

# 少数のレゾルベントの線形結合の多項式をフィルタとして用いた実対称定値一般固有値問題の解法

村上 弘<sup>1,a)</sup>

**概要:** 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あるすべての固有対の近似を得るために、その区間に対応する不変部分空間への射影を近似する線形作用素を構成してフィルタとして用いる。本研究で扱うフィルタは少数 2~4 個の複素シフトをもつレゾルベントの線形結合 (の実部) の多項式である。その構成には、まず元のフィルタは単一のレゾルベントの多項式であるとして、その伝達関数を表す有理関数に対して、うまく選んだ低次の有理関数を合成して、改善された伝達特性を表す新しい有理関数を得て、その新しい有理関数を伝達関数を持つ線形作用素がシフトが複素数の少数のレゾルベントの線形結合 (の実部) の多項式で表せることを用いる。

合成に用いる有理関数は、電気回路理論の典型 4 種類のフィルタであるバターワース型、チェビシェフ型、逆チェビシェフ型、楕円型に倣って選ぶ。最初の 3 種類に対応するものについては、これまで既に合成用の有理関数とフィルタの構成法を具体的に示し、それらのフィルタを実際に適用してみて実対称定値一般固有値問題の近似対がうまく求まることを示してきた。しかし最後の楕円型に対応するものはこれまでまだ扱っていなかった。そこで今回は楕円型に対応する場合の、合成用の有理関数とフィルタの構成法を具体的に示し、さらに構成されたフィルタを実際に用いて実対称定値一般固有値問題の近似対がうまく求まることをいくつかの実験例により示す。

**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, 楕円型

## Solution of Real Symmetric Definite Generalized Eigenproblems by using a Filter Which Is a Polynomial of a Linear Combination of a Small Number of Resolvents

**Abstract:** For a real symmetric-definite generalized eigenproblem to obtain all those approximate pairs whose eigenvalues are in a specified interval, we construct a linear operator as the filter which is an approximate projector to the invariant subspace corresponds to the interval.

In this study we use filters which are polynomials (of the real part) of a linear combination of a small number (2 to 4) of resolvents whose shifts are complex numbers. To construct the filter, we initially consider an original filter which is a polynomial of a single resolvent. And to the rational function which is the transfer function of the original filter we compose a low degree rational function which is chosen well to make a new rational function whose transfer function property is improved. Then, from the improved transfer function, we can construct a linear operator which is a polynomial of (the real part of) a linear combination of a few resolvents whose shifts are complex numbers.

We select the low degree rational function for the composition, by analogies to those four typical types of filters in the theory of electronic circuits namely *Butterworth*, *Chebyshev*, *inverse-Chebyshev* and *elliptic*. We have already shown analogies of the first three types, and already shown details of constructions of corresponding filters and also shown that approximate pairs of real symmetric-definite generalized eigenproblems can be solved by the use of them. However, we have not yet shown the analogy of the last one, the elliptic type. Thus in this paper, for the analogy of the elliptic type, we show details of the construction of the rational function used for the function composition and the construction of the corresponding filter. And we also show approximate pairs of real symmetric-definite generalized eigenproblems can be solved well by the use of filters constructed by the function composition method of elliptic type.

**Keywords:** filter diagonalization, eigenproblem, resolvent, polynomial, elliptic

## 1. はじめに

行列  $A$  と  $B$  が実対称で  $B$  は正定値である一般固有値問題  $A\mathbf{v} = \lambda B\mathbf{v}$  の固有対  $(\lambda, \mathbf{v})$  で固有値  $\lambda$  が指定された区間  $[a, b]$  にあるものをフィルタ対角化法 [5] を用いて近似して求める。フィルタとして用いる線形作用素  $\mathcal{F}$  をうまく構成して、固有値が区間  $[a, b]$  近傍にある固有ベクトルは良く通過させるが固有値が区間から離れている固有ベクトルは強く減衰させるようにする。

### 1.1 レゾルベントの線形結合型のフィルタ

フィルタの構成法として複素関数論の Cauchy の積分定理から導かれる複素平面内の閉曲線に沿った線積分の離散化近似に相当する方法がある ([6][7] [8][9][11][13])。これはフィルタの作用を曲線上の積分の離散化点に対応するシフト  $\rho_j$  を持つ複数のレゾルベント  $\mathcal{R}(\rho_j) \equiv (A - \rho_j B)^{-1}B$  の線形結合を用いて実現している。

別のアプローチとして、固有値  $\lambda$  の固有ベクトル  $\mathbf{v}$  に対する伝達関数  $f(\lambda)$  の特性が、なるべく望ましいものとなるように構成された線形作用素  $\mathcal{F}$  をフィルタとして用いる方法がある。いまフィルタの構成をレゾルベント (と恒等作用素) の線形結合  $\mathcal{F} = c_\infty I + \sum_j c_j \mathcal{R}(\rho_j)$  (あるいはその実部) に限れば、その伝達関数  $f(\lambda)$  は有理関数で無限遠では有界で極に重複の無い有理関数になる。そこでそのような性質を持つ有理関数であって理想的なフィルタの伝達特性を近似するものを定めてやり ([2][3]), それから伝達関数  $f(\lambda)$  を持つフィルタ  $\mathcal{F}$  を逆に構成すれば、各レゾルベントのシフト  $\rho_j$  と線形結合の係数  $c_j$  が決まる。こうして固有値が区間  $[a, b]$  にある固有ベクトルを抽出するのに良い特性を持ったフィルタが設計できる ([16][17][20][21][12])。たとえば構成を「楕円フィルタ」にすると、複素数をシフトとするレゾルベントの 8 個から 16 個程度の線形結合の実部として十分に特性の良いフィルタが得られる。しかも求めたい固有対の固有値を含む区間は、固有値分布の任意の位置に設定ができる (中間固有対用)。

また、フィルタをシフトがすべて実数のレゾルベントの線形結合により構成することで、計算に複素数を用いずに実数だけを用いる方法も考えられる ([18][19][25])。しかしそのようなフィルタは、求めようとする固有対の固有値が固有値分布の端にある場合にだけ安全に適用できるものであり、また得られるフィルタの伝達特性もシフトに虚数を用いる場合に比べて良くない。これらの方法はシフトとして実数だけを使うかそうでないかの違いはあるものの、どちらの場合もフィルタはレゾルベントの線形結合なので複数のレゾルベントを用いることになる。

### 1.2 レゾルベントと連立 1 次方程式

与えられたベクトルの組  $V$  に対するレゾルベント  $\mathcal{R}(\rho)$

の作用は、係数行列が  $C(\rho) = A - \rho B$  で右辺ベクトルの組が  $BV$  の連立 1 次方程式を解くことで実現する (行列  $C(\rho)$  はシフト  $\rho$  が実数ならば実対称で、虚数ならば複素対称である。行列  $C(\rho)$  の非零要素の分布は行列  $A$  と  $B$  の非零要素の分布の合併になるので、 $A$  と  $B$  が疎であれば  $C(\rho)$  も疎行列になり、 $A$  と  $B$  が帯行列であれば  $C(\rho)$  も帯行列になる)。いま係数  $C(\rho)$  の連立 1 次方程式を直接法で (行列分解を利用して) 解くことにすると、フィルタの作用を実現するには少なくともフィルタの構成に必要なレゾルベントと同数の行列分解が必要である (連立 1 次方程式を反復法を用いて解く場合にも、行列の不完全分解を前処理に用いるならば、フィルタを構成するレゾルベントと同数の行列の不完全分解が必要である)。

シフト  $\rho$  が実数で、それが固有値分布の下限あるいは上限であれば、行列  $C(\rho)$  は実対称定値なので、行列  $C(\rho)$  を係数とする連立 1 次方程式はピボット交換なしの改訂コレスキ法 (modified Cholesky method) を用いて安定に解ける。さらに  $C(\rho)$  が帯行列であれば帯用に特化した算法を用いることができる。

シフト  $\rho$  が虚数のときは、シフト行列  $C(\rho)$  は正則な複素対称行列 ( $C(\rho)^T = C(\rho)$ ) になるので、 $C(\rho)$  を係数とする連立 1 次方程式の解法として複素対称版の改訂コレスキ法を用いる。その際、もしもピボット交換を行わない分解計算では数値的な不安定性が生じたならば、行列  $C(\rho)$  の複素対称性の利用をやめて、解法として片側ピボット交換付きの複素版の  $LU$  分解を用いることができる (ただし必要な記憶量と計算量がかなり増えてしまう)。さらに  $C(\rho)$  が帯行列である場合には、帯用に特殊化した改訂コレスキ法や片側ピボット交換付きの  $LU$  分解を利用できる。

### 1.3 単一のレゾルベントの多項式型フィルタ

必要な行列分解の数を最少にするために、単一のレゾルベントを用いて、レゾルベントのシフトが実数の場合にはその作用の多項式で、シフトが複素数の場合にはその虚部の作用の多項式で、フィルタを構成する方法を試みてきた ([24][22][23][26] [27][28][29][30] [31][14][32])。レゾルベントの作用を与える連立 1 次方程式を直接法で解くことを前提にするとき、単一のレゾルベントを用いる方法は複数用いる方法に比べて、行列分解を行なう数が複数ではなく 1 つになるので、行列分解に掛かる演算量を減らせることが利点である。

いまフィルタ  $\mathcal{F}$  を単一のレゾルベント  $\mathcal{R}(\rho)$  から作られる実作用素の実多項式とすれば、その伝達関数  $f(\lambda)$  は実有理関数になる。そこで、必要な固有値を持つ固有ベクトルの抽出に適したフィルタの伝達関数を表すように実有理関数  $f(\lambda)$  をうまく調整する。そのようにして得られた実有理関数を伝達関数とするフィルタ  $\mathcal{F}$  を構成する。

本論文で扱う実対称定値一般固有値問題ではすべての固有値は実数なので、求めたい固有値の区間  $[a, b]$  が固有値分布の下端 (上端) にある場合には、レゾルベントのシフトを実数に制限しても、シフトの値を最小固有値の下端 (あるいは最大固有値の上端) に設定すれば、シフトと固有

<sup>1</sup> 首都大学東京・数理科学専攻  
Department of Mathematical Sciences, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan  
a) mrkmhrsh@tmu.ac.jp

値の一致や極端な近接はない。シフトに実数を用いる場合には、フィルタ対角化法の計算は実数の数値と演算だけを用いて実行できる。

求めたい固有値の区間が固有値分布に対して一般的な位置にある場合には、シフトに虚数を選ぶことで固有値との一致や極端な近接を避けることができる。実ベクトルに対してシフトが互いに複素共役対であるレゾルベントを作用させて得られる2つの複素ベクトルが互いの複素共役になることを用いると、実作用素であるフィルタはシフトが互いに複素共役であるレゾルベントの片方（たとえば虚部が正であるもの）だけを用いて構成できる。この場合にも必要なレゾルベントは1つだけとなり、レゾルベントに対応する連立1次方程式を直接法で解く場合には、係数行列の分解はただ1つになる。

しかし単一のレゾルベントの多項式として構成されたフィルタには、（多項式がチェビシェフ多項式である簡易型の場合は特に）難点がある。それは通過域における伝達関数の最大最小比を小さくするのが難しいことである（シフトが実数の場合の方が虚数の場合よりも難しい）。この最大最小比が大きいと（たとえば3~6桁）、近似固有ベクトルの精度の不均一さも同程度になる可能性がある。遷移域を広くすればこの最大最小比を小さくできるが、遷移域の広いフィルタは望ましいものではない。なぜならば遷移域を広くすれば、それに含まれる不要な固有値の数は増える傾向がある。固有値が遷移域にある不要な固有値が増えると、それに対応してフィルタを適用して得られた結果のベクトルに含まれる不要な固有ベクトルの数も増えるので、フィルタを適用するベクトルの数を増すことが必要になり、作業の効率が悪くなるからである。

#### 1.4 関数合成による特性の改良

本報告では、単一のレゾルベントの多項式であるフィルタを元にして、その伝達関数に対してうまく選んだ有理関数を合成することにより、遷移域の幅が縮小された（あるいは通過域における最大最小比の小さい）特性の改良された伝達関数となる有理関数が得られること、そうしてその合成で得られた有理関数を伝達関数とするフィルタを少数のレゾルベントの線形結合の多項式として実現できることを示す。

しかしレゾルベントの数は元の1つから少数（2~3個）に増えるため、レゾルベントの作用を与える連立1次方程式を直接法を用いて解く場合には、行列分解に掛かる演算量と行列分解を保持するための記憶量はどちらも元の2~3倍に増える。この不利益はフィルタの伝達特性の改良から得られる利益との交換になる。

## 2. 単一のレゾルベントの多項式によるフィルタ

実対称定値一般固有値問題  $A\mathbf{v} = \lambda B\mathbf{v}$  の固有対  $(\lambda, \mathbf{v})$  の固有値  $\lambda$  はどれも実数である。固有ベクトル  $\mathbf{v}$  もまたすべて実ベクトルに選べるのでそのように選ぶことにする。

さらに固有ベクトルの全体は全空間を張る  $B$ -正規直交系にとれる。この固有値問題に対するシフト  $\rho$  のレゾルベント  $\mathcal{R}$  を式 (1) で定義する。

$$\mathcal{R}(\rho) \equiv (A - \rho B)^{-1} B. \quad (1)$$

そのとき固有対  $(\lambda, \mathbf{v})$  に対して式 (2) が成り立つ。

$$\mathcal{R}(\rho)\mathbf{v} = \frac{1}{\lambda - \rho} \mathbf{v}. \quad (2)$$

すると  $Q(x)$  が  $x$  の多項式であれば式 (3) が成り立つ。

$$Q(\mathcal{R}(\rho))\mathbf{v} = Q\left(\frac{1}{\lambda - \rho}\right)\mathbf{v}. \quad (3)$$

いまフィルタ  $\mathcal{F}$  がレゾルベント  $\mathcal{R}(\rho)$  の多項式として式 (4) の形に書けたとする。

$$\mathcal{F} = Q(\mathcal{R}(\rho)). \quad (4)$$

そのとき固有対  $(\lambda, \mathbf{v})$  の固有ベクトル  $\mathbf{v}$  に対するフィルタ  $\mathcal{F}$  の作用は式 (5) で与えられるので、 $f(\lambda)$  はフィルタ  $\mathcal{F}$  の伝達関数である。

$$\mathcal{F}\mathbf{v} = f(\lambda)\mathbf{v}, \quad f(\lambda) = Q\left(\frac{1}{\lambda - \rho}\right). \quad (5)$$

いま座標  $\lambda$  に対する正規化座標  $t$  を、固有値の指定区間  $\lambda \in [a, b]$  から標準区間  $t \in [0, 1]$  への1次変換  $t = (\lambda - a)/(b - a)$  により定義する。そうして非負領域  $t \in [0, \infty)$  を定義域として式 (6) で表された2つのパラメタ  $\mu > 1$  と  $\sigma > 0$  を持つ有理関数  $x(t)$  を導入する。

$$x(t) \equiv \frac{\mu + \sigma}{t + \sigma}. \quad (6)$$

いまフィルタ  $\mathcal{F}$  の伝達関数  $f(\lambda)$  を正規化座標  $t$  を用いて表した有理関数  $g(t) = f(\lambda)$  が  $n$  次多項式  $P$  を用いて  $g(t) = P(x(t))$  と表せるならば、 $\lambda$  で表わした伝達関数は  $f(\lambda) = P(\gamma/(\lambda - \rho))$  であり、フィルタは  $\mathcal{F} = P(\gamma \mathcal{R}(\rho))$  となる。ただし  $\gamma \equiv (b - a)(\mu + \sigma)$ 、 $\rho \equiv a - (b - a)\sigma$  とおいた。

この場合のレゾルベントのシフト  $\rho$  は実数で  $a$  未満になる。シフト  $\rho$  がどの固有値にも一致や極端な近接をしないことを保証するため、区間の下端  $a$  は最小固有値  $\lambda_{\min}$  以下であるとする。つまり  $a$  は固有値分布の1つの下界であり、区間  $[a, b]$  が固有値分布の下端を含むように設定されていると仮定する。そうして  $x(t)$  が持つ2つのパラメタ  $\sigma$  と  $\mu$ 、および  $n$  次多項式  $P$  をうまく調整することで伝達関数  $g(t) = P(x(t))$  の特性をなるべく良いものにする。つまり  $g(t)$  に対応して構成されたフィルタ  $\mathcal{F}$  は固有値  $\lambda$  が区間  $[a, b]$  にある固有ベクトルは良く伝達するが、固有値が区間から離れた固有ベクトルは強く阻止するようにする。

## 3. 伝達関数 $g(t)$ と有理関数 $h(t)$ の合成

いま正規化座標  $t$  で表したフィルタ  $\mathcal{F}$  の伝達関数  $g(t)$  が半無限区間  $[0, \infty)$  で定義された有界な実有理関数であると



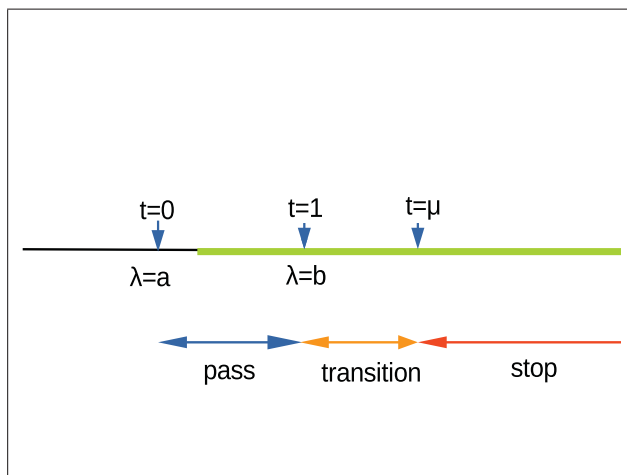


図 1 固有値  $\lambda$  の区間  $[a, b]$  と正規化座標  $t$  の関係と通過域  $t \in [0, 1]$ , 遷移域  $t \in (1, \mu)$ , 阻止域  $t \in [\mu, \infty)$

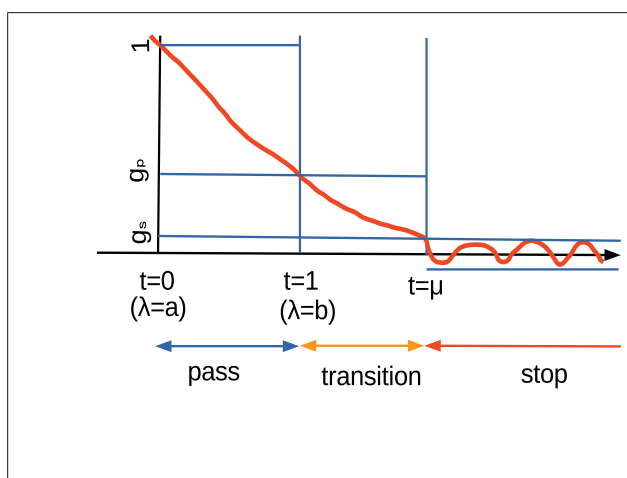


図 2 伝達関数  $g(t)$  の概形

し、その定義域を通過域 (passband), 遷移域 (transition-band), 阻止域 (stopband) に 3 分割してそれぞれ  $[0, 1]$ ,  $(1, \mu)$ ,  $[\mu, \infty)$  (ただし  $\mu > 1$ ) とする (図 1). そうして通過域における  $g(t)$  の最小値を  $g_p$  とし、通過域における  $g(t)$  の最大値を 1 に規格化し、阻止域における  $|g(t)|$  の上限値を  $g_s$  とする (図 2). また  $g_p$  と  $g_s$  を伝達率の閾値と呼ぶことにする. 通過域における伝達関数の最大最小比は  $1/g_p$  である.

### 3.1 有理関数 $h(t)$ の合成

合成を行うために導入する  $l$  次 (ただし  $l \geq 2$ ) の実有理関数  $h(t)$  は ( $t = \infty$  も含めた) 区間  $t \in [0, \infty)$  において以下の 3 つの性質を満たすものであるとする. ただし  $1 < \xi$  であるとして、関数値としては  $\pm\infty$  も含めて考える.

