみ出した固有値を持つ対については ITER0 では Δ の値が 10^{-2} 程度, ITER1 でも 10^{-6} になるものがあることが分かる。区間内の近似固有対について, ITER0 では Δ の値は 10^{-7} 程度以下, ITER1 では Δ の値は 10^{-10} 程度となっており, 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 957 秒と 86 秒であった。例題 1, 例題 2, 例題 3 の各場合と比べてフィルタ対角化法の経過時間が長い理由は主に, 他の場合の各次数 $n=17$, $n=26$, $n=12$ に比べて, 例題 4 の次数 $n=41$ が大きいためである。

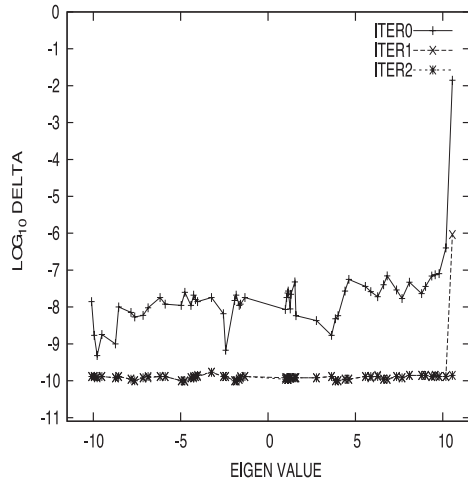


図 21 例題 4：近似固有対の品質（部分空間法による全対）

Fig. 21 Example 4: Qualities of approximated eigenpairs (all pairs by subspace method).

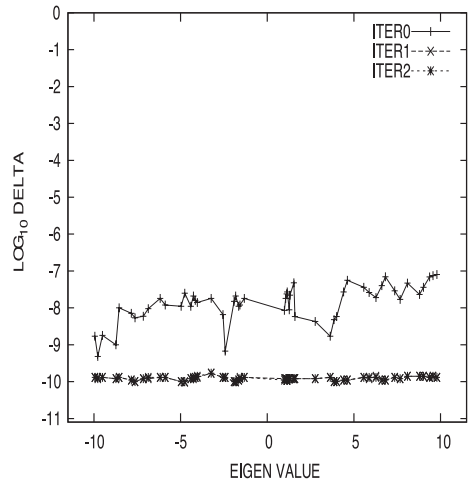


図 22 例題 4：近似固有対の品質（固有値が区間内のもの）

Fig. 22 Example 4: Qualities of approximated eigenpairs (whose eigenvalues are in the interval).

9. ま と め

典型的なフィルタである Butterworth, Chebyshev, Inverse Chebyshev, Elliptic の 4 種類に対して、フィルタの伝達関数（減衰率関数）の特性形状に対する要求からフィルタ作用素のパラメータを具体的に決定する手順を明確にした。

典型的な 4 種類のフィルタについては、特性形状の要求条件 3 個の組（正規化座標での透過帯域と阻止帯域の形状比 μ , 透過帯域における減衰率の最大値 A_{\max} , 阻止帯域における減衰率の下限値 A_{\min} ）を与えれば、フィルタの実現に必要な次数 n の最小値が求まる。

同じ要求条件の組に対してそれを満たせる次数 n の最小値は、Butterworth に比べると Chebyshev あるいは Inverse Chebyshev が小さく、さらに Elliptic が小さい傾向を確認して、その根拠を示した。フィルタの次数 n は（複素共役対称性を用いると）フィルタ作用素を実現する際に用いるレゾルベントの個数と等しい。そのため、同じ基準のフィルタ作用素の場合に必要なレゾルベントの個数は Elliptic の場合が最も少ない。フィルタ対角化法全体の演算量はレゾルベントの個数とともに増すし、並列処理をする場合の必要作業記憶量の合計も同時に扱うレゾルベントの個数とともに増すので、次数 n が小さくとれると計算を行ううえで有利になる。

典型的フィルタの場合は、その種類と 3 個の要求条件の組、それらから求まる必要最小次数以上の値を次数 n として指定すると、要求条件を満たす正規化座標 t での伝達関数 $g(t)$ が決まり、 $g(t)$ の極と極の係数は解析的表式を用いて計算できる（フィルタの種類と要求条件の組を与えると必要最小次数を出力し、その値以上の n を指定すると伝達関数 $g(t)$ の極と極の係数を数値的に求めて出力するプログラムを実装した。伝達関数 $g(t)$ の極と係数の値をあらかじめ求めておけば、与えられた任意の区間を透過帯域とするフィルタ作用素はただちに決まる。実際、標準区間 $[-1, 1]$ を任意の指定区間 $[\alpha, \beta]$ に移す線形一次変換を $\lambda = \mathcal{L}(t)$ とすると、伝達関数の極の座標 t_p を $\mathcal{L}(t)$ で変換したものがレゾルベントのシフト量 λ_p であり、極の係数 c_p を定数 \mathcal{L}' 倍したものがレゾルベントの結合係数 γ_p になる）。