- (1)  $[0, 1]$  から  $[0, 1]$  への全射.
- (2)  $(1, \xi)$  から  $(1, \mu)$  への全射で, 単調増加.
- (3)  $[\xi, \infty)$  から  $[\mu, \infty)$  への全射.

このとき  $g(t)$  と  $h(t)$  の合成関数  $\hat{g}(t) = g(h(t))$  を定義すると,  $\hat{g}(t)$  も実有理関数であり,  $\hat{g}(t)$  の定義域である  $[0, \infty)$  の通過域, 遷移域, 阻止域への分割をそれぞれ  $[0, 1]$ ,  $(1, \xi)$ ,  $[\xi, \infty)$  とすれば,  $g(t)$  と  $\hat{g}(t)$  の伝達率の閾値はそれぞれ一

致する, つまり通過域における伝達率の最小値はどちらも  $g_p$  であり, 阻止域における伝達率の大きさの上限値もどちらも  $g_s$  である. ただし遷移域の幅を与えるパラメタはそれぞれ  $\mu$  と  $\xi$  であり異なる. つまり伝達関数の形状パラメタの 3 つ組は, 元の伝達関数  $g(t)$  では  $(g_p, g_s, \mu)$  であり, 合成により得られた伝達関数  $\hat{g}(t) = g(h(t))$  では  $(g_p, g_s, \xi)$  である. 関係  $h(\xi) = \mu$  により決まる  $\xi$  が  $\mu$  よりも小さければ, 合成後の伝達関数  $\hat{g}(t)$  の遷移域の幅  $\xi - 1$  は元の伝達関数  $g(t)$  の遷移域の幅  $\mu - 1$  よりも狭くなる.

もしも  $h(t)$  が偶関数であれば, 合成された  $\hat{g}(t)$  も偶関数になるので,  $\hat{g}(t)$  の定義域は自然に実数全体に拡げることができて,  $\hat{g}(t)$  の通過域, 遷移域, 阻止域も原点  $t = 0$  に対称に拡大できて, それぞれ通過域  $[-1, 1]$ , 遷移域  $1 < |t| < \xi$ , 阻止域  $\xi \leq |t|$  となる. またさらに  $h(t)$  の  $[-1, 0]$  における値域が  $[0, 1]$  に含まれる場合には,  $\hat{g}(t)$  の通過域を  $[-1, 1]$  に拡げることができる.

#### 3.1.1 4 種類の合成用の有理関数 $h(t)$

本報告ではアナログ電気回路理論の典型フィルタの 4 種類としてよく知られているバターワース (Butterworth) 型, チェビシェフ (Chebyshev) 型, 逆チェビシェフ (inverse-Chebyshev) 型, 楕円 (elliptic) 型の各構成法 (文献 [2][3]) の模倣から得られる合成 (B-合成, C-合成, I-合成, E-合成) を利用してフィルタを構成する. それらの合成に用いる  $l$  次 (ただし  $l \geq 2$  とする) の実有理関数  $h(t)$  は, それぞれ式 (7) で与えられるものである.

$$h(t) = \begin{cases} t^l & \text{(B-合成)}, \\ \frac{1 + T_l(t)}{2} & \text{(C-合成)}, \\ \frac{1 + T_l(\xi)}{1 + T_l(\xi/t)} & \text{(I-合成)}, \\ \frac{L+1}{2} \times \frac{1 + R_l(\xi, t)}{L + R_l(\xi, t)} & \text{(E-合成)}. \end{cases} \quad (7)$$

ここで C-合成と I-合成に現れる  $T_l(z)$  は次数  $l$  の第 1 種チェビシェフ多項式を表す. また I-合成と E-合成に現れる実数パラメタ  $\xi$  の値は 1 より大きい. E-合成の場合に現れる  $R_l(\xi, z)$  はパラメタ  $\xi$  を持ち変数が  $z$  である次数  $l$  の「楕円有理関数」を表し,  $L$  は実数で  $L = R_l(\xi, \xi)$  を満たし 1 より大きい (E-合成についての説明は A.1.1 節に記述).

#### 3.1.2 4 種類の合成について

4 つの合成用の有理関数  $h(t)$  は次数が  $l = 2$  の場合にはどれも多項式  $t^2$  となり一致する. また有理関数  $h(t)$  は B-合成と C-合成の場合は常に多項式であるが, I-合成と E-合成は ( $l \geq 3$  では) 多項式ではない.

B-合成と I-合成はどちらも, 原点付近で  $h(t) = O(t^l)$  なので, 合成された伝達関数  $\hat{g}(t) = g(h(t))$  の原点における  $(l-1)$  階までの導関数はすべて零になり, 原点付近で高度な平坦性 (いわゆるバターワース特性) を持つ.

上記の 4 通りの合成 (B-合成, C-合成, I-合成, E-合成) に対する有理関数  $h(t)$  はどれも次数  $l$  が偶数であれば偶関数になる. 次数  $l$  が奇数の場合には, 式 (7) で与えた  $h(t)$  の C-合成と E-合成については, 有理関数  $h(t)$  の  $t \in [-1, 0]$  と  $t \in [0, 1]$  における値域は一致することが示せるが, B-合

成とI-合成については区間  $t \in [-1, 0]$  における  $h(t)$  の値域は1よりも大きい値を含むので区間  $[0, 1]$  には収まらないことが示せる(そのため  $l$  が奇数の場合のB-合成やI-合成において、通過域を  $t \in [-1, 1]$  に広げるためには、拡張された通過域  $t \in [-1, 1]$  において  $h(t)$  の値が  $[0, 1]$  に収まるように  $h(t)$  の定義式を修正することが必要になる)。

以上のことから、 $l$  が偶数である場合には4通りの合成(B-合成、C-合成、I-合成、E-合成)のいずれの場合も伝達関数  $\hat{g}(t)$  の通過域、遷移域、阻止域をそれぞれ原点对称に自然に広げて  $t \in [-1, 1]$ ,  $1 < |t| < \xi$ ,  $\xi \leq |t|$  にできることがわかる。他方で、 $l$  が奇数である場合には、B-合成とI-合成については  $t$  が原点から負の側に向かうときに合成された伝達関数  $\hat{g}(t)$  は値が1を越えてしまうので通過域を ( $h(t)$  の式を変更せずに) 自然に広げる (たとえば  $t \in [-1, 1]$  にする) ことはできない。そこでたとえば通過域、遷移域、阻止域の各区域をそれぞれ元のままの  $[0, 1]$ ,  $(1, \xi)$ ,  $[\xi, \infty]$  とする。他方でC-合成とE-合成については  $l$  の偶奇に依らずに通過域を  $t \in [-1, 1]$  に広げることが自然にできるので、通過域、遷移域、阻止域の各区域はそれぞれ  $t$  について  $[-1, 1]$ ,  $(1, \xi)$ ,  $[\xi, \infty]$  にできることがわかる(C-合成とE-合成では  $l$  が奇数の場合には合成された伝達関数  $\hat{g}(t)$  は値が  $-1$  未満の実数の極を1つだけ持つ)。

合成用の次数  $l$  が偶数である場合には、4通りの合成で得られるフィルタの伝達関数はどれも実数の極を持たないので、固有値を求める区間  $[a, b]$  の位置は任意に設定できる。そうして4通りの合成(B-合成、C-合成、I-合成、E-合成)のどの場合にも固有値の区間  $\lambda \in [a, b]$  と拡張された通過域  $t \in [-1, 1]$  とを対応させることができる(ただし、B-合成とI-合成では、そのように対応させると、伝達関数の持つ原点  $t = 0$  付近における高度な平坦性(バターワース特性)は、区間の中央  $(a + b)/2$  付近の固有値に対するものになる。また  $l$  が偶数である場合には、B-合成とI-合成の伝達関数は原点  $t = 0$  で最大で  $t \in [0, 1]$  では単調減少である。そこで通過域を元の  $t \in [0, 1]$  に制限すると、固有値が区間  $[a, b]$  の下端付近の固有対がより良く近似される傾向を持つ。そこで、もしも区間の下端付近に固有値を持つ固有対の精度を重視する場合には、 $l$  が偶数の場合にも、B-合成やI-合成の通過域として  $t \in [-1, 1]$  ではなくて、あえて  $t \in [0, 1]$  を使うこともできる。すなわち用途によっては、固有値の区間  $\lambda \in [a, b]$  に対して  $t \in [0, 1]$  を対応させることがありえる。)

他方で  $l$  が奇数の場合には、4種類のフィルタの伝達関数は通常は通過域の下端よりも下側に実数の極を1つ持つので、そのようなフィルタを安全に適用できるのは固有値分布の下端付近の固有値を持つ固有対を求める場合だけである。いま  $a$  が最小固有値  $\lambda_{\min}$  以下の実数であるとする。C-合成とE-合成では  $l$  が奇数の場合であっても広げた通過域  $t \in [-1, 1]$  を固有値の区間  $\lambda \in [a, b]$  に対応させることができる。他方でB-合成とI-合成では、 $l$  が奇数の場合に、 $\lambda \in [a, b]$  に  $t \in [-1, 1]$  を対応させると ( $h(t)$  の定義式を修正しなければ)  $t \in [-1, 0]$  においてはフィルタの伝達関数  $\hat{g}(t)$  の値は1を越える(たとえば実験例5のフィルタ

の伝達関数のグラフ図 A-42 を参照)。そこで保守的な立場では、固有値の区間  $\lambda \in [a, b]$  に対応させる通過域を区間  $t \in [0, 1]$  に制限する。その場合には、B-合成とI-合成は区間の下端付近でバターワース特性を持ち、またフィルタの伝達関数は区間  $\lambda \in [a, b]$  において単調減少になる。

有理関数の合成により伝達関数  $\hat{g}(t) = g(h(t))$  が得られたら、次はそれに対応するフィルタ  $\hat{F}$  をレゾルベントの線形結合(の実部)の多項式として構成する。本報告では、 $n$  次多項式がチェビシェフ多項式  $T_n$  を用いて表される簡易型のフィルタだけを扱うことにする。

### 3.1.3 各合成における $\mu$ と $\xi$ の対応

4種類の合成(B-合成、C-合成、I-合成、E-合成)の各場合に、 $\mu$  は  $\xi$  を用いた式(8)により表せる(C-合成とI-合成の場合の式は一致する)。

$$\mu = \begin{cases} \xi^\ell & (\text{B-合成}), \\ \frac{1 + T_\ell(\xi)}{2} & (\text{C-合成とI-合成}), \\ \frac{(L+1)^2}{4L} & (\text{E-合成}). \end{cases} \quad (8)$$

ただし式(8)でE-合成の場合は  $L \equiv 1/q^{-1}(\{q(1/\xi)\}^\ell)$  である。ここで  $q(k)$  は楕円 nome 関数  $q(k) \equiv \exp\{-\pi K'(k)/K(k)\}$  を表し、 $q^{-1}(x)$  はその逆関数を表す。また  $K(k)$  は母数  $k$  の第1種の完全楕円積分であり、 $K'(k) \equiv K(k')$  であり、 $k' \equiv \sqrt{1-k^2}$  は母数  $k$  の補母数を表す(これらの関数や楕円積分についてはA.1.1節に記述)。

あるいは  $\xi$  を  $\mu$  を用いて表すと式(9)になる(これもC-合成とI-合成に対する式は一致する)。

$$\xi = \begin{cases} \mu^{1/\ell} & (\text{B-合成}), \\ \cosh\{(2/\ell)\sinh^{-1}\sqrt{\mu-1}\} & (\text{C-合成とI-合成}), \\ 1/q^{-1}(\{q(1/L)\}^{1/\ell}) & (\text{E-合成}). \end{cases} \quad (9)$$

ただし式(9)でE-合成の場合は  $L \equiv (\sqrt{\mu} + \sqrt{\mu-1})^2$  である。

## 3.2 簡易型の伝達関数と $h(t)$ の合成

伝達関数  $g(t)$  がレゾルベントの線形結合(の実部)の多項式で、その多項式はチェビシェフ多項式  $T_n$  を用いて表せるとき、そのような伝達関数をここでは簡易型と呼ぶ。

合成を行う前の簡易型の伝達関数を式(10)とする。この式が含むパラメタの組は  $(n, \mu, \sigma)$  であり、本報告ではそれを標準パラメタの3つ組と呼ぶことにする。

$$g(t) = g_s T_n(2x(t) - 1), \quad x(t) = \frac{\mu + \sigma}{t + \sigma}. \quad (10)$$

有理関数  $h(t)$  を合成した後の伝達関数  $\hat{g}(t) = g(h(t))$  は式(11)で与えられる。

$$\hat{g}(t) = g_s T_n(2\hat{x}(t) - 1), \quad \hat{x}(t) = \frac{\mu + \sigma}{h(t) + \sigma}. \quad (11)$$

標準パラメタの3つ組  $(n, \mu, \sigma)$  から閾値  $g_s$  と  $g_p$  は式(12)

を順に計算すれば求まる。

$$\begin{cases} g_s \leftarrow 1 / \cosh \left( 2n \sinh^{-1} \sqrt{\frac{\mu}{\sigma}} \right), \\ g_p \leftarrow g_s \cosh \left( 2n \sinh^{-1} \sqrt{\frac{\mu-1}{\sigma+1}} \right). \end{cases} \quad (12)$$

関数  $h(t)$  を合成することで、閾値  $g_s$  と  $g_p$  は変えずに、関係式 (9) により遷移域の幅を  $\mu-1$  から  $\xi-1$  に縮小できる。

#### 4. 合成された伝達関数を持つフィルタの構成

$P$  が多項式であるときに、合成で得られた関数  $\hat{g}(t) = P(\hat{x}(t))$  を伝達関数として持つフィルタ  $\hat{F}$  は以下の手順により、構成できる

重複する極を持たない有理関数  $\hat{x}(t)$  の表式 (13) から、式 (14) の形の (重複極を持たない場合の) 部分分数分解を作る (4.1 節参照)。

$$\hat{x}(t) \equiv x(h(t)) = \frac{\mu + \sigma}{h(t) + \sigma} \quad (13)$$

$$\hat{x}(t) = c_\infty + \sum_{j=1}^{\ell} \frac{c_j}{t - t_j}. \quad (14)$$

すると  $\hat{x}(t)$  に対応する作用素  $\hat{X}$  は、式 (15) で表された恒等作用素  $I$  と  $\ell$  個のレゾルベントの線形結合になる。ここでシフト  $\rho_j$  に対応する規格化座標が  $t_j$  である。

$$\hat{X} = c_\infty I + \sum_{j=1}^{\ell} \gamma_j \mathcal{R}(\rho_j). \quad (15)$$

さらにいま扱っている実対称定値一般固有値問題では、 $\hat{X}$  は実線形作用素、つまり実ベクトルから実ベクトルへの作用素であることから、複素共役対称性を利用することで、複素ベクトルの実部を取り出す処理と組み合わせると、たとえば式 (15) はシフトの虚部の符号が非負であるレゾルベントの項だけが現れる形に書き直せる。それにより、4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) の場合には、(たとえば) 虚部が正である虚数をシフトとするレゾルベントの項を  $[\ell/2]$  個、実数をシフトとするレゾルベントの項を  $\ell$  が奇数のときにだけ 1 個、それぞれ用いて表せる (4.2 節参照)。

伝達関数が  $\hat{g}(t) = P(\hat{x}(t))$  である場合のフィルタは  $\hat{F} = P(\hat{X})$  となる。特に伝達関数が簡易型で  $\hat{g}(t) = g_s T_n(2\hat{x}(t)-1)$  である場合のフィルタは  $\hat{F} = g_s T_n(2\hat{X}-I)$  となる (4.3 節参照)。

##### 4.1 各合成に対する $\hat{x}(t)$ の部分分数分解の構成

4 種類の合成 (B-合成, C-合成, I-合成, E-合成) のそれぞれで用いる関数  $h(t)$  は次数  $\ell$  の実有理関数 (B-合成と C-合成の場合は特に多項式) である。それぞれの場合に、フィルタをレゾルベントを用いて構成するためには式 (13) で表わされる実有理関数  $\hat{x}(t)$  から式 (14) の形の部分分数分解を具体的に構成する必要がある (これら 4 種類

の合成の場合には  $\hat{x}(t)$  の極には重複がなく全部で  $\ell$  個ある)。ここで定数項の係数  $c_\infty$  は実数であり、極  $t_j$  と係数  $c_j$ ,  $j=1, 2, \dots, \ell$  は複素数である。

部分分数分解の定数項  $c_\infty$  は  $c_\infty = \lim_{t \rightarrow \infty} \hat{x}(t)$  であり、極  $t_j$  は方程式  $h(t) + \sigma = 0$  を解いて得られ、極  $t_j$  の係数  $c_j$  は式  $c_j = \lim_{t \rightarrow t_j} \{(t - t_j)\hat{x}(t)\}$  を計算すれば求められる。

##### 4.1.1 $\hat{x}(t)$ の部分分数分解 (B-合成の場合)

ここでは、B-合成の場合の部分分数分解 (14) を構成する。定数項の係数  $c_\infty$  は B-合成の場合の  $h(t)$  が多項式であることから常に零になる。

複素数の範囲での  $\ell$  個の極  $t = t_j$  は、方程式  $t^\ell + \sigma = 0$  を満たし、式 (16) から計算できて重複するものは無い。

$$t_j = \sigma^{1/\ell} \omega^{2j-1}, j=1, 2, \dots, \ell, \quad \omega \equiv \exp(\pi\sqrt{-1}/\ell). \quad (16)$$

それらのうちで虚部が正の極の添字は前半のもの  $j=1, 2, \dots, [\ell/2]$  である。実数の極  $t_R$  は  $\ell$  が奇数のときに限り唯一存在して、添字は  $j = (\ell + 1)/2$  で値は負で  $t_R \equiv t_{(\ell+1)/2} = -\sigma^{1/\ell}$  である (その存在範囲は  $\sigma$  が 1 と比べて大きい、等しい、小さいに従ってそれぞれ  $t_R < -1$ ,  $t_R = -1$ ,  $-1 < t_R$  となる)。

極の係数  $c_j$ ,  $j=1, 2, \dots, \ell$  は式 (17) で与えられる。

$$c_j = -\frac{\mu + \sigma}{\sigma \ell} \times t_j. \quad (17)$$

そうして  $\ell$  が奇数のときに限り唯一存在する実数の極  $t_R \equiv t_{(\ell+1)/2}$  の係数  $c_R \equiv c_{(\ell+1)/2}$  は実数である。

##### 4.1.2 $\hat{x}(t)$ の部分分数分解 (C-合成の場合)

ここでは、C-合成の場合の部分分数分解 (14) を構成する。部分分数分解の係数  $c_\infty$  は C-合成の場合の  $h(t)$  が多項式であることから常に零になる。

次に  $\hat{x}(t)$  の極を求めるためには、その分母が零になる条件  $h(t) + \sigma = 0$ , つまり方程式  $T_\ell(t) = -(1 + 2\sigma)$  を解けばよい。そこでまず 3 つの正の実数  $\alpha$ ,  $d_R$ ,  $d_I$  をそれぞれ式 (18) を用いて順に計算する。式の形から  $d_R > 1$  であることがわかる。

$$\begin{cases} \alpha \leftarrow (\sqrt{\sigma} + \sqrt{1 + \sigma})^{2/\ell}, \\ d_R \leftarrow (\alpha + 1/\alpha)/2, \\ d_I \leftarrow (\alpha - 1/\alpha)/2. \end{cases} \quad (18)$$

あるいは  $\alpha$  を作る計算の中で  $2/\ell$  冪乗を求めるために対数の計算と指数の計算を用いるのであれば、むしろ  $\log \alpha \leftarrow (2/\ell) \log(\sqrt{\sigma} + \sqrt{1 + \sigma})$ ,  $d_R \leftarrow \cosh(\log \alpha)$ ,  $d_I \leftarrow \sinh(\log \alpha)$  として計算する方がよいであろう。

方程式のすべての複素数の解  $t_j$  は式 (19) で表わされて、それらに重複が無いこともわかる。

$$t_j = d_R \cos \frac{(2j-1)\pi}{\ell} + \sqrt{-1} \cdot d_I \sin \frac{(2j-1)\pi}{\ell}, \quad j = 1, 2, \dots, \ell. \quad (19)$$

なお、虚部が正である極の添字は前半のもの  $j=1, 2, \dots, [\ell/2]$  である。実数の極は  $\ell$  が奇数のときに



限り唯一存在し、添字は  $j = (\ell+1)/2$  であり、値は  $t_R \equiv t_{(\ell+1)/2} = -d_R$  であって  $-1$  よりも小さい。

そうして極  $t = t_j$  の係数  $c_j$  は式 (20) により計算できる。

$$c_j = \lim_{t \rightarrow t_j} \{(t - t_j) \hat{x}(t)\} = \frac{\mu + \sigma}{\frac{d}{dt} h(t)|_{t=t_j}}. \quad (20)$$

この式で  $(d/dt)h(t)|_{t=t_j}$  の値は、公式  $(d/dz)T_m(z) = mU_{m-1}(z)$  を用いて求められる。ただし  $U_m(z)$  は  $m$  次の第2種チェビシェフ多項式を表し、その値は漸化式 (21) を用いて計算できる。

$$\begin{cases} U_0(z) = 1, U_1(z) = 2z, \\ U_m(z) = 2zU_{m-1}(z) - U_{m-2}(z) \quad (m \geq 2). \end{cases} \quad (21)$$

すると  $(d/dt)h(t)|_{t=t_j}$  の値は  $(\ell/2)U_{\ell-1}(t_j)$  となるので、そのことから極  $t_j$  の係数  $c_j$  は式 (22) により計算できる。

$$c_j = \frac{2(\mu + \sigma)}{\ell} \times \frac{1}{U_{\ell-1}(t_j)}. \quad (22)$$

#### 4.1.3 $\hat{x}(t)$ の部分分数分解 (I-合成の場合)

ここでは、I-合成の場合の部分分数分解 (14) を構成する。

まず定数項  $c_\infty$  の値は、次数  $\ell$  の 4 を法とする剰余により場合が分かれて、式 (23) により与えられる。

$$c_\infty = \begin{cases} \frac{\mu + \sigma}{2\mu + \sigma} & (\text{mod}(\ell, 4) = 1, 3 \text{ の場合}), \\ 1 & (\text{mod}(\ell, 4) = 0 \text{ の場合}), \\ 0 & (\text{mod}(\ell, 4) = 2 \text{ の場合}). \end{cases} \quad (23)$$

そうして極  $t_j$  は  $h(t) + \sigma = 0$  の解  $t$  であることから  $T_\ell(\xi/t) = -(1 + 2\mu/\sigma)$  を満たす。そこでいま  $z = \xi/t$  とおくと、 $z$  の方程式は  $T_\ell(z) = -(1 + 2\mu/\sigma)$  である。すると C-合成の場合の計算と同様に、まず 3 つの正の実数  $\alpha$ ,  $d_R$ ,  $d_I$  を式 (24) により順に計算する (式の形から  $d_R > 1$  であることがわかる)。

$$\begin{cases} \alpha \leftarrow (\sqrt{\mu/\sigma} + \sqrt{1 + \mu/\sigma})^{2/\ell}, \\ d_R \leftarrow (\alpha + 1/\alpha)/2, \\ d_I \leftarrow (\alpha - 1/\alpha)/2. \end{cases} \quad (24)$$

あるいは C-合成の場合と同様に、 $\log \alpha \leftarrow (2/\ell) \log(\sqrt{\mu/\sigma} + \sqrt{1 + \mu/\sigma})$ ,  $d_R \leftarrow \cosh(\log \alpha)$ ,  $d_I \leftarrow \sinh(\log \alpha)$  により計算することもできる。

すると (今度は  $z_j$  の式の虚部の符号を C-合成の場合とは逆に) 式 (25) により、複素数の範囲における有理関数  $\hat{x}(t)$  のすべての極  $t_j$ ,  $j=1, 2, \dots, \ell$  が求まる。なお次数が  $\ell$  のとき  $\mu$  と  $\xi$  の間には  $\xi = \cosh\{(2/\ell) \sinh^{-1} \sqrt{\mu - 1}\}$  の関係がある。

$$\begin{cases} z_j = d_R \cos \frac{(2j-1)\pi}{\ell} - \sqrt{-1} \cdot d_I \sin \frac{(2j-1)\pi}{\ell}, \\ t_j = \xi/z_j, \quad j = 1, 2, \dots, \ell \end{cases} \quad (25)$$

虚部が正である極の添字は前半のもの  $j=1, 2, \dots, \lfloor \ell/2 \rfloor$  である。実数の極は  $\ell$  が奇数の場合に限り唯一存在して、そ

の添字は  $j = (\ell+1)/2$  で、値は  $t_R \equiv t_{(\ell+1)/2} = -\xi/d_R$  となる。

$\ell$  が奇数のとき、 $t_R$  は代数方程式  $T_\ell(\xi/t) + 1 + 2\mu/\sigma = 0$  が持つ唯一の実根である。いまこの代数方程式の左辺を  $\phi(t)$  と書くとき、 $\phi(-\xi) = 2\mu/\sigma$  は正である。また  $\phi(-1) = -T_\ell(\xi) + 1 + 2\mu/\sigma$  であるが、 $\mu = \frac{1}{2}\{1 + T_\ell(\xi)\}$  であることから、 $\phi(-1) = 2\{1 - \mu(1 - 1/\sigma)\}$  となり、もしも  $\mu > 1$  や  $\sigma > 0$  がある程度大きければその値は負になり、そのとき方程式  $\phi(t) = 0$  の唯一の実根  $t_R$  は区間  $(-\xi, -1)$  に含まれるから  $-1$  よりも小さい。

また極  $t_j$  の係数  $c_j$  は、第2種チェビシェフ多項式  $U_m(z)$  を用いて式 (26) で与えられることが式計算により示せる。

$$c_j = \frac{2(\sigma + \mu)\mu}{\ell\sigma^2\xi} \times \frac{t_j^2}{U_{\ell-1}(\xi/t_j)}. \quad (26)$$

$\ell$  が奇数のときに限り唯一存在する実数の極  $t_R \equiv t_{(\ell+1)/2}$  の係数  $c_R$  は実数で、 $c_R \equiv c_{(\ell+1)/2}$  である。

#### 4.1.4 $\hat{x}(t)$ の部分分数分解 (E-合成の場合)

部分分数分解 (14) を E-合成の場合について構成する手順の説明は、それに用いる楕円関数や楕円有理関数などの記述を含んで長いので、別の場所に移した (A.1 節, A.1.4 節)。

#### 4.2 $\hat{x}(t)$ の部分分数分解に対応する作用素 $\hat{\mathcal{X}}$ の構成

ここまでで、4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) に対する次数  $\ell$  の有理関数  $h(t)$  による  $\hat{x}(t) = x(h(t))$  の式 (14) の形の部分分数分解の具体的な構成法を示した。各合成の場合に対して、虚部が正である極の添字は前半のもの  $j = 1, 2, \dots, \lfloor \ell/2 \rfloor$  であり、実数の極は  $\ell$  が奇数であるときに限り唯一存在して、その添字は  $j = (\ell+1)/2$  であり、値を  $t_R$  と表す。

複素共役対称性から、部分分数分解の式 (14) を実部を取り出す操作  $\text{Re}$  を用いて書き換えることで虚部が非負の極だけが現れる形の式 (27) にできる (複素共役な虚数の極の項の対の寄与はどちらか片方の寄与の実部として表せるから、その片方を選ぶ手段として、ここでは極の虚部の符号を用いている。)

$$\hat{x}(t) = \begin{cases} c_\infty + \sum_{j=1}^{\ell/2} \text{Re} \left( \frac{2c_j}{t - t_j} \right) & (\ell \text{ が偶数}), \\ c_\infty + \frac{c_R}{t - t_R} + \sum_{j=1}^{(\ell-1)/2} \text{Re} \left( \frac{2c_j}{t - t_j} \right) & (\ell \text{ が奇数}). \end{cases} \quad (27)$$

すると実有理関数  $\hat{x}(t)$  に対応する実線形作用素  $\hat{\mathcal{X}}$  は、 $\ell$  が偶数の場合には式 (28) と (30) の組により、 $\ell$  が奇数の場合には式 (31) と (33) の組により、それぞれ与えられる。

このように  $\hat{\mathcal{X}}$  の作用を複素ベクトルの実部をとる操作  $\text{Re}$  を用いて表すことで、フィルタの構成に必要なレゾルベントの数を  $\ell$  から減らせる。次数  $\ell$  が偶数の場合にはシフトが虚数のレゾルベント  $\ell/2$  個でよく、 $\ell$  が奇数の場合にはシフトが虚数のレゾルベント  $(\ell-1)/2$  個とシフトが

実数のレゾルベント 1 個でよい (表 1).

表 1 次数  $\ell$  に対するフィルタの構成に必要なレゾルベントの数

$\ell$	シフトが虚数のもの	シフトが実数のもの
2	1	0
3	1	1
4	2	0
5	2	1
6	3	0
7	3	1
8	4	0

#### 4.2.1 次数 $\ell$ が偶数の場合

求めたい固有対の固有値が存在する区間を  $\lambda \in [a, b]$  とする.  $\ell$  が偶数の場合はその区間の位置は任意であり, 固有値分布の内側であっても良い. 固有値の区間  $\lambda \in [a, b]$  と正規化座標で表わした通過域  $t \in [-1, 1]$  を線形変換  $\lambda = \frac{1}{2}(a+b) + \frac{1}{2}(b-a)t$  を用いて対応させるものとする. いま部分分数分解 (27) 中の虚数の極  $t_j$  とその係数  $c_j$  の項については  $c_j/(t-t_j) = \gamma_j/(\lambda-\rho_j)$  の関係がある. すると, レゾルベントの虚数のシフト  $\rho_j$  とその係数  $\gamma_j$  はそれぞれ式 (28) で与えられる.

$$\rho_j \equiv \frac{1}{2}(a+b) + \frac{1}{2}(b-a)t_j, \quad \gamma_j \equiv \frac{1}{2}(b-a)c_j. \quad (28)$$

あるいは,  $\ell$  が偶数の場合の B-合成や I-合成で, 通過域を  $[-1, 1]$  に選べる場合であっても, たとえば固有値が区間  $[a, b]$  の下端側にある固有対をなるべく精度良く求めたいという理由から, 通過域を拡げずに元のままの  $t \in [0, 1]$  にするなどの場合は, レゾルベントの虚数のシフト  $\rho_j$  とその係数  $\gamma_j$  は式 (29) により与えられる.

$$\rho_j \equiv a + (b-a)t_j, \quad \gamma_j \equiv (b-a)c_j. \quad (29)$$

レゾルベントのシフトと係数が決まれば, 次数  $\ell$  が偶数の場合の実線形作用素  $\hat{\mathcal{X}}$  は式 (30) で与えられる.

$$\hat{\mathcal{X}} = c_\infty I + \sum_{j=1}^{\ell/2} \text{Re}\{2\gamma_j \mathcal{R}(\rho_j)\}. \quad (30)$$

#### 4.2.2 次数 $\ell$ が奇数の場合

いま求めようとする固有値を含む固有値分布の下端の区間を  $\lambda \in [a, b]$  とする. そうして  $a$  は最小固有値  $\lambda_{\min}$  以下であるとする.

#### B-合成と I-合成の場合

B-合成と I-合成の場合には通過域を  $t \in [0, 1]$  にとり, それを固有値の区間  $\lambda \in [a, b]$  に線形変換  $\lambda = a + (b-a)t$  により対応させる. いま部分分数分解 (27) 中の虚数の極  $t_j$  とその係数  $c_j$  の項に対しては,  $\ell$  が偶数の場合と同様に  $c_j/(t-t_j) = \gamma_j/(\lambda-\rho_j)$  の関係がある. すると, レゾルベントの虚数のシフト  $\rho_j$  とその係数  $\gamma_j$ , および実数の極  $\rho_R$  とその実数係数  $\gamma_R$  は, それぞれ式 (31) で与えられる.

$$\begin{cases} \rho_j \equiv a + (b-a)t_j, & \gamma_j \equiv (b-a)c_j, \\ \rho_R \equiv a + (b-a)t_R, & \gamma_R \equiv (b-a)c_R. \end{cases} \quad (31)$$

#### C-合成と E-合成の場合

C-合成と E-合成の場合には,  $\ell$  が奇数の場合であっても  $h(t)$  の性質から通過域を  $t \in [-1, 1]$  に拡げることができるので, それを線形変換  $\lambda = \frac{1}{2}(a+b) + \frac{1}{2}(b-a)t$  により固有値の区間  $\lambda \in [a, b]$  に対応させることができる. その場合のレゾルベントの虚数のシフト  $\rho_j$  とその係数  $\gamma_j$ , および実数の極  $\rho_R$  とその実数係数  $\gamma_R$  はそれぞれ式 (32) で与えられる.

$$\begin{cases} \rho_j \equiv \frac{1}{2}(a+b) + \frac{1}{2}(b-a)t_j, & \gamma_j \equiv \frac{1}{2}(b-a)c_j, \\ \rho_R \equiv \frac{1}{2}(a+b) + \frac{1}{2}(b-a)t_R, & \gamma_R \equiv \frac{1}{2}(b-a)c_R. \end{cases} \quad (32)$$

レゾルベントのシフトと係数が決まれば, 次数  $\ell$  が奇数の場合の実線形作用素  $\hat{\mathcal{X}}$  は式 (33) で与えられる.

$$\hat{\mathcal{X}} = c_\infty I + \gamma_R \mathcal{R}(\rho_R) + \sum_{j=1}^{(\ell-1)/2} \text{Re}\{2\gamma_j \mathcal{R}(\rho_j)\}. \quad (33)$$

#### 4.3 合成から導かれる簡易型フィルタの作用の計算

関数合成で得られた伝達関数  $\hat{g}(t) = g_s T_n(2\hat{x}(t) - 1)$  に対応する簡易型のフィルタ  $\hat{\mathcal{F}}$  は式 (34) で与えられる.

$$\hat{\mathcal{F}} = g_s T_n(2\hat{\mathcal{X}} - I). \quad (34)$$

いま  $\hat{\mathcal{Y}} \equiv 2\hat{\mathcal{X}} - I$  とおいて, 与えられた  $m$  個の縦ベクトルの組を列として並べた  $N \times m$  行列  $V$  に対して  $V^{(k)} \equiv T_k(\hat{\mathcal{Y}})V$  と定義すると,  $V$  から開始して 3 項漸化式 (35) を用いることにより  $V^{(n)}$  が求められる.

$$\begin{cases} V^{(0)} = V, \\ V^{(1)} = \hat{\mathcal{Y}}V, \\ V^{(k)} = 2\hat{\mathcal{Y}}V^{(k-1)} - V^{(k-2)} \quad (k \geq 2 \text{ の場合}). \end{cases} \quad (35)$$

そうしてフィルタ  $\hat{\mathcal{F}}$  の  $V$  に対する作用は,  $V$  から始めて 3 項漸化式を用いて計算された  $V^{(n)}$  に対して定数  $g_s$  を乗じる形の式 (36) として与えられる.

$$\hat{\mathcal{F}}V = g_s V^{(n)} \quad (36)$$

### 5. 合成で得られる簡易型のフィルタの設計法

合成前のフィルタ  $\mathcal{F}$  の伝達関数  $g(t)$  は,  $n$  次のチェビシェフ多項式を用いる簡易型の設計では  $g(t) = g_s T_n(2x(t) - 1)$ ,  $x(t) = (\mu + \sigma)/(t + \sigma)$  で与えられる. いまこの  $g(t)$  の式に含まれるパラメタの組  $(n, \mu, \sigma)$  を「標準パラメタの 3 つ組」と呼ぶことにする (係数の  $g_s$  は通過域での  $g(t)$  の最大値が 1 という規格化条件, 今の場合は  $g(0) = 1$  から自然に決まる). 他方でフィルタの伝達特性を特徴づける 3 つの形状パラメタは  $g_p, g_s, \mu$  である.

また, この伝達関数  $g(t)$  と  $\ell$  次の有理関数  $h(t)$  の関数合成で得られる伝達関数は  $\hat{g}(t) = g(h(t))$  であり,  $g(t)$  と  $\hat{g}(t)$  の両方で伝達率の閾値である  $g_p$  と  $g_s$  は共通になるが,



阻止域の端を指定するパラメタ（あるいは遷移域の幅を決めるパラメタ）は元の伝達関数  $g(t)$  では  $\mu$  であるが、合成後の伝達関数  $\hat{g}(t)$  では  $\xi$  となる（ただし  $\mu = h(\xi)$  の関係がある）。合成されたフィルタの伝達特性を特徴付ける3つの形状パラメタは  $g_p, g_s, \xi$  であり、フィルタの設計に際してこれら3つのすべてあるいはその一部を指定したい場合がある。

そこで簡易型の伝達関数あるいはそれに合成を施したものについて、指定されたパラメタの組から伝達関数とそれに対応するフィルタを具体的に決定する方法を以下に記述する。ただし合成の種類（B-合成, C-合成, I-合成, E-合成）とその合成に用いる有理関数  $h(t)$  の次数  $l$  を先に決めてそれから設計を行うものとする。

### 5.1 標準パラメタの組 $(n, \mu, \sigma)$ を指定する場合

チェビシェフ多項式の次数  $n$  とパラメタ  $\mu (> 1)$  と  $\sigma (> 0)$  の3つ、すなわち標準パラメタの3つ組  $(n, \mu, \sigma)$  を指定する場合は、まず式 (12) を計算して  $g_s$  と  $g_p$  の値を求める。すると合成前の  $g(t)$  とその形状パラメタの3つ組  $(g_p, g_s, \mu)$  が決まる。

元の伝達関数  $g(t)$  に  $h(t)$  を合成して得られる  $\hat{g}(t) = g(h(t))$  は  $g(t)$  と閾値の値  $g_s$  と  $g_p$  を共有するが、遷移域の幅に関係する  $\xi$  の値は関係 (9) を用いて  $\mu$  から計算して求める。すると合成された  $\hat{g}(t)$  の形状パラメタの3つ組は  $(g_p, g_s, \xi)$  である。

$\mu$  と  $\sigma$  と合成の種類と合成の次数  $l$  が決まれば  $\hat{x}(t)$  が決まり、それから  $\hat{x}(t)$  の部分分数分解が決まり、それに対応してレゾルベントの線形結合の実部を用いて表された作用素  $\hat{x}$  が決まり、それとチェビシェフ多項式の次数  $n$  と  $g_s$  の値を用いて合成された伝達関数  $\hat{g}(t)$  に対応するフィルタが  $\hat{F} = g_s T_n(2\hat{x} - I)$  と決まる。

### 5.2 組 $(n, g_p, g_s)$ を指定する場合

チェビシェフ多項式の次数  $n$  と伝達関数の閾値  $g_p, g_s$  を指定する場合には、式 (12) の関係を逆に解くことで  $\mu$  と  $\sigma$  の値を求める。それには式 (37) を順に計算すればよい。

$$\begin{cases} w_1 \leftarrow \frac{1}{2n} \cosh^{-1} \frac{1}{g_s}, \\ w_2 \leftarrow \frac{1}{2n} \cosh^{-1} \frac{g_p}{g_s}, \\ \sigma \leftarrow \frac{\cosh^2 w_2}{\sinh(w_1 + w_2) \sinh(w_1 - w_2)}, \\ \mu \leftarrow \sigma \sinh^2 w_1. \end{cases} \quad (37)$$

これにより、指定されたパラメタの3つ組  $(n, g_p, g_s)$  から標準パラメタの3つ組  $(n, \mu, \sigma)$  が決まるので、フィルタ  $\hat{F}$  の構成法は前副節 (5.1) の「標準パラメタの3つ組  $(n, \mu, \sigma)$  を指定する場合」に帰着する。