数値実験の例は、フィルタ対角化法に用いるフィルタ作用素をフィルタ特性の形状により指定することが有効であることを示している。

今後は、これらのフィルタをフィルタ対角化に用いてフィルタ特性の形状と特異値分析の閾値の設定との関連性などについて、さらに実験研究を行い、理論的解明を試みるつもりである。

付記 1: 伝達関数の位相特性について

アナログ電気回路の周波数フィルタの伝達関数は複素数値で絶対値と位相を持つ。フィルタを通過する際の各周波数成分ごとの強度変化のみが問題である場合は伝達関数の絶対値のみに着目してフィルタを設計すればよい。フィルタの位相特性はフィルタを通過する際の各周波数成分の時間的遅延や、複数の周波数成分が混ざった信号の周波数成分の間の位相が変化して生じる波形の歪みを問題とする場合には重要となる。

しかし、今回のフィルタ対角化法ではフィルタ作用素は入力ベクトルを不変部分空間へ近似的に射影するためのもので、入力ベクトル中の不要な固有ベクトルは強度を著しく減衰させるが、必要な固有ベクトルは強度をあまり変化させずに出力しさえすればよい、つまり伝達関数の絶対値だけが重要で、伝達関数の位相は重要ではない。もちろん今回のような本来は実数の中で閉じて解ける問題には、計算の途中で極力実数のみを用いて計算するのが好都合である。レゾルベントには複素数のシフト量を入れるので複素演算が含まれるが、複素共役対どうしを組み合わせることで、その組み合わせた計算のブロックごとには実数の入力に対して実数の出力を返す計算の形にできる。

さらに今回の典型的な 4 種類のフィルタの伝達関数 $f(\lambda)$ は値がつねに非負の実数なので、位相特性を持たない。なぜならば、今回の実対称定値一般固有値問題では固有値はすべて実数だから固有ベクトルに対する λ として実数値だけを考えればよく、また t も λ を実係数の線形変換した量だから実数で、 $f(\lambda) = g(t)$ の逆数である減衰率関数 $A(t)$ は、今回紹介した典型的な 4 種類の場合の関数形はどれも実数 t に対して正値を持つ定義となっているからである（ただし、Inverse Chebyshev と Elliptic のフィルタでは、次数 n が偶数の場合は $g(t)$ の部分分数展開の定数項 c_∞ の値は零ではないけれども、通常きわめて小さい正の実数なので、定数項を省略しても実質的にはかまわないことを述べた。そのように修正された伝達関数 $\tilde{f}(\lambda) = g(t) - c_\infty$ の値は、阻止帯域にある固有値 λ の値によってはきわめて微小だが負になって位相を反転することになる。しかし（修正された）伝達関数 $\tilde{f}(\lambda)$ の値がきわめて微小となるような λ を固有値に持つ固有ベクトルは、除去の対象として強い減衰を受けるので、出力にはあまり影響しない）。

LCR 電気回路の場合には、 ω を正規化された周波数とすると、 $2n$ 次の有理関数である減衰率関数を $A(\omega) = H^*(\omega)H(\omega)$ のように分解し、 $H(\omega)$ の零点として $A(\omega)$ の零点のうちで（回路の安定性条件から）虚部が正のものだけを集め、 $H(\omega)$ の有限の位置にある極は A の（すべて偶数位の）有限の位置にある極を位数を半分にして極として集めて作る。こうして作った n 次の有理関数 $H(\omega)$ の逆数を伝達関数とする回路を実現すれば、その伝達関

数は複素数値の関数で、絶対値は $1/\sqrt{A(\omega)}$ 、位相は $\arg\{1/H(\omega)\}$ となる。数値計算による今回の論文のフィルタ作用素では、回路の安定性のような制約は存在せず、また複素共役対称性を利用すると $2n$ 個のレゾルベントのうち半分の n 個だけを使って $2n$ 次のフィルタの計算が簡単にできる。もしも電気回路のフィルタ構成法を忠実に採用して実関数の A を分解し、複素数値の n 次有理関数 H を作り、その逆数の複素数値の伝達関数を作ると、それに対応するフィルタ作用素は実ベクトルの入力に対しても複素ベクトルを出力し、 n 個のレゾルベントを用いる点はまったく同じだが減衰率も分解をしない場合の平方根になるなどで、特に利点は見当たらない。