### 5.3 形状パラメタの組 $(g_p, g_s, \xi)$ を指定する場合

合成されたフィルタ  $\hat{F}$  の伝達関数  $\hat{g}(t)$  の形状パラメタの3つ組  $(g_p, g_s, \xi)$  が指定された場合には、それから条件を満たす標準パラメタの3つ組  $(n, \mu, \sigma)$  を決めれば、フィ

ルタの具体的な構成は前々副節 (5.1) の「標準パラメタの3つ組  $(n, \mu, \sigma)$  を指定する場合」に帰着する。

しかしながら、形状パラメタの組  $(g_p, g_s, \xi)$  が指定されたときに、それを正確に実現する標準パラメタの組  $(n, \mu, \sigma)$  を決めることは、次数  $n$  は整数であって連続的に変えられる量ではないことから、一般には不可能である。そのため、ここでは形状パラメタの3つのうちの2つについては値を指定して正確に一致させるが、残りの1つの値については不等式による制約条件を指定することにする。

以下の3通りのそれぞれの場合について、標準パラメタの3つ組  $(n, \mu, \sigma)$  を求める方法を以下に示す。

- (1)  $g_p$  と  $\xi$  は値を、 $g_s$  は上限値  $G_s$  をそれぞれ指定
- (2)  $g_s$  と  $\xi$  は値を、 $g_p$  は下限値  $G_p$  をそれぞれ指定
- (3)  $g_p$  と  $g_s$  は値を、 $\xi$  は上限値  $\Xi$  をそれぞれ指定

#### 5.3.1 $g_p$ と $\xi$ は値を $g_s$ は上限値 $G_s$ を指定する場合

この場合は、次数  $n$  には条件  $g_s \leq G_s$  を満たせる最小のものを採用することにする。さらに  $n$  の上限値  $n_{\max}$  (たとえば50とする) を適切に設定して、 $n = 1$  から始めて、以下の方法で条件を満たせる  $n$  の探索を行う。

まず式 (8) を用いて  $\xi$  から計算して  $\mu$  を求める。そうして、既に決まっている  $g_p$  と  $\mu$  の値および探索中に仮定する整数  $n$  の値を用いて、非線形方程式 (38) を解いて  $\sigma$  の値を決める。

$$F(\sigma) \equiv \frac{\cosh \left( 2n \sinh^{-1} \sqrt{\frac{\mu-1}{\sigma+1}} \right)}{\cosh \left( 2n \sinh^{-1} \sqrt{\frac{\mu}{\sigma}} \right)} - g_p = 0. \quad (38)$$

この非線形方程式  $F(x) = 0$  は適切に設定された  $x$  の非負の初期区間  $[x_L, x_H]$  から始めて2分法を用いることで解ける（まず  $F(0) = -g_p < 0$  である。そうして  $x$  を順次増大させてゆき（たとえば小さい正の値から始めて毎回2倍ずつにする）そうして初めて  $F(x) \geq 0$  が成り立つ  $x$  の値を見つけたならその  $x$  の値を初期区間の上端  $x_H$  とし、その直前のまだ  $F(x) < 0$  であった最後の  $x$  の値を初期区間の下端  $x_L$  にとればよい。あるいは簡単に  $x_L = 0$  と設定することもできる）。そうして2分法により  $\sigma$  を求めたら、式 (39) を用いて、 $g_s$  の値を  $n$  と  $\mu$  と  $\sigma$  から計算する。

$$g_s \leftarrow 1 / \cosh \left( 2n \sinh^{-1} \sqrt{\frac{\mu}{\sigma}} \right). \quad (39)$$

このようにして求めた  $g_s$  の値が  $G_s$  以下であれば条件を満たせたことになるので探索を終える。そうでなければ  $n$  の値を1つ増やして、 $n$  の値があらかじめ設定した上限値を越えたら探索は失敗とするが、そうでなければ増やした  $n$  について上記の方法により繰り返し調べる。

#### 5.3.2 $g_s$ と $\xi$ は値を、 $g_p$ は下限値 $G_p$ を指定する場合

ここでは次数  $n$  として、指定された条件を満たすことのできる最小のものを採用することにする。さらに  $n$  の値の上限値  $n_{\max}$  をあらかじめ適切に決めておき（たとえば50とする）、条件を満たす標準パラメタの3つ組  $(n, \mu, \sigma)$  をその範囲の  $n$  についてだけ探索することにする。

まず式 (8) を用いて  $\xi$  から  $\mu$  の値を計算で求める。そうして既に決まっている  $g_s$  と  $\mu$  の値と探索でいま仮定している整数  $n$  の値から、式 (40) を順に計算することで  $\sigma$  と  $g_p$  の値を求める。その計算により得られた  $g_p$  の値が条件  $g_p \geq G_p$  を満たすような  $n$  の最小値を探索で求める。

$$\begin{cases} \sigma \leftarrow \mu / \sinh^2 \left( \frac{1}{2n} \cosh^{-1} \frac{1}{g_s} \right), \\ g_p \leftarrow g_s \cosh \left( 2n \sinh^{-1} \sqrt{\frac{\mu-1}{1+\sigma}} \right). \end{cases} \quad (40)$$

最小値の探索法は、たとえば整数  $n$  を 1 から始めて  $n_{\max}$  まで順に上昇させて行くことができる。あるいは  $n$  についての 2 分法を用いて最初 1 と  $n_{\max}$  をそれぞれ上端と下端にして探索することもできる。なお、もしも  $n = n_{\max}$  において不等式  $g_p \geq G_p$  が満たせなければ、 $n_{\max}$  以下には条件を満たせる  $n$  は存在しない。

最小値  $n$  とその時の  $\sigma$  の値および  $\mu = h(\xi)$  を並べて作った標準パラメタの 3 つ組  $(n, \mu, \sigma)$  が得られたならば、合成されたフィルタ  $\hat{F}$  の構成は前々副節 (5.1) の「標準パラメタの 3 つ組  $(n, \mu, \sigma)$  を指定する場合」に帰着される。(なお本報告の趣旨からは外れるが、合成の次数  $\ell$  として大きい値を選び、それにより  $n = 1$  でもって条件が満たされるようにした場合は、それから得られる  $\hat{F}$  は選んだ合成の種類による「レゾルベントの線形結合の実部」として構成されたフィルタになる。)

### 5.3.3 $g_p$ と $g_s$ には値を、 $\xi$ は上限値 $\Xi$ を指定する場合

この場合、次数  $n$  として条件  $\xi \leq \Xi$  を満たせる最小のものを採用することにする。またさらに  $n$  の上限値  $n_{\max}$  を適切に設定して探索を行う。

まず  $\Xi$  から関係  $M = h(\Xi)$  を用いて  $\mu$  の上限値  $M$  を求める。そうして、既に決まっている  $g_p$  と  $g_s$  の値および探索中に仮定する整数  $n$  の値を用いて、以前に挙げた式 (37) を順に計算して  $\mu$  と  $\sigma$  の値を求める。このようにして求めた  $\mu$  の値が  $M$  以下であれば条件を満たせたことになるが、そうでなければ  $n$  のより大きい値を調べる。関数合成により得られたフィルタの伝達関数を表すためには、 $n$  にこのようにして求めた  $\mu$  と  $\sigma$  を並べた標準パラメタの 3 つ組  $(n, \mu, \sigma)$  が決まれば十分である。しかし必要ならば式 (9) を用いることで、 $\mu$  から求めた  $\xi$  が上限値  $\Xi$  を越えていないか調べることができる。

## 6. おわりに

単一のレゾルベントの多項式で、多項式に  $n$  次のチェビシェフ多項式を用いた場合の簡易型のフィルタは、次数  $n$  を増すことにより阻止域における伝達率の大きさの上限値  $g_s$  を容易に小さくできる。しかしそれと引き換えに、通過域における伝達率の最大最小比  $1/g_p$  が増大する。この比の値が大きいと (たとえば大きさが 4~6 桁)、得られる近似解の精度がそれだけバラつく可能性があるのが好ましくない。簡易型の伝達関数を用いる場合は特に、 $g_s$  を増大させずに  $1/g_p$  を減少させようとするには、遷移域の幅を決めるパラメタ  $\mu$  の値を大きくする必要がある。しかし遷移

域の幅  $\mu - 1$  が広くなれば固有値が遷移域にある不要な固有対の数が (通常は) 増す。不変部分空間の近似空間の基底をうまく構成するためには、通過域と遷移域を合併した区間に固有値が入る固有ベクトルの数よりも多くのベクトルに対してフィルタを適用する必要があるので、 $\mu$  が大きいとそれだけ多くのベクトルにフィルタを適用する必要性が生じて計算の手間が増えてしまい実用性が低下する。

今回の関数合成を用いる方法により遷移域の幅は縮小できるが、それと引き換えにフィルタの構成に必要なレゾルベントの数は元の 1 つから複数に増える。

とりあげた B-合成あるいは特に C-合成や I-合成、さらに E-合成では、たとえば  $\ell = 4$  の場合は虚数シフトのレゾルベントを 2 つだけ用いてそれらの線形結合の実部のチェビシェフ多項式で構成されたフィルタは、遷移域の幅が元のものよりもかなり狭くなり良好な特性が得られる。 $\ell = 6$  の場合には虚数シフトのレゾルベントを 3 つ用いて、さらに良好な特性が得られる。

そうして  $\ell$  が偶数の場合には、求めたい固有対の固有値の区間  $[a, b]$  は任意に位置を設定できるので中間固有対を求める場合に使う。ただし、連立 1 次方程式の係数行列の分解に掛かる演算量とその分解結果を保持するための記憶量は、用いるレゾルベントの数に比例する。それでもフィルタを複素シフトのレゾルベント 8~16 個程度の線形結合の実部で構成する場合に比べると、複素シフトのレゾルベントを少数 2~4 個用いて構成できるので、行列分解に掛かる演算量と特に行列の分解結果を保持する記憶量を減らす手段としては有効である。

今回新しく追加した E-合成によるフィルタの伝達特性は、次数  $\ell (> 2)$  を同じに揃えた場合には 4 種類の合成 (B, C, I, E) のうちで遷移域の幅が最も狭い急峻な遮断特性を実現できる。言い換えれば E-合成は急峻な遮断特性を最も小さい次数  $\ell$  で実現できる。そのことは 4 種類の合成によるフィルタのうちで構成に必要なレゾルベントの数が最も少ないのは E-合成によるものであることを意味する。すると、レゾルベントの作用を与える連立 1 次方程式を直接法で解く場合に、行列分解を行う演算量や分解結果を保持するための記憶量が最も少ないのは 4 種類の合成によるフィルタのうちで E-合成によるものであることになる。

付録の A.3 章「実験例」の中で、今回追加した E-合成 (楕円型合成) によるフィルタを用いて実対称定値一般固有値問題の近似対を求める実験を行なった例を示しておいた (例題 1~9)。問題の規模が大きくなって全体の計算の手間の中で行列分解の占める割合が増えると、他の合成 (B-合成, C-合成, I-合成) に比べて E-合成が経過時間が相対的に短く有利になる傾向を確認できる。

## 付 録

### A.1 E-合成について

これまで電気回路理論の典型フィルタの4種類を模倣する合成法として、バターワース型 (Butterworth) に対応するB-合成, チェビシェフ型 (Chebyshev) に対応するC-合成, 逆チェビシェフ型 (inverse Chebyshev) に対応するI-合成の3種類を提案して扱ってきた。今回の報告ではそれらに加えて, これまでに実装や実験をしていなかった電子回路論の典型フィルタの楕円型 (elliptic) に対応するE-合成について述べる。

#### A.1.1 楕円有理関数の紹介

まずE-合成の構成に用いる「楕円有理関数」(elliptic rational function) について簡単に紹介する。

##### A.1.1.1 楕円有理関数の楕円関数による媒介変数表示

楕円有理関数  $R_\ell(\xi, t)$  は  $t$  についての  $\ell$  次の実有理関数であり, 次数以外には実数のパラメタ  $\xi (> 1)$  を1つ持ち, ヤコビの楕円関数  $\text{cd}$  を用いた媒介変数表示 (A.1) を持つ [2], [3] (ここでは複素数の媒介変数として  $\phi$  を用いている)。

$$\begin{cases} R_\ell(\xi, t) &= \text{cd} \{ \ell \phi K(L^{-1}), L^{-1} \}, \\ t &= \text{cd} \{ \phi K(\xi^{-1}), \xi^{-1} \}. \end{cases} \quad (\text{A.1})$$

この表示に含まれる2つの定数パラメタは  $\xi > 1$  と  $L > 1$  であり,  $K(k)$  は式 (A.2) で定義される母数  $k$  の第1種完全楕円積分である。

$$K(k) \equiv \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}. \quad (\text{A.2})$$

なお関数  $\text{cd}$  はヤコビの楕円  $\text{sn}$  関数から実数周期の4分の1だけ座標をシフトした  $\text{cd}(u, k) \equiv \text{sn}(u + K(k), k)$  により定義される偶関数である (式 (A.1) は, Chebyshev 多項式の媒介変数  $\theta$  を用いた偶関数の三角関数  $\cos$  を用いた表示  $T_n(t) = \cos n\phi$ ,  $t = \cos \phi$  の類似になっている)。式 (A.1) に含まれている2つの定数パラメタ  $L > 1$  と  $\xi > 1$  は独立ではなく, 媒介変数表示が  $\ell$  次有理関数を表すための「次数方程式」と呼ばれる条件を満たす必要がある。

##### A.1.1.2 パラメタ $L$ と $\xi$ の満たすべき条件 (次数方程式)

媒介変数表示 (A.1) による関数  $R_\ell(\xi, t)$  が  $\ell$  次有理関数を表すためには, 表示に含まれるパラメタ  $L$  と  $\xi$  は独立ではなくて「次数方程式」(degree equation) と呼ばれる関係式 (A.3) を満たす必要がある [2], [3]。

$$K'(L^{-1})/K(L^{-1}) = \ell \times K'(\xi^{-1})/K(\xi^{-1}). \quad (\text{A.3})$$

ここで  $K'(k)$  は  $K'(k) \equiv K(k')$  で定義される関数であり,  $k' \equiv \sqrt{1 - k^2}$  は母数  $k$  の補母数を表している。

母数  $k$  に対する楕円 nome 関数  $q(k)$  の定義は式 (A.4) である。

$$q(k) \equiv \exp \{ -\pi K'(k)/K(k) \}. \quad (\text{A.4})$$

この楕円 nome 関数  $q(k)$  を用いて「次数方程式」(A.3) を書き直せば式 (A.5) になる。

$$q(L^{-1}) = \{ q(\xi^{-1}) \}^\ell. \quad (\text{A.5})$$

##### A.1.1.3 「次数方程式」の解法

楕円有理関数の次数が  $\ell$  であるとき, 「次数方程式」に等価な式 (A.5) を楕円 nome 関数とその逆関数を用いて解くことで,  $\xi$  を与えて  $L$  を求めることができる。また同様に  $L$  を与えて  $\xi$  を求めることもできる。楕円 nome 関数  $z = q(k)$  の逆関数  $k = q^{-1}(z)$  は巾級数展開を用いた式 (A.6) を用いて計算できる。この級数の収束は  $z (< 1)$  の値が小さければ速い (次数方程式の解法には楕円 nome 関数とその逆関数を用いない方法も存在する [2], [3])。

$$\begin{aligned} q^{-1}(z) &= \left\{ \frac{\sum_{m=1}^{\infty} z^{(m-\frac{1}{2})^2}}{1 + 2 \sum_{m=1}^{\infty} z^{m^2}} \right\}^2 \\ &= 4\sqrt{z} \left\{ \frac{\sum_{m=1}^{\infty} z^{m(m-1)}}{1 + 2 \sum_{m=1}^{\infty} z^{m^2}} \right\}^2 \\ &= 4\sqrt{z} \left( \frac{1 + z^2 + z^6 + z^{12} + \dots}{1 + 2z + 2z^4 + 2z^9 + \dots} \right)^2. \end{aligned} \quad (\text{A.6})$$

##### A.1.1.4 「楕円有理関数」の性質

楕円有理関数  $R_\ell(\xi, t)$  は  $t$  の  $\ell$  次の実有理関数で, 以下の性質を満たす ( $\xi > 1$  と  $L > 1$  は既に次数方程式を満たしているものとする) [2], [3]。

- (1)  $R_\ell(\xi, t)$  は次数  $\ell$  の偶奇に従って偶関数, 奇関数。
- (2)  $t \in [0, 1]$  で等リプル性を持ち,  $|R_\ell(\xi, t)| \leq 1$ 。
- (3)  $t \in (1, \xi)$  で単調増加。  $R_\ell(\xi, 1) = 1$ ,  $R_\ell(\xi, \xi) = L$ 。
- (4)  $t \in [\xi, \infty)$  では関数値の逆数が等リプル性を持ち,  $L \leq |R_\ell(\xi, t)|$ 。
- (5)  $R_\ell(\xi, \xi/t) = L/R_\ell(\xi, t)$  を満たす。

区間  $[-1, 1]$  で値の絶対値が1以下,  $t = 1$  での値が1である偶または奇の有理関数のうちで楕円有理関数が  $t = 1$  での増加率を最大にするという性質を利用すれば, 遷移域  $(1, \xi)$  で最も急峻に値が変化する伝達関数を実現できる。

##### A.1.1.5 楕円有理関数の明示的な有理関数表示

パラメタ  $\xi$  を持つ  $\ell$  次楕円有理関数  $R_\ell(\xi, t)$  は, すべての零点と極を用いて因子に分解された明示的な有理関数表示 (A.7) で表わすことができる [2], [3]。

$$R_\ell(\xi, t) = C t^{\text{mod}(\ell, 2)} \prod_{j=1}^{\lfloor \ell/2 \rfloor} \frac{t^2 - x_j^2}{t^2 - \tilde{x}_j^2} \quad (\text{A.7})$$

ここで零点  $\pm x_j$  と極  $\pm \tilde{x}_j$ ,  $j = 1, 2, \dots, \lfloor \ell/2 \rfloor$  はすべて実数であり, 実数引数のヤコビ楕円関数  $\text{sn}$  を用いて式 (A.8) で表される ( $\ell$  が奇数のとき零点  $t = 0$  と極  $t = \infty$  も持つ)。

$$\begin{cases} x_j &= \text{sn} \{ [2j - 1 + \text{mod}(\ell, 2)] K(\xi^{-1})/\ell, \xi^{-1} \}, \\ \tilde{x}_j &= \xi/x_j. \end{cases} \quad (\text{A.8})$$



定数  $C$  の逆数は、規格化条件  $R_\ell(\xi, 1) = 1$  を満たすように式 (A.9) で与えられる。

$$C^{-1} = \prod_{j=1}^{\lfloor \ell/2 \rfloor} \frac{1 - x_j^2}{1 - \tilde{x}_j^2}. \quad (\text{A.9})$$

### A.1.2 有理関数 $h(t)$ (E-合成)

いま E-合成に用いる  $\ell$  次の実有理関数  $h(t)$  を  $\ell$  次の楕円有理関数を用いた式 (A.10) により定義する。ただし  $\ell \geq 2$  であるとする。また  $L$  と  $\xi$  は次数条件を満たすとする。

$$h(t) \equiv \frac{L+1}{2} \times \frac{1 + R_\ell(\xi, t)}{L + R_\ell(\xi, t)} \quad (\text{A.10})$$

条件  $\mu = h(\xi)$  を満たすとするれば、 $\mu (> 1)$  と  $L (> 1)$  の関係を表す式 (A.11) が得られる。

$$\mu = \frac{(L+1)^2}{4L} \quad \text{または} \quad L = (\sqrt{\mu} + \sqrt{\mu-1})^2. \quad (\text{A.11})$$

#### A.1.2.1 有理関数 $h(t)$ の性質 (E-合成)

式 (A.10) で表される関数  $h(t)$  は、楕円有理関数  $R_\ell(\xi, t)$  が持つ性質から導かれる以下の性質を持つ。

- (1)  $\ell$  が偶数のときは偶関数。
- (2)  $t \in [0, 1]$  では等リプル性を持ち、 $[0, 1]$  への全射。
- (3)  $t \in (1, \xi)$  で単調増加で  $(1, \mu)$  への全射。
- (4)  $t \in [\xi, \infty)$  では振動して極を持ち、 $[\mu, \infty)$  への全射で、すべての極小値は  $\mu$  に一致。

すると  $h(t)$  は合成用の有理関数が満たすべき性質 (3.1 節) を備えていることがわかる。

なお  $\ell \geq 2$  であれば  $\ell$  が偶数でも奇数でも、この E-合成で用いる合成用の関数  $h(t)$  の  $t \in [0, 1]$  と  $t \in [-1, 0]$  の両方に対する値域はどちらも  $[0, 1]$  であって一致する。このことから、 $\ell \geq 2$  の場合には、 $\ell$  の偶奇によらず常に E-合成で得られるフィルタの通過域を  $t \in [0, 1]$  から  $t \in [-1, 1]$  に広げることができる。

### A.1.3 伝達関数の設計方法の例 (E-合成)

Chebyshev 多項式の次数  $n$  と伝達率の閾値  $g_p$  と  $g_s$  の 3 つを指定すれば簡単な式計算により  $\sigma$  と  $\mu$  の値が求まる。逆に  $n, \sigma, \mu$  の 3 つを与えれば  $g_p$  と  $g_s$  が決まる。そうして  $\mu$  からは式 (A.11) により  $L$  が決まる。するとあとは楕円有理関数の次数  $\ell$  が決まれば  $h(t)$  は完全に決まる。

次数  $\ell$  を指定する場合には、「次数方程式」を解いて  $L$  から  $\xi$  が決まるので、 $h(t)$  に含まれる  $L, \ell, \xi$  がすべて決まるので合成関数  $\hat{x}(t) = x(h(t))$  の決定が完了する。次数  $\ell$  を指定するのではなくて  $\xi$  の上限値  $\xi_{\max}$  を指定する場合には、実数値  $\left\{ \frac{K'(L^{-1})}{K(L^{-1})} \right\} / \left\{ \frac{K'(\xi_{\max}^{-1})}{K(\xi_{\max}^{-1})} \right\}$  を切り上げて得られる整数値を  $\ell_{\min}$  として、それ以上の整数である次数  $\ell$  を選ぶと、その  $\ell$  に対する「次数方程式」を解いて  $L$  から決まる  $\xi$  の値は必ず  $\xi_{\max}$  以下となり、それで  $\hat{x}(t)$  の決定は完了する。ただし構成するフィルタが中間固有対を求めためのものである場合には、次数  $\ell$  として奇数は選ばないので、 $\ell$  は  $\ell_{\min}$  以上の偶数に選ぶ必要がある。

ここまでで  $\ell, L, \xi$  は決まったので、楕円有理関数  $R_\ell(\xi, t)$  が定まり、 $h(t) \equiv \frac{L+1}{2} \times \frac{1 + R_\ell(\xi, t)}{L + R_\ell(\xi, t)}$  が定まり、さらに  $\sigma$  と  $\mu$  も使って  $\hat{x}(t) \equiv \frac{\mu + \sigma}{h(t) + \sigma}$  が定まる。

以上の手順、まず次数  $n$  と閾値  $g_p$  と  $g_s$  を指定して  $\sigma, \mu, L$  を決め、さらに次数  $\ell$  あるいは  $\xi_{\max}$  を指定することで  $\ell$  と  $\xi$  の値を決め、それにより有理関数  $\hat{x}(t) = x(h(t))$  が決まるまでを図 A.1 に示す。

- 1) 次数  $n$  と閾値  $g_p$  と  $g_s$  を「指定する」。
- 2)  $w_1 \leftarrow \frac{1}{2n} \cosh^{-1} \frac{1}{g_s}$  と  $w_2 \leftarrow \frac{1}{2n} \cosh^{-1} \frac{g_p}{g_s}$  を計算。
- 3)  $\sigma \leftarrow \frac{\cosh^2 w_2}{\sinh(w_1 + w_2) \sinh(w_1 - w_2)}$ ,  $\mu \leftarrow \sigma \sinh^2 w_1$  を計算。
- 4)  $L \leftarrow (\sqrt{\mu} + \sqrt{\mu-1})^2$  とする。
- 5)  $\ell$  の決め方は、最初から  $\ell$  を「指定する」、あるいは  $\xi$  の上限  $\xi_{\max}$  を「指定して」、実数値  $\left\{ \frac{K'(L^{-1})}{K(L^{-1})} \right\} / \left\{ \frac{K'(\xi_{\max}^{-1})}{K(\xi_{\max}^{-1})} \right\}$  以上の整数を  $\ell$  に設定する。
- 6)  $q(k)$  とその逆関数を用いて  $\xi \leftarrow 1/q^{-1}(\{q(1/L)\}^{1/\ell})$ 。
- 7)  $h(t) \equiv \frac{L+1}{2} \times \frac{1 + R_\ell(\xi, t)}{L + R_\ell(\xi, t)}$ ,  $\hat{x}(t) \equiv \frac{\mu + \sigma}{h(t) + \sigma}$ 。

図 A.1 次数  $n$  と閾値  $g_p$  と  $g_s$  および次数  $\ell$  あるいは  $\xi_{\max}$  を与えて  $\hat{x}(t) = x(h(t))$  を決定する手順 (E-合成の場合)

#### A.1.3.1 遷移域の幅 $\xi-1$ の値の例 (E-合成)

E-合成で、 $g_s = 10^{-15}$  と固定して、Chebyshev 多項式の次数  $n$ 、閾値  $g_p$ 、次数  $\ell$  を指定したいいくつかの組み合わせについて、(拡大された) 通過域  $t \in [-1, 1]$  の半値幅 1 に対する (片方の) 遷移域  $t \in (1, \xi)$  の幅の相対値である  $\xi-1$  の値を表 A.1 に示す。次数  $\ell$  が増えるにつれて急速に遷移域の幅が狭くなる傾向が読み取れる。

表 A.1 E-合成による遷移域の幅  $\xi-1$  (相対値) の例

$g_p$	楕円有理関数の次数 $\ell = 4$			楕円有理関数の次数 $\ell = 5$		
	$n=10$	$n=15$	$n=20$	$n=8$	$n=10$	$n=12$
$10^{-1}$	0.408	0.251	0.200	0.250	0.168	0.127
$10^{-2}$	0.238	0.133	0.101	0.144	0.090	0.064
$10^{-3}$	0.155	0.081	0.058	0.092	0.054	0.037
$g_p$	楕円有理関数の次数 $\ell = 6$			楕円有理関数の次数 $\ell = 8$		
	$n=6$	$n=8$	$n=10$	$n=6$	$n=8$	$n=10$
$10^{-1}$	0.221	0.115	0.073	0.0599	0.0263	0.0147
$10^{-2}$	0.133	0.061	0.035	0.0316	0.0117	0.0057
$10^{-3}$	0.087	0.036	0.019	0.0184	0.0059	0.0026

また図 A.2 から図 A.8 までの 7 枚の図はそれぞれ、合成用の有理関数の次数  $\ell$  を 2 から 7 まで変えた各場合について横軸に  $\mu$  の常用対数の値を、縦軸に  $\xi$  の常用対数の値をとって  $\mu$  と  $\xi$  の対応関係をプロットしたグラフを 4 種類の合成 (B-合成, C-合成, I-合成, E-合成) について描いたものである。ただし C-合成と I-合成のグラフは常に同じものになるので 1 つにまとめている。4 種類の合成は次数が  $\ell = 2$  のときにはすべて  $h(t) = t^2$  で  $\mu$  と  $\xi$  の対応

関係は  $\xi = \sqrt{\mu}$  になるので、各合成のグラフの線は完全に重なっている (図 A.2)。パラメタ  $\mu$  と  $\xi$  はそれぞれ合成前と合成後の遷移域の端の位置を表すものであり ( $\mu - 1$  と  $\xi - 1$  がそれぞれ遷移域の幅である)、4種類の合成はどの場合にも  $\log_{10} \xi$  の値が  $\log_{10} \mu$  よりも小さく、つまり  $\xi$  が  $\mu$  よりも小さくなっており、合成により伝達関数の遷移域の幅が縮小される様子を表している。これらのグラフから、遷移域の幅を縮小する能力が最も優れているのは E-合成であり、その次が C-合成およびそれと等しい結果になる I-合成であり、最後が B-合成であることがわかる。

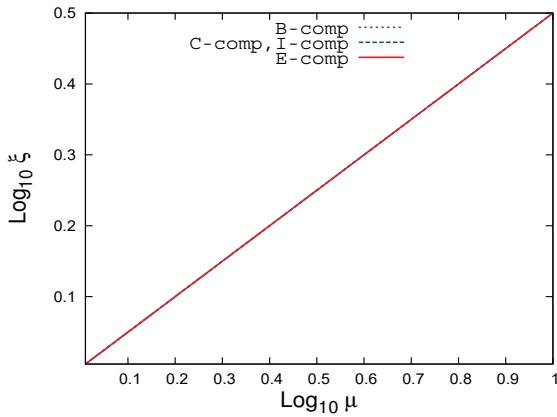


図 A.2 各合成での  $\mu$  と  $\xi$  の対応 (次数  $l = 2$ )

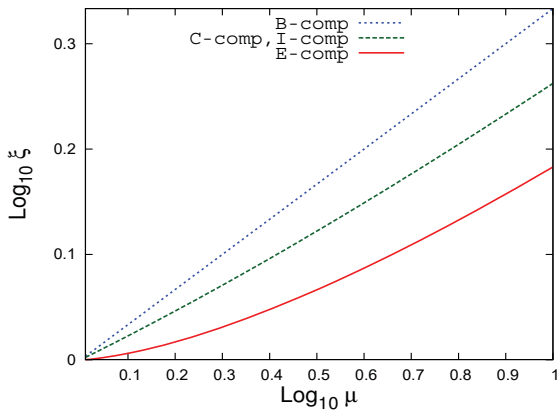


図 A.3 各合成での  $\mu$  と  $\xi$  の対応 (次数  $l = 3$ )

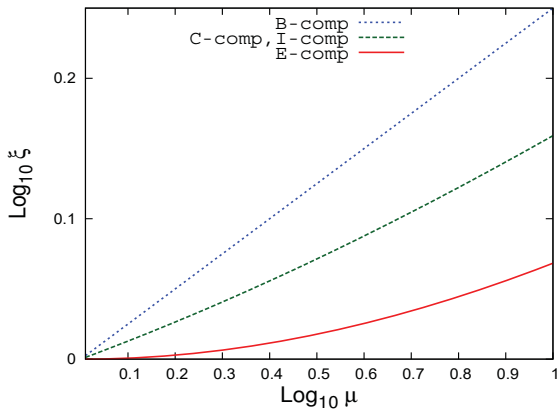


図 A.4 各合成での  $\mu$  と  $\xi$  の対応 (次数  $l = 4$ )

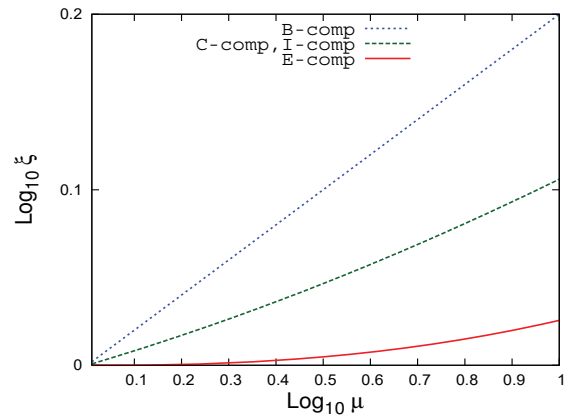


図 A.5 各合成での  $\mu$  と  $\xi$  の対応 (次数  $l = 5$ )

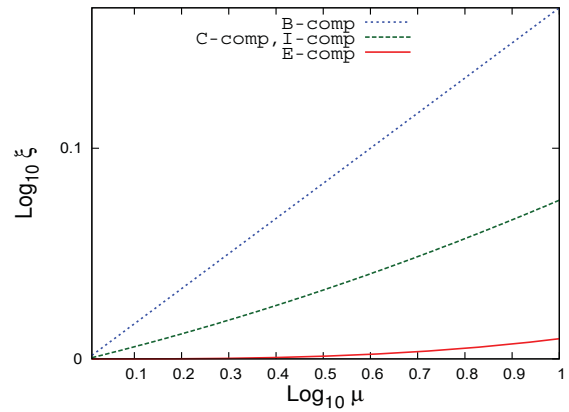


図 A.6 各合成での  $\mu$  と  $\xi$  の対応 (次数  $l = 6$ )

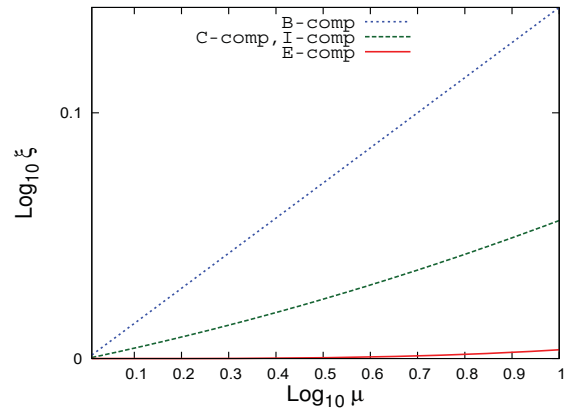


図 A.7 各合成での  $\mu$  と  $\xi$  の対応 (次数  $l = 7$ )

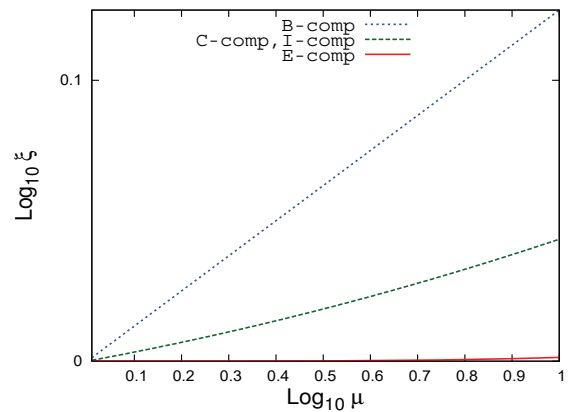


図 A.8 各合成での  $\mu$  と  $\xi$  の対応 (次数  $l = 8$ )

また、図 A-9、図 A-10、図 A-11 の各グラフからも、B-合成、C-合成 (I-合成)、E-合成のそれぞれの場合について、次数  $\ell$  を増すとき遷移域の幅  $\xi - 1$  が縮小される様子を読みとることができる。これら 3 枚の図のグラフを見比べると遷移域の幅を縮小する能力は E-合成の場合が最も優れており、次が C-合成と I-合成、そして最後が B-合成であることを再確認できる。

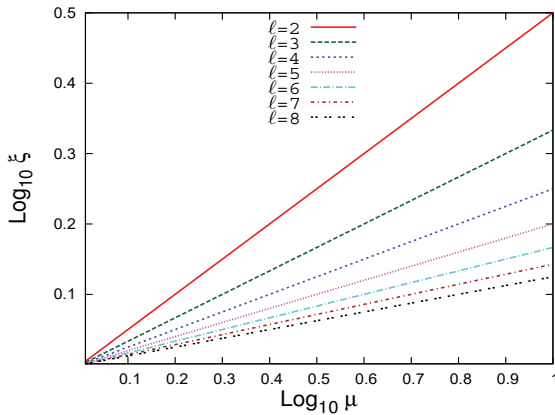


図 A-9 各次数  $\ell$  に対する  $\mu$  と  $\xi$  の対応 (B-合成)

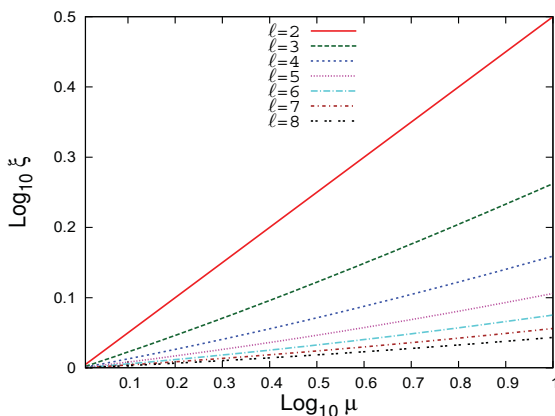


図 A-10 各次数  $\ell$  に対する  $\mu$  と  $\xi$  の対応 (C-合成または I-合成)

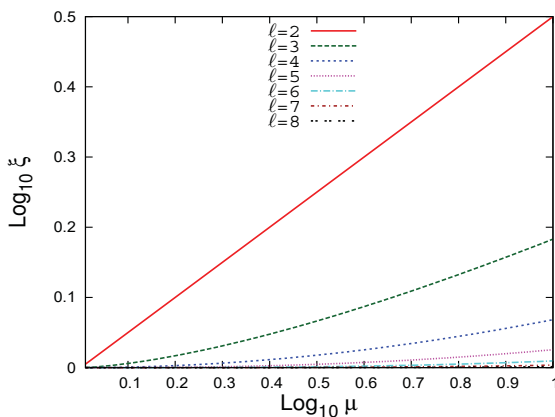


図 A-11 各次数  $\ell$  に対する  $\mu$  と  $\xi$  の対応 (E-合成)

#### A.1.4 $\hat{x}(t)$ の部分分数分解 (E-合成)

E-合成の場合の実有理関数  $\hat{x}(t) = x(h(t))$  の式は (A.12) である。

$$\hat{x}(t) = \frac{\mu + \sigma}{h(t) + \sigma}, \quad h(t) = \frac{L+1}{2} \times \frac{1 + R_\ell(\xi, t)}{L + R_\ell(\xi, t)}. \quad (\text{A.12})$$

いま複素数の範囲での  $\hat{x}(t)$  の部分分数分解を式 (A.13) として、その定数項  $c_\infty$ 、極  $t_j$ 、極の係数  $c_j$ ,  $j = 1, 2, \dots, \ell$  をそれぞれ求める。

$$\hat{x}(t) = c_\infty + \sum_{j=1}^{\ell} \frac{c_j}{t - t_j}. \quad (\text{A.13})$$

##### A.1.4.1 定数項 $c_\infty$ の計算 (E-合成)

まず最初に定数項  $c_\infty$  を求める。それにはまず  $\lim_{t \rightarrow \infty} R_\ell(\xi, t) = \lim_{t \rightarrow \infty} L/R_\ell(\xi, \xi/t) = L/R_\ell(\xi, 0)$  であるが、 $\ell$  が奇数の場合には  $R_\ell(\xi, t)$  は奇関数なので  $R_\ell(\xi, 0) = 0$  である。また  $\ell$  が偶数の場合にはさらに  $\ell/2$  が偶数であれば  $R_\ell(\xi, 0) = 1$  であり、 $\ell/2$  が奇数であれば  $R_\ell(\xi, 0) = -1$  であることもわかる。

すると  $\lim_{t \rightarrow \infty} h(t)$  の値は、 $\ell$  が奇数の場合は  $\frac{1}{2}(L+1)$  であるが、 $\ell$  が偶数の場合にはさらに  $\ell/2$  が偶数ならば  $\mu$  になり、 $\ell/2$  が奇数ならば  $\infty$  になる。よって  $c_\infty = \lim_{t \rightarrow \infty} \hat{x}(t)$  の値は、 $\ell$  が奇数の場合は  $(\mu + \sigma) / \{\frac{1}{2}(L+1) + \sigma\}$  であり、 $\ell$  が偶数の場合にはさらに  $\ell/2$  が偶数ならば 1 であり、 $\ell/2$  が奇数ならば 0 になる。

以上をまとめると、定数項  $c_\infty$  の値は以下の式 (A.14) になる。

$$c_\infty = \begin{cases} \frac{2(\mu + \sigma)}{L + (2\sigma + 1)} & (\text{mod}(\ell, 4) = 1, 3 \text{ の場合}), \\ 1 & (\text{mod}(\ell, 4) = 0 \text{ の場合}), \\ 0 & (\text{mod}(\ell, 4) = 2 \text{ の場合}). \end{cases} \quad (\text{A.14})$$