付記 2: 減衰率の演算精度による限界について

浮動小数点数の精度が 10 進で D 桁のとき、固有値が阻止帯域にある固有ベクトルの強度を減衰させる効果は $10D$ [dB] の程度で飽和し、 A_{\min} の値をそれ以上に設定しても必要なフィルタの次数 n が増すだけで非効率であることを、次のような実験により確かめることができる。

$g(t)$ の値を部分分数分解の式を用いて計算し、その逆数として計算される減衰率のグラフを t について描いてみる。その際に極の座標 t_p や極の係数 c_p の数値の精度、または部分分数分解の式の演算精度が十分でなければ、阻止帯域 $|t| \geq \mu$ で実際に到達できる減衰率の最小値が減少する。つまり不要なベクトルの削減が十分にできない状況になることを示す。そこで極の複素数値と極の係数の複素数値をわざと 10 進で有効桁数 d に丸めて、それを $g(t)$ の式に用いて減衰率を計算する。

- 図 23 は、Chebyshev フィルタに対して、要求 $\mu=1.1$, $A_{\max}=3$ [dB], $A_{\min}=150$ [dB] を与えて、それを満たす最小の値 41 を次数 n として用いた場合の減衰率のグラフである。演算はすべて倍精度で行った。パラメータを丸めた際の有効桁数 d は図中で上側のグラフから順に 15, 12, 9, 6, 3 である。
- 図 24 は、Chebyshev フィルタに対して、要求 $\mu=1.1$, $A_{\max}=3$ [dB], $A_{\min}=300$ [dB] を与えて、それを満たす最小値 80 を次数 n として用いた場合の減衰率のグラフである。演算はすべて 4 倍精度で行った。パラメータを丸めた際の有効桁数 d は図中で上側のグラフから順に 30, 25, 20, 15, 10, 5 である。
- 図 25 は、Elliptic フィルタに対して、要求 $\mu=1.1$, $A_{\max}=3$ [dB], $A_{\min}=150$ [dB] を与えて、それを満たす最小値 17 を次数 n として用いた場合の減衰率のグラフである。演算はすべて倍精度で行った。パラメータを丸めた際の有効桁数 d は図中で上側のグ

19 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

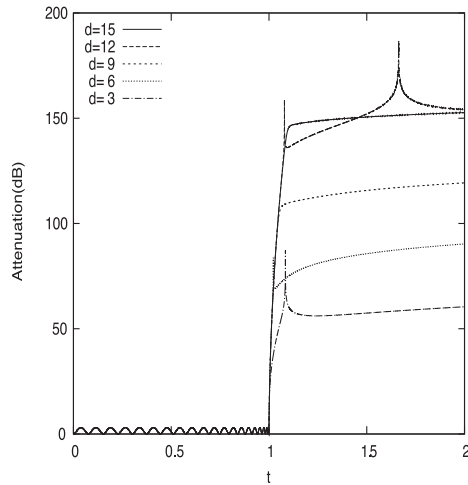


図 23 Chebyshev フィルタのパラメータを丸めた場合の減衰率 (倍精度計算)
 Fig. 23 Attenuation of Chebyshev Filter with rounded parameters (DP calculation).
 ($\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=150$ [dB]), $n=41$.

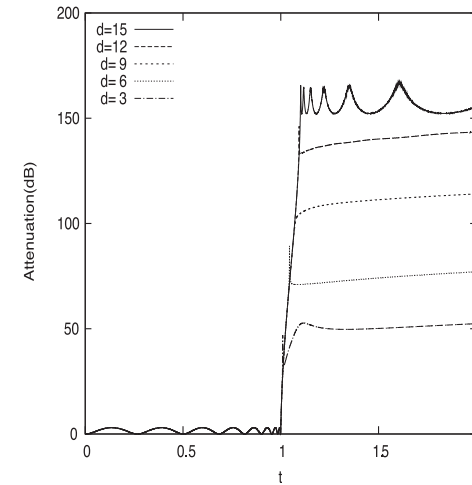


図 25 Elliptic フィルタのパラメータを丸めた場合の減衰率 (倍精度計算)
 Fig. 25 Attenuation of Elliptic Filter with rounded parameters (DP calculation).
 ($\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=150$ [dB]), $n=17$.