##### A.1.4.2 部分分数分解の極 $t_j$ の計算 (E-合成)

式 (A.12) で与えられる実有理関数  $\hat{x}(t)$  の  $\ell$  個の極を複素数の範囲で求める。極  $t$  は  $h(t) + \sigma = 0$  の解であることから以下の方程式 (A.15) の解になる。

$$R_\ell(\xi, t) = -\Gamma, \quad \Gamma \equiv 1 + \frac{2(L-1)\sigma}{L + (2\sigma + 1)} > 1. \quad (\text{A.15})$$

楕円有理関数  $R_\ell(\xi, t)$  のヤコビ楕円関数  $\text{cd}$  による  $\phi$  を媒介変数とする表示 (A.1) を用いると、(A.16) が得られる (上側の式を満たす  $\phi$  の値を求めて、それを下側の式中の  $\phi$  とすれば  $t$  の値が得られる)。ただし  $L$  と  $\xi$  は次数  $\ell$  に対する「次数方程式」(A.3) を満たしているものとする。

$$\begin{cases} -\Gamma = \text{cd} \{ \ell \phi K(L^{-1}), L^{-1} \}, \\ t = \text{cd} \{ \phi K(\xi^{-1}), \xi^{-1} \}. \end{cases} \quad (\text{A.16})$$

いま (A.16) の上側の式を、楕円関数  $\text{cd}$  の性質を用いて符号を変えて式の値が正となるように (引数を実数周期



$4K(L^{-1})$  の半分だけずらして) 書き直すと以下の式 (A.17) となる.

$$\text{cd}\{(\ell\phi + 2)K(L^{-1}), L^{-1}\} = \Gamma (> 1). \quad (\text{A.17})$$

楕円関数  $\text{cd}$  が実数周期  $4K(L^{-1})$  を持つ周期関数であることにより, 式 (A.17) の左辺の式の値は任意の整数  $j$  に対して  $\phi$  を  $\phi + 4j/\ell$  に置き換えても変わらない. すると式 (A.17) を満たす  $\phi$  の 1 つの値  $\phi_0$  を得たならば, それから式 (A.18) により  $t_j$  の  $\ell$  個の値を生成すれば, (生成された値  $t_j$  がすべて異なれば) 必要な  $\ell$  個すべての解が得られる.

$$t_j = \text{cd}\{(\phi_0 + 4j/\ell)K(\xi^{-1}), \xi^{-1}\}, \quad j=1, 2, \dots, \ell. \quad (\text{A.18})$$

ヤコビの楕円関数  $\text{cd}$  の絶対値は実数の引数に対しては常に 1 以下であるが, いま  $\Gamma > 1$  であるので, (A.17) の  $\phi$  は実数ではないことがわかる. 関数  $\text{cd}$  は偶関数なので, 純虚数の引数に対しては実数値を与える. そこでいま  $y$  を実数として, 式 (A.17) の関数  $\text{cd}$  の (第 1 の) 引数を純虚数  $-y\sqrt{-1}$  に等しいとおいて式 (A.19) をたてる.

$$(\ell\phi + 2)K(L^{-1}) = -y\sqrt{-1}. \quad (\text{A.19})$$

すると (A.17) から (A.20) が得られる.

$$\text{cd}(-y\sqrt{-1}, L^{-1}) = \Gamma. \quad (\text{A.20})$$

ヤコビの楕円関数  $\text{cd}$  に対しては虚数変換の公式 (A.21) [1][10] がある.

$$\text{cd}(u\sqrt{-1}, k) = \frac{1}{\text{dn}(u, k')}. \quad (\text{A.21})$$

この虚数変換の公式を  $u = -y$ ,  $k = L^{-1}$  として (A.20) にあてはめると式 (A.22) が得られる. ここで  $(L^{-1})'$  は  $L^{-1}$  の補母数  $\sqrt{1 - L^{-2}}$  である.

$$\text{dn}\{-y, (L^{-1})'\} = \Gamma^{-1}. \quad (\text{A.22})$$

ヤコビの楕円関数の間の恒等関係式  $\text{dn}^2(u, k) = 1 - k^2 \text{sn}^2(u, k)$  から,  $\text{sn}^2(u, k) = (1/k^2)\{1 - \text{dn}^2(u, k)\}$  であることを用いると, 式 (A.23) が得られる. ここで複号は任意だが以下ではマイナス符号を採用することにする (この選択は極  $t_j$  の虚部の符号の分布の都合である).

$$\text{sn}\{-y, (L^{-1})'\} = \pm\sqrt{(1 - \Gamma^{-2})/(1 - L^{-2})}. \quad (\text{A.23})$$

関係  $\text{sn}(u, k) = x$  を第 1 種楕円積分を用いて表せば  $u = F(\sin^{-1} x, k)$  であるから, 実数  $y$  の値は以下の式 (A.24) により表わせる.

$$y = F\left(\sin^{-1} \sqrt{(1 - \Gamma^{-2})/(1 - L^{-2})}, (L^{-1})'\right). \quad (\text{A.24})$$

第 1 種楕円積分  $F(\varphi, k)$  の定義は式 (A.25) である.

$$F(\varphi, k) \equiv \int_0^\varphi \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}. \quad (\text{A.25})$$

ここでさらに  $F'(\varphi, k) \equiv F(\varphi, k')$  を定義する. ただし  $k'$  は母数  $k$  の補母数  $k' \equiv \sqrt{1 - k^2}$  を表している.  $F(\varphi, k)$  の値を算術幾何平均の方法を用いて求める算法 [4] を僅かに変更することで, 母数  $k$  が小さい場合に  $F'(\varphi, k)$  の値を良い精度で計算できる方法が得られる. そうして  $k = L^{-1} \ll 1$  である場合にはそのような方法を用いる方が, 第 1 種楕円積分  $F(\varphi, k)$  の値を求める通常の計算ルーチンに  $\varphi$  と  $(L^{-1})' = \sqrt{1 - L^{-2}}$  の 2 つの値を引数として渡して計算するよりも精度の良い結果が得られる.

得られた  $y$  の値を用いて式 (A.19) を解くことで,  $\phi$  の 1 つの解  $\phi_0$  が式 (A.26) で与えられる.

$$\phi_0 = \frac{1}{\ell} \left\{ -2 - \frac{y}{K(L^{-1})} \sqrt{-1} \right\}. \quad (\text{A.26})$$

そうして式 (A.18) を用いると, 部分分数分解に現れるすべての極  $t_j$  は複素引数の楕円関数  $\text{cd}$  あるいは  $\text{sn}$  の値を用いた式 (A.27) により表せる.

$$\begin{cases} \omega_j &= \frac{1}{\ell} \left\{ 4j - 2 - \frac{y}{K(L^{-1})} \sqrt{-1} \right\}, \\ t_j &= \text{cd}\{\omega_j K(\xi^{-1}), \xi^{-1}\} \\ &= -\text{sn}\{(\omega_j - 1)K(\xi^{-1}), \xi^{-1}\}. \end{cases} \quad (\text{A.27})$$

引数が複素数である楕円関数  $\text{sn}$  の値は, 楕円関数の加法公式により, 引数の実部と虚部それぞれに対する実引数の楕円関数の値を組み合わせることで計算できる ([1][10]). 複素数の引数である  $(\omega_j - 1)K(\xi^{-1})$  の虚部が添字  $j$  を含んでいないことを用いると, 複素引数に対する楕円関数の値の計算において手間を減らすことができる.

式 (A.27) で表わされる複素数の極  $t_j$  は, 添字が  $j = 1, 2, \dots, \lfloor \ell/2 \rfloor$  のものは (式 (A.23) の複号でマイナスの符号を選択したことにより) 虚部が正であり,  $\ell$  が奇数の場合に限り唯一存在する実数の極の添字は  $j = (\ell + 1)/2$  であり, その値は  $-1$  より小さい.

極を計算する式をまとめると式 (A.28) になる. まず  $\sigma$  と  $L$  から  $\sin \varphi$  あるいは  $\cos \varphi$  を計算して  $\varphi$  を逆三角関数を用いて求める ( $\sin \varphi$  と  $\cos \varphi$  の式はどちらも 1 に極めて近い値をとらないことを示せるが,  $\varphi$  の値は  $\cos \varphi$  の値から求めた方がよさそうである). 次に実数  $y (> 0)$  の値を, 第 2 引数が母数  $k$  の補母数  $k'$  である第 1 種楕円積分  $F(u, k')$  を求めるルーチン  $F'(u, k)$  を用いて,  $y = F'(\varphi, L^{-1})$  と計算して実数  $y$  を求める. そうして添字  $j$  に対して複素数の値  $\omega_j$  を計算して, それを用いて  $t_j = -\text{sn}\{(\omega_j - 1)K(\xi^{-1}), \xi^{-1}\}$  により引数が複素数である楕円関数  $\text{sn}$  の値を計算すれば複素数の極  $t_j$  が求まる. この式において正の虚部を持つ極は, 前半の添字  $j = 1, 2, \dots, \lfloor \ell/2 \rfloor$  に対応している. 実数の極は  $\ell$  が奇数のときに限り唯一存在して添字  $j = (\ell + 1)/2$  を持ち, その値  $t_R = t_{(\ell+1)/2}$  は必ず  $t_R < -1$  である.

$$\begin{cases} \sin \varphi = \sqrt{\frac{1 - \Gamma^{-2}}{1 - L^{-2}}} = \frac{2L\sqrt{\sigma(\sigma+1)}}{(2\sigma+1)L+1}, \\ \cos \varphi = \sqrt{\frac{\Gamma^{-2} - L^{-2}}{1 - L^{-2}}} = \frac{\sqrt{L^2 + 2(2\sigma+1)L + 1}}{(2\sigma+1)L+1}, \\ y = F(\varphi, (L^{-1})') = F'(\varphi, L^{-1}), \\ \omega_j = \frac{1}{\ell} \left\{ 4j - 2 - \frac{y}{K(L^{-1})} \sqrt{-1} \right\}, \\ t_j = -\operatorname{sn} \{ (\omega_j - 1)K(\xi^{-1}), \xi^{-1} \}. \end{cases} \quad (\text{A.28})$$

例として、E-合成の場合の  $\hat{x}(t)$  の極の複素平面内での配置を示す図を4つ示す(図 A.12 から図 A.15). 図中で黒い横線は区間  $t \in [-1, 1]$  である.

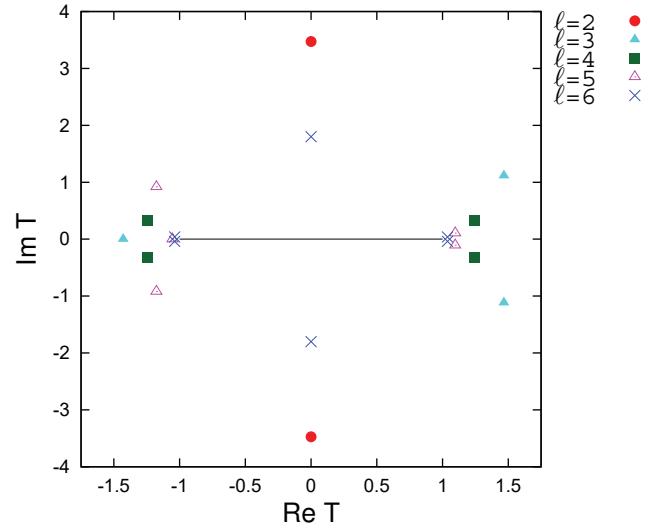


図 A.14  $\hat{x}(t)$  の極の配置の例 (E-合成)  $n=20, g_s=10^{-15}, g_p=10^{-1}$

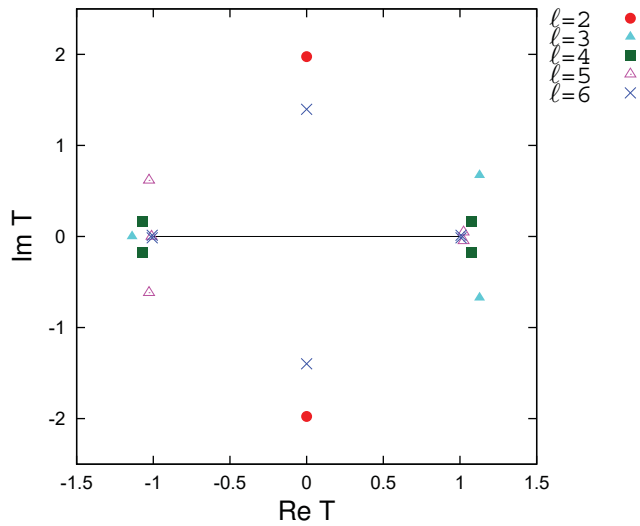


図 A.12  $\hat{x}(t)$  の極の配置の例 (E-合成)  $n=20, g_s=10^{-15}, g_p=10^{-3}$

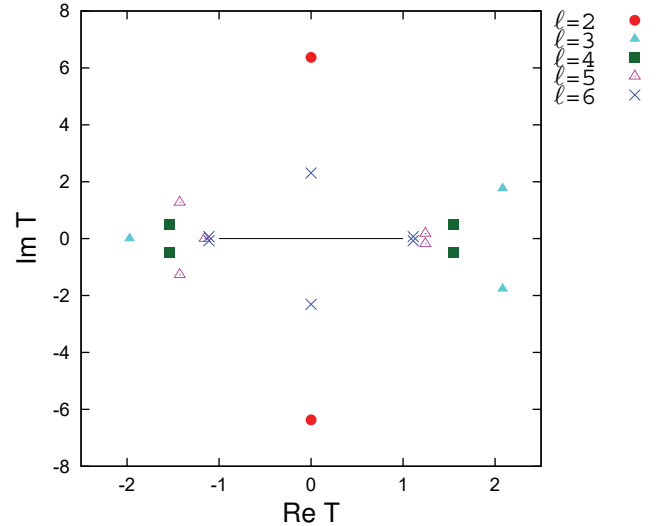


図 A.15  $\hat{x}(t)$  の極の配置の例 (E-合成)  $n=20, g_s=10^{-15}, g_p=0.5$

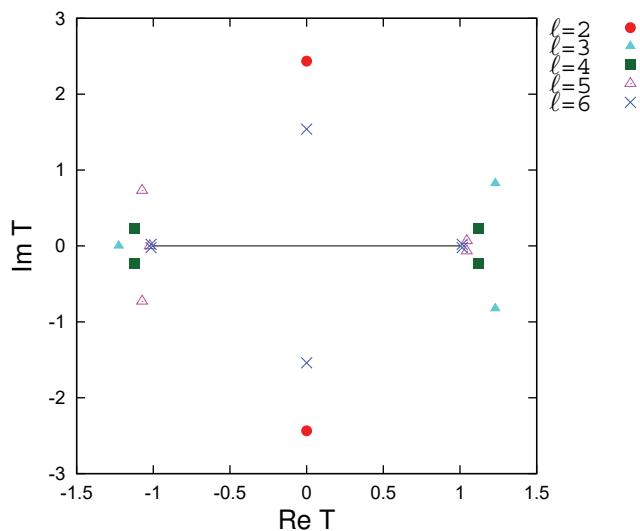


図 A.13  $\hat{x}(t)$  の極の配置の例 (E-合成)  $n=20, g_s=10^{-15}, g_p=10^{-2}$

#### A.1.4.3 部分分数分解の極の係数 $c_j$ の計算 (E-合成)

楕円有理関数の明示的な有理関数表示式 (A.7) およびその表示式に現れる (すべて実数である) 零点と極の値の明示表式 (A.8) を用いれば、楕円有理関数  $R(\xi, t)$  の対数微分  $\Psi(t) \equiv \frac{d}{dt} \log R_\ell(\xi, t) = R'_\ell(\xi, t) / R_\ell(\xi, t)$  の値を計算するのに便利な式 (A.29) が導かれ、それを用いると与えられた複素数  $t$  に対する  $\Psi(t)$  の値を計算できる ( $R'_\ell(\xi, t)$  はここでは  $R_\ell(\xi, t)$  の導関数を表している).

$$\Psi(t) = \frac{\operatorname{mod}(\ell, 2)}{t} + 2t \sum_{i=1}^{\lfloor \ell/2 \rfloor} \left( \frac{1}{t^2 - x_i^2} - \frac{1}{t^2 - \tilde{x}_i^2} \right). \quad (\text{A.29})$$

いま  $h(t) = \frac{L+1}{2} \times \frac{1 + R_\ell(\xi, t)}{L + R_\ell(\xi, t)}$  より、合成された有

理関数は  $\hat{x}(t) = x(h(t)) = \frac{\mu + \sigma}{h(t) + \sigma} = \frac{2(\mu + \sigma)}{L + (2\sigma + 1)} \times \frac{L + R_\ell(\xi, t)}{R_\ell(\xi, t) + \Gamma}$  である。これを用いて式 (A.13) における極の係数  $c_j$  を式 (A.30) のように計算する (途中の変形では、 $R_\ell(\xi, t_j) = -\Gamma$  であることや、 $\frac{L - \Gamma}{\Gamma} = \frac{L^2 - 1}{(2\sigma + 1)L + 1}$  となることを用いている)。その結果、極の係数  $c_j$  は対数微分  $\Psi(t)$  の  $t = t_j$  での値を用いて表されることがわかる。

$$\begin{aligned} c_j &= \lim_{t \rightarrow t_j} (t - t_j) \hat{x}(t) \\ &= \frac{2(\mu + \sigma)(L - \Gamma)}{L + (2\sigma + 1)} \times \lim_{t \rightarrow t_j} \frac{t - t_j}{R_\ell(\xi, t) + \Gamma} \\ &= \frac{2(\mu + \sigma)(L - \Gamma)}{L + (2\sigma + 1)} \times \frac{1}{R'_\ell(\xi, t_j)} \\ &= -\frac{2(\mu + \sigma)}{L + (2\sigma + 1)} \times \frac{L - \Gamma}{\Gamma} \times \frac{1}{R'_\ell(\xi, t_j)/R_\ell(\xi, t_j)} \\ &= -\frac{2(\mu + \sigma)}{L + (2\sigma + 1)} \times \frac{L^2 - 1}{(2\sigma + 1)L + 1} \times \frac{1}{\Psi(t_j)}. \end{aligned} \quad (\text{A.30})$$

まとめると、E-合成の場合の  $\hat{x}(t)$  の部分分数分解の式 (A.13) に現れる極  $t_j$  に対する係数  $c_j$ ,  $j = 1, 2, \dots, \ell$  は式 (A.31) を計算することで求められる。

$$c_j \leftarrow -\frac{2(\mu + \sigma)(L^2 - 1)}{\{L + (2\sigma + 1)\}\{(2\sigma + 1)L + 1\}} \times \frac{1}{\Psi(t_j)}. \quad (\text{A.31})$$

## A.2 実験について

### A.2.1 使用した計算機システム環境

実験に使用した計算機は、東京大学情報基盤センターに設置された Oakbridge-CX の 1 ノード (富士通 PRIMERGY CX2550M5) である。その 1 ノードは密結合の 2 つの CPU で構成されている。ノード内の主記憶は DDR4 で容量は 192GiB (ユーザが利用できるのは 168GiB) であり、メモリのバンド幅は最大 281.6GiB/s である。CPU は Intel Xeon Platinum 8280 で、1 つの CPU は 28 コア、クロック 2.7GHz, L2 キャッシュ 1MiB/コア, L3 キャッシュ 38.5MiB/CPU, CPU のメモリチャネル数は 6, Intel AVX-512 拡張命令セットで、AVX-512 FMA ユニット数は 2. Hyper-Thread 機能や Turbo-Boost 機能 (最大周波数 4.0GHz) が有効になっているかどうかは不明である。OS は CentOS 7 for x86\_64 (64bit 版) である。

プログラムは Fortran90 と OpenMP ディレクティブを入れてコーディングをし、コンパイラは Intel Fortran v19.0.4.243 for x86\_64 で、コンパイラのオプションに "-fast -openmp -axCORE-AVX512" を指定して OpenMP により 2CPU の合計 56 スレッド並列で実行した。

### A.2.2 計算の主要部

計算の主要部は、行列  $X$  と  $Y$  を実あるいは複素の  $N$  次の縦ベクトル  $m$  個を列方向に並べた  $N \times m$  行列とするときに、与えられた  $X$  に対してレゾルベント  $\mathcal{R}(\rho_j)$  を適用

した結果である  $Y \leftarrow \mathcal{R}(\rho_j)X$  を求めるところである。今回のフィルタは  $\hat{\mathcal{F}} = g_s T_n(\hat{Y})$  であり、 $\hat{Y} \equiv 2\hat{X} - I$  は以下の式 (A.32) により表される実線形作用素 (実ベクトルに対して実ベクトルを与える線形作用素) である。

$$\hat{Y} = \begin{cases} (2c_\infty - 1)I + \sum_{j=1}^{\ell/2} \text{Re}\{4\ell_j \mathcal{R}(\rho_j)\} & (\ell \text{ が偶数}), \\ (2c_\infty - 1)I + 2\ell_R \mathcal{R}(\rho_R) + \sum_{j=1}^{(\ell-1)/2} \text{Re}\{4\ell_j \mathcal{R}(\rho_j)\} & (\ell \text{ が奇数}). \end{cases} \quad (\text{A.32})$$

$A$  と  $B$  を  $N$  次の実対称行列としてシフトが  $\rho_j$  のレゾルベントは  $\mathcal{R}(\rho_j) \equiv (A - \rho_j B)^{-1} B$  なので、 $C^{(j)} \equiv A - \rho_j B$  とすると、 $Y^{(j)} \leftarrow \mathcal{R}(\rho_j)X$  の計算は、各  $j$  について右辺が共通な値  $BX$  を持つ連立 1 次方程式  $C^{(j)}Y^{(j)} = BX$  を  $Y^{(j)}$  について解くことに帰着する。行列  $C^{(j)}$  は対称で、シフト  $\rho_j$  が実数であるか虚数であるかによって実行列あるいは複素行列になるが、さらに以下の例題のように  $A$  と  $B$  が帯行列であれば  $C^{(j)}$  も帯行列となる。

今回の実験に用いたプログラムは、簡単化のために係数が実あるいは複素の帯対称行列である連立 1 次方程式を帯の内部は実際には非常に疎であるのに密であるかのようにして扱い、領域分割法やブロック化手法などは用いないで、古典的な BLAS-2 レベルの帯用の改訂コレスキ法で解いている。それを Fortran90 で記述したものに OpenMP の指示を挿入している。これは現段階での研究の主な目的は、フィルタ対角化法の手法の数値的な性質の検証であり、まだ使用する計算機システムの性能を高度に絞り出す段階ではないからである。今回の各実験結果に対して掲げた経過時間の測定値はそのような性格のものである。各種の計算機システム上で帯対称行列の連立 1 次方程式を効率良く解く方法の研究はまた別の課題になる。

連立 1 次方程式  $C^{(j)}Y^{(j)} = BX$  を解くには、各  $j$  について共通の右辺ベクトルの組  $BX$  を作るために、まず  $X$  に帯行列  $B$  を乗じる処理  $Z \leftarrow BX$  を行う。そうして対称帯行列用の改訂コレスキ法による行列分解  $C^{(j)} \Rightarrow L^{(j)}D^{(j)}(L^{(j)})^T$  の処理と、その分解結果を用いて前進消去と後退代入を行なって  $Y^{(j)}$  を求める処理  $Y^{(j)} \leftarrow (L^{(j)})^{-T} \{(D^{(j)})^{-1}(L^{(j)})^{-1}Z\}$  が要る。複素版の改訂コレスキ法は通常の実数版の算法中の数値と四則演算をそのまま複素数のものに置き換えることで得られる。改訂コレスキ分解の結果は残しておいて繰り返し用いる。

#### 注 1

シフト  $\rho_j$  が実数で  $C^{(j)}$  が実対称正定値である場合には、ピボット選択をしなくても (改訂) コレスキ法の計算は数値安定であるが、シフト  $\rho_j$  が複素数で  $C^{(j)}$  が複素対称である場合にはそのような保証は一般には無い。そのため  $C^{(j)}$  が複素対称である場合には数値安定性を保証するために行列の対称性を利用することをやめて、片側ピボット選択を入れた帯行列用の LU 分解法を利用できる。しか



しそのようにすると  $LU$  分解の記憶量はピボット選択なしの (改訂) コレスキ法を用いる場合に比べて (行選択により上帯幅は 2 倍にまで広がる可能性がある) 3 倍になり、計算に用いる演算量もかなり増えるので、実験例では係数行列  $C^{(j)}$  が複素対称の場合にはピボット選択なしの帯行列用の (複素版の) 改訂コレスキ法を用いている。

## 注 2

今回の実験ではまず  $j$  について順次に行列  $C^{(j)}$  の (改訂) コレスキ分解を作って保持しておく。それから  $j$  について順次に連立 1 次方程式  $C^{(j)}Y^{(j)} = Z$  を解くには既に作っておいた行列  $C^{(j)}$  の分解結果を利用して前進消去と後退代入を行うことで解  $Y^{(j)}$  を求める。前進消去と後退代入の処理の中では、 $m$  個のベクトルの組をまとめて計算することにより、マルチコア計算機の主記憶 (共有メモリ) に置かれた帯行列  $C^{(j)}$  の分解で得た帯行列  $L^{(j)}$  に対する記憶参照の量を節約している。

### A.2.3 例題の一般固有値問題について

実験の例題として用いた行列の実対称定値一般固有値問題  $A\mathbf{v} = \lambda B\mathbf{v}$  は、一辺の長さが  $\pi$  で辺が座標軸に沿った 3 次元立方体  $[0, \pi] \times [0, \pi] \times [0, \pi]$  の内部を領域として、その表面に於いて零-ディリクレ境界条件を課したときの (符号を逆にした) 3 次元ラプラシアン  $-\nabla^2$  の固有値問題、それを有限要素法により離散化近似して得られるものである。

用いた要素分割は立方体領域の各辺方向をそれぞれ  $N_1 + 1, N_2 + 1, N_3 + 1$  の等間隔の小区間に分割したもので、要素内での展開基底関数には各辺方向の 3 重線形関数を用いた。この有限要素法の離散化で得られる行列  $A$  と  $B$  の次数は  $N = N_1 N_2 N_3$  であり、( $N_1 \leq N_2 \leq N_3$  であるとして) 行列の帯幅がなるべく小さくなるように基底関数に適切に番号を付けると、各行列の (対角を含まない) 半帯幅 (下帯幅) は  $w_L = 1 + N_1 + N_1 N_2$  にできる。

今の場合に、辺に沿った 3 方向の 1 次元 FEM の展開基底である区分線形関数はそれぞれ  $N_1$  個、 $N_2$  個、 $N_3$  個あり、3 方向の区分線形関数に対してその頂上の位置の増加順につけた順番をそれぞれ  $i_1, i_2, i_3$  とする ( $1 \leq i_k \leq N_k, k = 1, 2, 3$ )。3 次元 FEM の展開基底である 3 重線形関数は各方向の展開基底の積なので 3 重添字 ( $i_1, i_2, i_3$ ) により指定できる。3 重線形関数の 3 重添字に対しては  $i_1 + N_1(i_2 - 1) + N_1 N_2(i_3 - 1)$  の値により順番を付けて、それに基づいて 3 次元 FEM の係数行列  $A$  や  $B$  を組み立てている。

そうして、固有値  $\lambda$  が区間  $[a, b]$  に含まれる固有対の近似をフィルタ対角化法を適用して求めた。このテスト例題の一般固有値問題の固有値は簡単な数式で表わせるので、式の計算により厳密な値を容易に求めることができる。またそのことを用いて、固有値の厳密値を列挙して値順に並べることで、固有値が区間  $[a, b]$  にある固有対の正しい数も求められる。

#### A.2.3.1 例題の固有値の厳密値を与える式

1 次元の区間  $[0, \pi]$  における零-Dirichlet 境界条件の Laplacian (但し符号を逆にした  $-\nabla^2$ ) を有限要素法で離

散化して得られる「行列の一般固有値問題」の固有値の全体は、区間  $[0, \pi]$  を  $N + 1$  等分して得られる幅  $h = \pi / (N + 1)$  の小区間による分割の区分線形関数を基底とする場合には、 $\phi_k \equiv hk$  とおくと以下の式 (A.33) で与えられる。

$$E(N, k) = \frac{6k^2 \{\sin(\phi_k) / \phi_k\}^2}{(1 + \cos \phi_k)(2 + \cos \phi_k)}, \quad k = 1, 2, \dots, N. \quad (\text{A.33})$$

そうして 3 次元立方体領域  $[0, \pi] \times [0, \pi] \times [0, \pi]$  の各座標軸方向の区間をそれぞれ  $N_1 + 1, N_2 + 1, N_3 + 1$  に等分割して各方向の 1 次元の基底の直積型の基底関数 (3 重線形関数) を用いた場合の有限要素法では、 $k_i = 1, 2, \dots, N_i$  とするとき  $(k_1, k_2, k_3)$  で識別される 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。

### A.2.4 フィルタ対角化法の概要

フィルタ対角化法では、まず最初に  $m$  個のランダムなベクトルの組 ( $N \times m$  行列) を乱数で生成して、それらを  $B$ -正規直交化して  $m$  個のベクトルの組  $X$  を作る。そうして区間  $[a, b]$  を通過域とするフィルタを  $X$  に作用させて  $m$  個のベクトルの組  $Y \leftarrow \mathcal{F}X$  を作る。

この  $X$  と  $Y$  をもとにして、使用したフィルタの伝達特性も考慮に入れて、元の一般固有値問題のある「不変部分空間」を近似する空間の基底を  $Y$  の列の適切な線形結合として構成する [17]。その「不変部分空間」は、区間  $[a, b]$  のある適切な近傍に入る固有値を持つ固有ベクトルの全体で張られるものである。その「不変部分空間」の近似空間の基底に Rayleigh-Ritz 法を適用して得られた Ritz 対を元の一般固有値問題の近似対として採用する。得られた近似対については (近似対に改良を加える場合はその後) に、近似固有値を検査して指定区間  $[a, b]$  にないものは棄却することができる。

### A.2.5 近似対の品質の評価方法

近似対  $(\lambda, \mathbf{v})$  に対する相対残差  $\Theta$  の定義を式 (A.34) とする。ベクトルのノルム  $\|\cdot\|$  には 2-ノルムを採用した。

$$\Theta \equiv \frac{\|A\mathbf{v} - \lambda B\mathbf{v}\|}{\|\lambda B\mathbf{v}\|}. \quad (\text{A.34})$$

この  $\Theta$  の値はベクトル  $\mathbf{v}$  の規格化には依らないし、共通の非零の値で  $A$  と  $B$  をスケールしても変わらない。そうして幾何学的には、 $N$  次元ユークリッド空間内の 2 つのベクトル  $A\mathbf{v}$  と  $\lambda B\mathbf{v}$  の挟む角の大きさを  $\phi$  とするとき、 $\sin \phi \leq \Theta$  であることが示せる。

複数の近似対に対する相対残差は、近似ベクトルを集めて並べた行列  $V$  から ( $A$  や  $B$  の帯性や疎性を利用して) 行列の積  $AV$  と  $BV$  を計算すれば  $A$  と  $B$  の記憶全体に対する参照がそれぞれ 1 回ずつになり、効率良く求められる。この相対残差  $\Theta$  は、計算量や記憶の参照量が少ないので、近似対の品質の評価法として適用が容易である。

### A.3 実験例

#### A.3.1 例題 1 (中間固有対を求めた例, その 1)

例題 1 は有限要素法 (FEM) により要素分割を変えながら、離散化された立方体のラプラス固有値問題の固有値が区間  $[a, b] = [1020, 1025]$  にある中間固有対を求める問題である。使用したフィルタは、4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) に対してはいずれも伝達関数の 3 つの形状に対する条件として  $g_p = 0.1$ ,  $g_s \leq 10^{-16}$ ,  $\xi = 1.1$  を課した。合成用の有理関数の次数  $\ell$  は形状の要求を満たせる最小の偶数を選んだ。ただし中間固有対を求める場合には  $\ell$  を偶数にする必要がある。なおチェビシェフ多項式の次数  $n$  は 50 以下に制限している。

この例題 1 で各合成の場合に実現されたフィルタのパラメタを表 A.2 に示す。

表 A.2 例題 1: 使用したフィルタの構成パラメタ

フィルタ	$\ell$	$n$	実現した $g_s$ の値
B-合成	24	36	9.18E-17
C-合成	8	48	9.57E-17
I-合成	8	48	9.57E-17
E-合成	6	10	1.45E-17

また、各合成によるフィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$  を図 A.16 に示す。

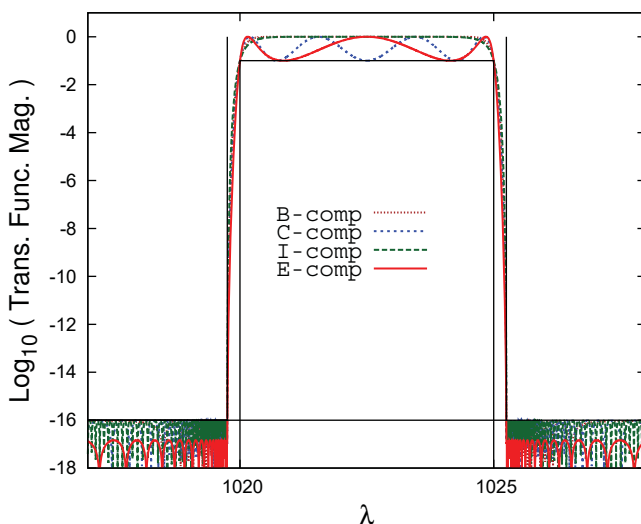


図 A.16 例題 1: フィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$

#### A.3.1.1 例題 1-1

これは FEM の要素分割を  $(N_1, N_2, N_3) = (20, 30, 40)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 24,000$ ,  $w_L = 621$  である。固有値が区間  $[1020, 1025]$  にある対の真の数は 64 個である。フィルタを適用するベクトルの数は  $m = 100$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.3 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.3 例題 1-1: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	1.20E-13	85.9
C-合成 ( $\ell=8, n=48$ )	2.33E-13	39.0
I-合成 ( $\ell=8, n=48$ )	2.47E-13	38.4
E-合成 ( $\ell=6, n=10$ )	1.23E-13	8.7

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.17 に示す。

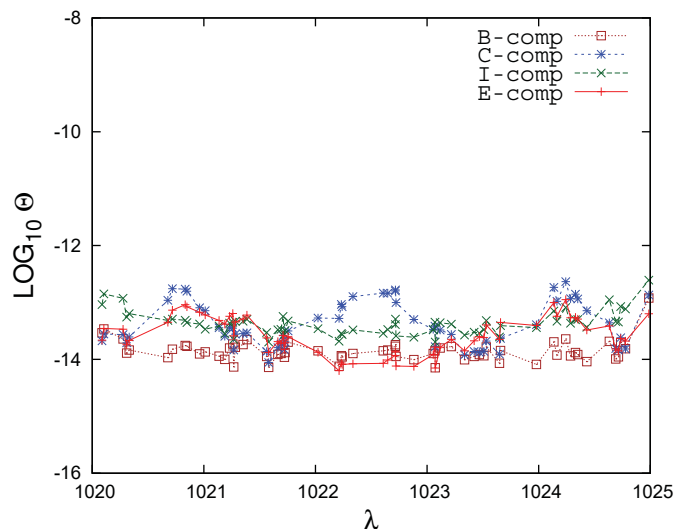


図 A.17 例題 1-1: 各合成による近似対の相対残差  $\theta$  の分布

### A.3.1.2 例題 1-2

これは FEM の要素分割を  $(N_1, N_2, N_3) = (30, 40, 50)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 60,000$ ,  $w_L = 1,231$  である。固有値が区間  $[1020, 1025]$  にある対の真の数は 77 個である。フィルタを適用するベクトルの数は  $m = 100$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.4 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.4 例題 1-2: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	3.43E-13	346.0
C-合成 ( $\ell= 8, n=48$ )	4.02E-13	155.8
I-合成 ( $\ell= 8, n=48$ )	6.57E-13	157.8
E-合成 ( $\ell= 6, n=10$ )	3.21E-13	35.1

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.18 に示す。

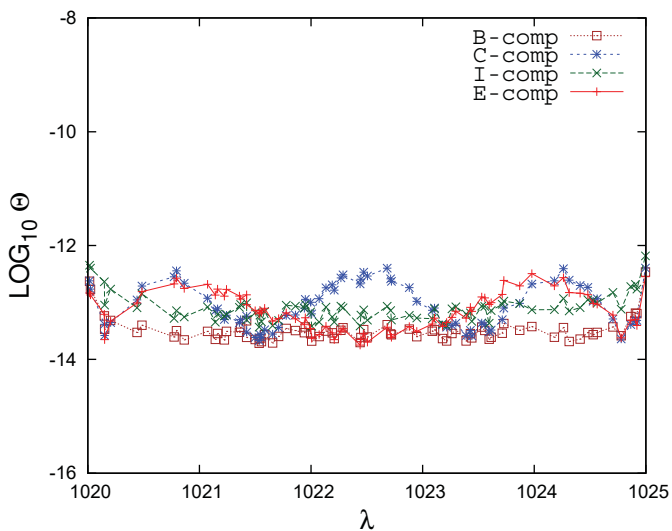


図 A.18 例題 1-2: 各合成による近似対の相対残差  $\theta$  の分布

### A.3.1.3 例題 1-3

これは FEM の要素分割を  $(N_1, N_2, N_3) = (40, 50, 60)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 120,000$ ,  $w_L = 2,041$  である。固有値が区間  $[1020, 1025]$  にある対の真の数は 79 個である。フィルタを適用するベクトルの数は  $m = 100$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.5 に示す。表からは E-合成の場合の経過時間が最も短かったことがわかる。

表 A.5 例題 1-3: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	4.09E-13	1,042.2
C-合成 ( $\ell= 8, n=48$ )	5.16E-13	465.0
I-合成 ( $\ell= 8, n=48$ )	7.42E-13	462.0
E-合成 ( $\ell= 6, n=10$ )	3.86E-13	101.0

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.19 に示す。

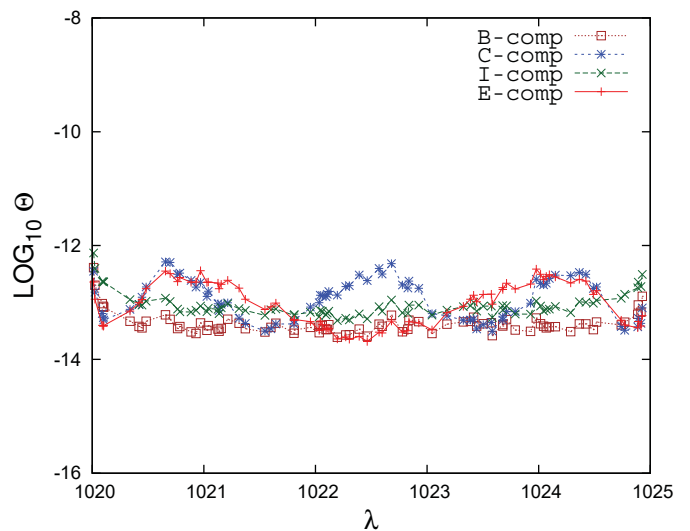


図 A.19 例題 1-3: 各合成による近似対の相対残差  $\theta$  の分布



**A.3.1.4 例題 1-4**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (50, 60, 70)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 210,000$ ,  $w_L = 3,051$  である。固有値が区間  $[1020, 1025]$  にある対の真の数は 73 個である。フィルタを適用するベクトルの数は  $m = 100$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.6 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.6 例題 1-4: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	1.90E-13	3,126.7
C-合成 ( $\ell= 8, n=48$ )	4.72E-13	1,333.5
I-合成 ( $\ell= 8, n=48$ )	4.97E-13	1,324.7
E-合成 ( $\ell= 6, n=10$ )	4.48E-13	393.4