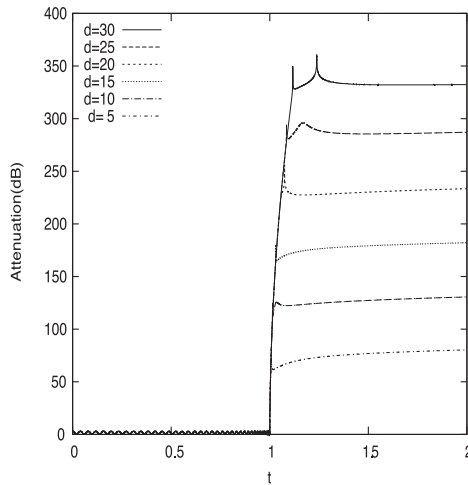


図 24 Chebyshev フィルタのパラメータを丸めた場合の減衰率 (4倍精度計算)
 Fig. 24 Attenuation of Chebyshev Filter with rounded parameters (QP calculation).
 ($\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=300$ [dB]), $n=80$.

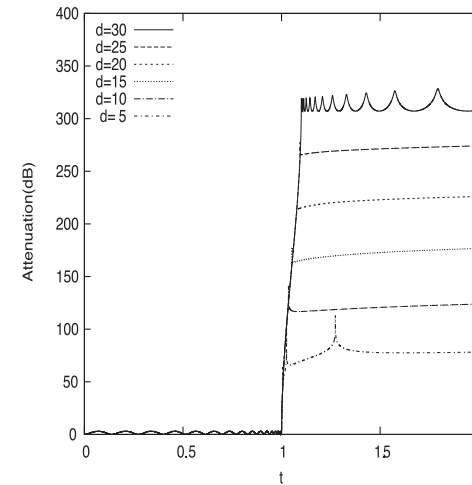


図 26 Elliptic フィルタのパラメータを丸めた場合の減衰率 (倍精度計算)
 Fig. 26 Attenuation of Elliptic Filter with rounded parameters (QP calculation).
 ($\mu=1.1$, $\mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=300$ [dB]), $n=33$.

表 7 Elliptic フィルタの減衰率 (倍精度計算)

Table 7 Attenuation of Elliptic Filter (DP calculation)

 $\mu=1.1, \mathcal{A}_{\max}=3$ [dB].

\mathcal{A}_{\min} ([dB])	n	$\min \mathcal{A}(t)$ ([dB]) in $\mu \leq t \leq 2$
100	12	104
110	13	114
120	14	123
130	15	133
140	16	143
150	17	152
160	18	162
170	18	162*
180	34*	166*
190	34*	166*
200	34*	166*

表 8 Elliptic フィルタの減衰率 (4 倍精度計算)

Table 8 Attenuation of Elliptic Filter (QP calculation)

 $\mu=1.1, \mathcal{A}_{\max}=3$ [dB].

\mathcal{A}_{\min} ([dB])	n	$\min \mathcal{A}(t)$ ([dB]) in $\mu \leq t \leq 2$
150	17	152
160	18	162
170	19	172
180	20	181
190	21	191
200	22	201
210	23	210
220	24	220
230	26	239
240	27	249
250	28	259
260	29	269
270	30	278
280	31	288
290	32	298
300	33	307
310	34	317
320	35	327
330	36	336
340	37	347
350	72*	--

ラフから順に 15, 12, 9, 6, 3 である .

- 図 26 は, Elliptic フィルタに対して, 要求 $\mu=1.1, \mathcal{A}_{\max}=3$ [dB], $\mathcal{A}_{\min}=300$ [dB] を与えて, それを満たす最小値 33 を次数 n として用いた場合の減衰率のグラフである . 演算はすべて 4 倍精度で行った . パラメータを丸めた際の有効桁数 d は図中で上側のグラフから順に 30, 25, 20, 15, 10, 5 である .

これらのグラフから, 有効数字の桁数を丸めて極と係数の精度を減らすと, 阻止帯域 $|t| \geq \mu$ において実現可能な減衰率の下界が減少する傾向が分かる . つまり, 演算の有効桁数により「阻止帯域に於ける減衰率の下界」として実現できる値は制約を受けるので, それを超える値を \mathcal{A}_{\min} に指定しても数値的には実現できない .

さらに, Elliptic フィルタについて, 形状パラメータ $\mu=1.1$ と $\mathcal{A}_{\max}=3$ [dB] を固定し, \mathcal{A}_{\min} の値 ([dB]) を与えて最小の次数 n を計算し, $g(t)$ の部分分数展開の極とその係数を決めて, その部分分数展開を用いて計算した $g(t)$ の値の逆数を減衰率 $\mathcal{A}(t)$ として阻止帯域に属する $\mu \leq |t| \leq 2$ での十分多くの分点に対する最小値を求めた . すべての演算を倍精度で行った場合 (表 7), 170 [dB] 以上の \mathcal{A}_{\min} の値で $\mathcal{A}(t)$ は $\mu \leq |t| \leq 2$ での最小値が要求を正しく満たせていないし, 180 [dB] 以上の \mathcal{A}_{\min} の値において急に n の最小値の計算値が 34 に増えているが, それは倍精度演算では演算精度が不足したからである . 同様にすべての演算を 4 倍精度で行った場合 (表 8), \mathcal{A}_{\min} の値が 350 [dB] で急に n の最小値の計算値が 72 に増えているのも同様で, 4 倍精度演算では演算精度が不足したからである .

参 考 文 献

- 1) Mandelshtam, V. and Taylor, H.: Spectral projection approach to the quantum scattering calculations, *J. Chem. Phys.*, Vol.102, pp.7390–7399 (1995).
- 2) Wall, M. and Neuhauser, D.: Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal mode frequencies from a small number of residues or a short-time segment of a signal. 1. Theory and application to a quantum-dynamics model, *J. Chem. Phys.*, Vol.102, pp.8011–8022 (1995).
- 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. and Taylor, H.: A low-storage filter diagonalization method for quantum eigenenergy calculation for spectral analysis of time signals, *J. Chem. Phys.*, Vol.106, pp.5085–5090 (1996).
- 5) Mandelshtam, V. and Taylor, H.: The quantum resonance spectrum of the H3+ molecular ion for $J = 0$. An accurate calculation using filter diagonalization, *J.*

Chem. Soc. Faraday Trans., Vol.93, pp.847–860 (1997).

- 6) Alacid, M., Leforestier, C. and Moiseyev, N.: Bound and resonance states by a time-independent filter diagonalization method for larger Hamiltonian systems, *Chem. Phys. Lett.*, Vol.305, pp.258–262 (1999).
- 7) 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).
- 8) 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).
- 9) Saad, Y.: Diagonalization algorithms in real space methods for electronic structure calculations, Technical Report PMAA-06, IRISA, Rennes, Dept. Comput. Sci. and Eng., Univ. of Minnesota (2006).
- 10) Zhou, Y., Saad, Y., Tiago, M. and Chelikowsky, J.: Self-Consistent-Field Calculations using Chebyshev Filtered Subspace Iteration, *J. Comput. Phys.*, Vol.219, No.1, pp.172–184 (2006).
- 11) Murakami, H.: The filter Diagonalization Method by the Shifted Inverses, *Proc. ICCM2007 held at Hiroshima (also the CD-ROM file p126_G7-8_proc.pdf)*, p.126 (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 コンピューティングシステム (ACS21)，Vol.49, No.SIG2, pp.66–87 (2008).
- 13) Polizzi, E.: Density-matrix-based algorithm for solving eigenvalue problems, *Phys. Rev. B*, Vol.79, No.11, p.115112 (6 pages) (2009).
- 14) 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).
- 15) 多田野寛人，櫻井鉄也：一般化固有値問題で現れる複素対称連立一次方程式に対する反復解法の性能評価，第 35 回数値解析シンポジウム講演予稿集，pp.61–64 (2006).

- 16) Daniels, R.: *Approximation Methods for Electronic Filter Design*, McGraw-Hill (1974).
- 17) Lutovac, M., Tošić, D. and Evans, B.: *Filter Design for Signal Processing*, Prentice Hall (2001).
- 18) Thompson, W.: *Atlas for Computing Mathematical Functions*, Wiley-Interscience (1997).
- 19) Abramowitz, M. and Stegun, I.: *Handbook of Mathematical Function with Formulas, Graphs, and Mathematical Tables*, Dover (1965).
- 20) Martin, R. and Wilkinson, J.: Solution of symmetric and unsymmetric band equations and the calculations of eigenvectors of band matrices, *Numer. Math.*, Vol.9, pp.279–301 (1967).
- 21) 村田健郎，小国 力，唐木幸比古：スーパーコンピュータ科学技術計算への適用，丸善 (1985).
- 22) 小国力編，村田健郎，三好俊郎，ドンガラ，J.J.，長谷川秀彦：行列計算ソフトウェア，丸善 (1991).

(平成 22 年 1 月 26 日受付)

(平成 22 年 5 月 1 日採録)



村上 弘 (正会員)

現在，首都大学東京・数理情報科学専攻の准教授．研究分野は，数学または科学の問題の，数値的あるいは記号的方法による，効率の良いあるいは精度の良い解法，およびその並列化．1992 年に北海道大学の理学博士号 (化学第二学専攻) を取得．