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.20 に示す。

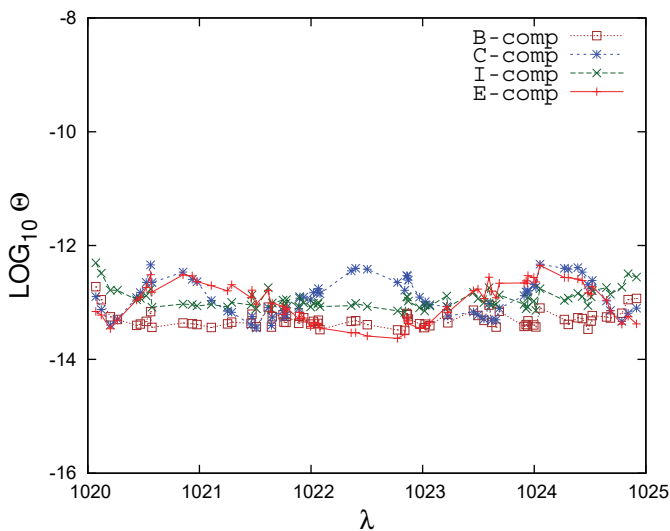


図 A.20 例題 1-4: 各合成による近似対の相対残差  $\theta$  の分布

**A.3.1.5 例題 1-5**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (60, 70, 80)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 336,000$ ,  $w_L = 4,261$  である。固有値が区間  $[1020, 1025]$  にある対の真の数は 87 個である。フィルタを適用するベクトルの数は  $m = 110$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.7 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。B-合成については、システムで利用可能な記憶量の制限値 168GiB を越えたため計算を実施できなかった。

表 A.7 例題 1-5: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	(記憶量の制限を超過)	
C-合成 ( $\ell= 8, n=48$ )	6.83E-13	3,643.4
I-合成 ( $\ell= 8, n=48$ )	3.38E-12	3,612.0
E-合成 ( $\ell= 6, n=10$ )	2.58E-12	1,405.6

また、(計算が実施できなかった B-合成を除く) 各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.21 に示す。

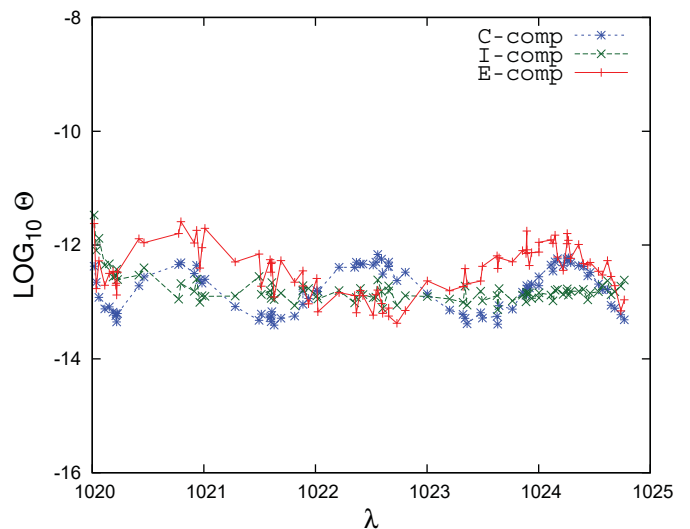


図 A.21 例題 1-5: 各合成による近似対の相対残差  $\theta$  の分布

### A.3.2 例題 2 (中間固有対を求めた例, その 2)

例題 2 は有限要素法 (FEM) により要素分割を変えながら, 離散化された立方体のラプラス固有値問題の固有値が区間  $[a, b] = [100, 200]$  にある中間固有対を求める問題である. 使用したフィルタは, 4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) のどれも伝達関数の 3 つの形状として  $g_p \geq 0.1$ ,  $g_s = 10^{-16}$ ,  $\xi = 1.1$  を要求した. 合成用の有理関数の次数  $\ell$  は形状の要求を満たせる最小の偶数を選んだ. 中間固有対を求める場合には  $\ell$  は偶数であることが要求される. なおチェビシェフ多項式の次数  $n$  は 50 以下に制限している.

この例題 2 で各合成の場合に実現されたフィルタのパラメタを表 A.8 に示す.

表 A.8 例題 2: 使用したフィルタの構成パラメタ

フィルタ	$\ell$	$n$	実現した $g_p$ の値
B-合成	24	36	0.1007
C-合成	8	48	0.1003
I-合成	8	48	0.1003
E-合成	6	10	0.1444

また, 各合成の場合のフィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$  を図 A.22 に示す.

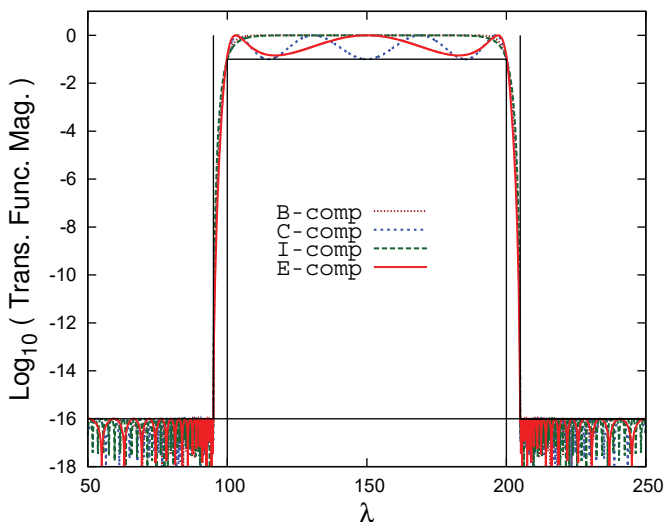


図 A.22 例題 2: フィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$

### A.3.2.1 例題 2-1

これは FEM の要素分割を  $(N_1, N_2, N_3) = (20, 30, 40)$  とした場合で, 固有値問題の行列の次数と下帯幅は  $N = 24,000$ ,  $w_L = 621$  である. 固有値が区間  $[100, 200]$  にある対の真の数は 684 個である. フィルタを適用するベクトルの数は  $m = 800$  とした. 各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.9 に示す. 表からは E-合成の場合の経過時間が最も短いことがわかる.

表 A.9 例題 2-1: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	9.07E-12	248.1
C-合成 ( $\ell= 8, n=48$ )	5.05E-12	119.1
I-合成 ( $\ell= 8, n=48$ )	1.51E-11	120.8
E-合成 ( $\ell= 6, n=10$ )	6.67E-12	34.5

また, 各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.23 に示す.

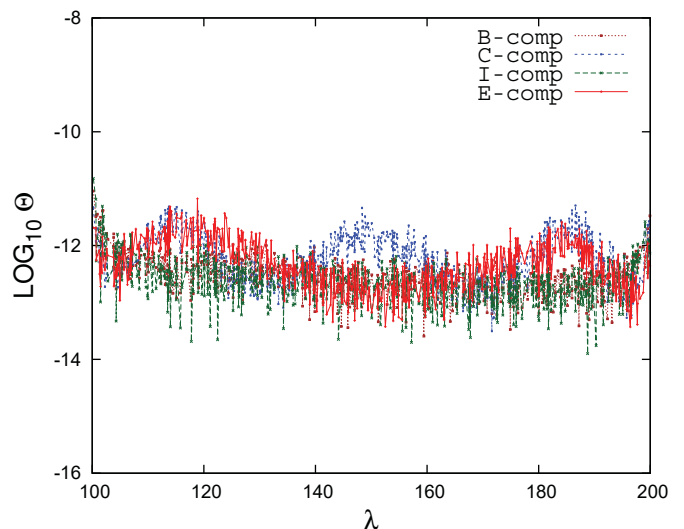


図 A.23 例題 2-1: 各合成による近似対の相対残差  $\theta$  の分布

### A.3.2.2 例題 2-2

これは FEM の要素分割を  $(N_1, N_2, N_3) = (30, 40, 50)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 60,000$ ,  $w_L = 1,231$  である。固有値が区間  $[100, 200]$  にある対の真の数は 750 個である。フィルタを適用するベクトルの数は  $m = 900$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.10 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.10 例題 2-2: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	8.62E-12	1,257.7
C-合成 ( $\ell= 8, n=48$ )	4.04E-12	574.0
I-合成 ( $\ell= 8, n=48$ )	2.72E-12	578.4
E-合成 ( $\ell= 6, n=10$ )	1.30E-11	120.9

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.24 に示す。

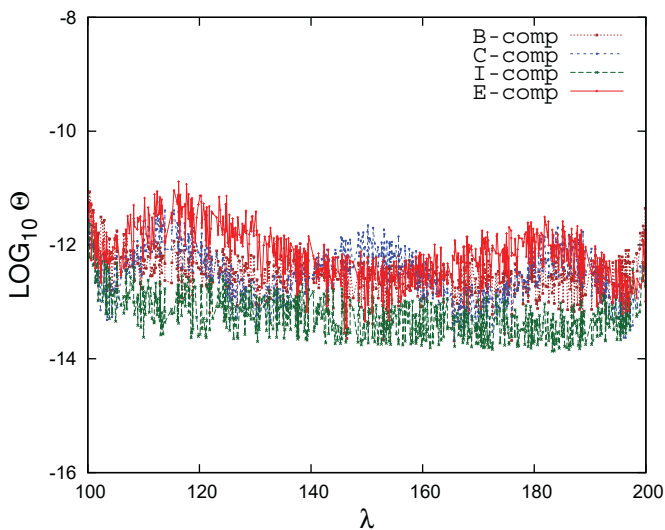


図 A.24 例題 2-2: 各合成による近似対の相対残差  $\Theta$  の分布

### A.3.2.3 例題 2-3

これは FEM の要素分割を  $(N_1, N_2, N_3) = (40, 50, 60)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 120,000$ ,  $w_L = 2,041$  である。固有値が区間  $[100, 200]$  にある対の真の数は 788 個である。フィルタを適用するベクトルの数は  $m = 1,000$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.11 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.11 例題 2-3: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	7.30E-12	4,285.1
C-合成 ( $\ell= 8, n=48$ )	9.14E-12	1,960.1
I-合成 ( $\ell= 8, n=48$ )	6.27E-12	1,964.3
E-合成 ( $\ell= 6, n=10$ )	3.31E-12	374.4

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.25 に示す。

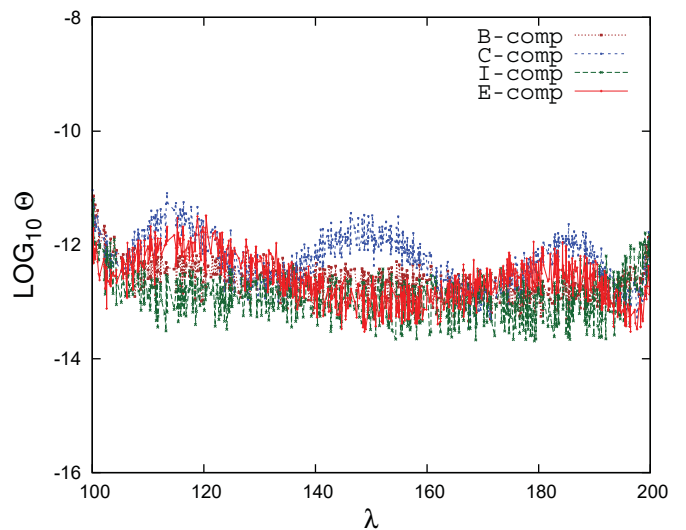


図 A.25 例題 2-3: 各合成による近似対の相対残差  $\Theta$  の分布



### A.3.2.4 例題 2-4

これは FEM の要素分割を  $(N_1, N_2, N_3) = (50, 60, 70)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 210,000$ ,  $w_L = 3,051$  である。固有値が区間  $[100, 200]$  にある対の真の数は 801 個である。フィルタを適用するベクトルの数は  $m = 1,000$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.12 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.12 例題 2-4: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	4.55E-12	11,350.5
C-合成 ( $\ell=8, n=48$ )	1.84E-12	5,204.4
I-合成 ( $\ell=8, n=48$ )	1.39E-12	5,229.2
E-合成 ( $\ell=6, n=10$ )	1.47E-11	1,059.7

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.26 に示す。

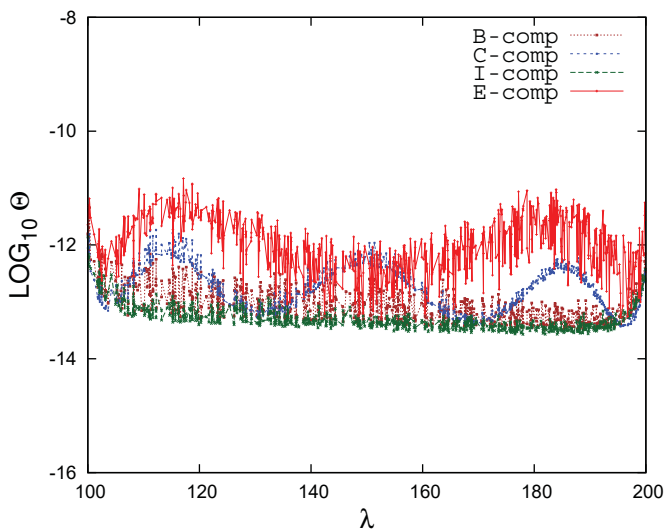


図 A.26 例題 2-4: 各合成による近似対の相対残差  $\Theta$  の分布

### A.3.2.5 例題 2-5

これは FEM の要素分割を  $(N_1, N_2, N_3) = (60, 70, 80)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 336,000$ ,  $w_L = 4,261$  である。固有値が区間  $[100, 200]$  にある対の真の数は 809 個である。フィルタを適用するベクトルの数は  $m = 1,000$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.13 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。なお、B-合成の場合はシステムで使用可能な記憶量の制限値 168GiB を越えたので計算を実施できなかった。

表 A.13 例題 2-5: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	(記憶量の制限を超過)	
C-合成 ( $\ell=8, n=48$ )	5.64E-12	14,968.1
I-合成 ( $\ell=8, n=48$ )	3.19E-12	14,763.0
E-合成 ( $\ell=6, n=10$ )	1.83E-11	3,204.6

また、(計算が実施できなかった B-合成の場合を除く) 各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.27 に示す。

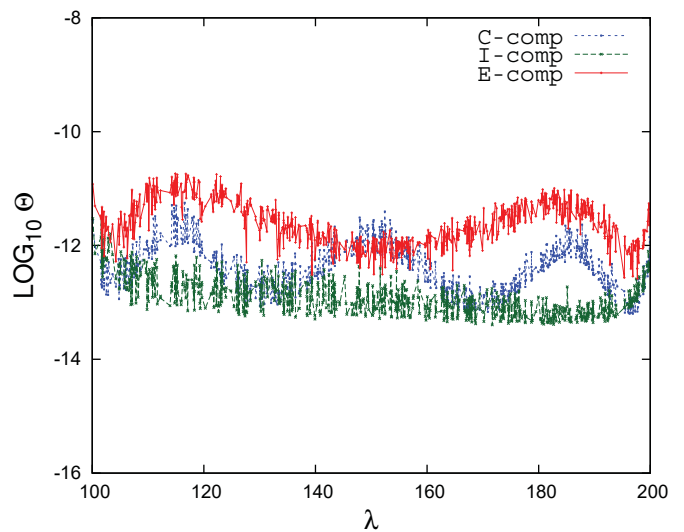


図 A.27 例題 2-5: 各合成による近似対の相対残差  $\Theta$  の分布

### A.3.3 例題 3 (中間固有対を求めた例, その 3)

例題 3 は有限要素法 (FEM) により要素分割を変えながら、離散化された立方体のラプラス固有値問題の固有値が区間  $[a, b] = [70, 80]$  にある中間固有対を求める問題である。使用したフィルタは、4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) に対してはいずれも伝達関数の 3 つの形状に対する条件として  $g_p = 0.1$ ,  $g_s \leq 10^{-16}$ ,  $\xi = 1.3$  を課した。合成用の有理関数の次数  $\ell$  は形状の要求を満たせる最小の偶数を選んだ。中間固有対を求める場合には  $\ell$  は偶数にしなければならない。なおチェビシェフ多項式の次数  $n$  は 50 以下に制限している。

この例題 3 で各合成の場合に実現されたフィルタのパラメタを表 A.14 に示す。

表 A.14 例題 3 : 使用したフィルタの構成パラメタ

フィルタ	$\ell$	$n$	実現した $g_s$ の値
B-合成	10	20	6.97E-17
C-合成	6	13	8.35E-17
I-合成	6	13	8.35E-17
E-合成	4	15	2.40E-17

また、各合成によるフィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$  を図 A.28 に示す。

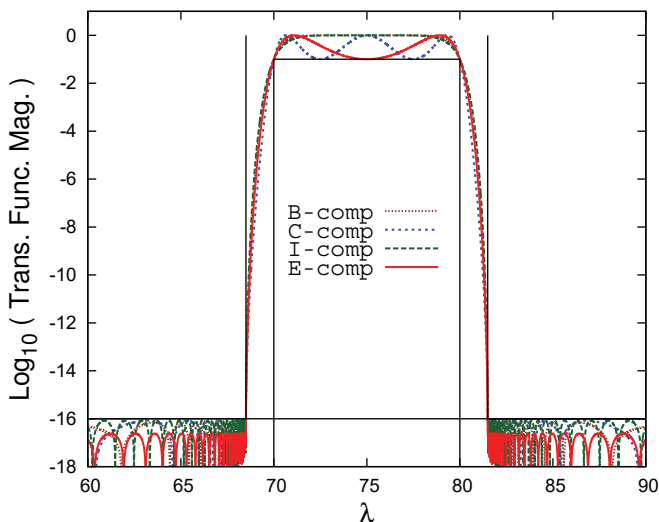


図 A.28 例題 3 : フィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$

### A.3.3.1 例題 3-1

これは FEM の要素分割を  $(N_1, N_2, N_3) = (20, 30, 40)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 24,000$ ,  $w_L = 621$  である。固有値が区間  $[70, 80]$  にある対の真の数は 55 個である。フィルタを適用するベクトルの数は  $m = 100$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.15 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.15 例題 3-1 : 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=10, n=20$ )	7.49E-14	22.8
C-合成 ( $\ell= 6, n=13$ )	1.03E-13	10.5
I-合成 ( $\ell= 6, n=13$ )	1.20E-13	10.4
E-合成 ( $\ell= 4, n=15$ )	6.69E-14	7.8

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.29 に示す。

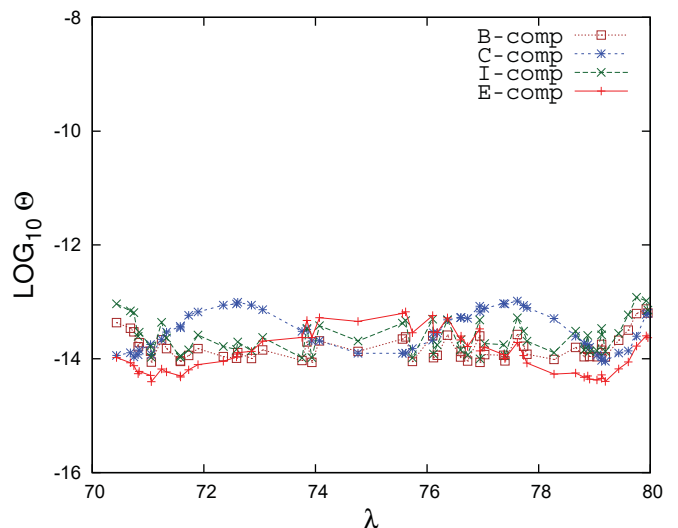


図 A.29 例題 3-1 : 各合成による近似対の相対残差  $\theta$  の分布

### A.3.3.2 例題 3-2

これは FEM の要素分割を  $(N_1, N_2, N_3) = (30, 40, 50)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 60,000$ ,  $w_L = 1,231$  である。固有値が区間  $[70, 80]$  にある対の真の数は 58 個である。フィルタを適用するベクトルの数は  $m = 100$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.16 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.16 例題 3-2: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=10, n=20$ )	1.00E-13	92.7
C-合成 ( $\ell= 6, n=13$ )	3.05E-13	41.0
I-合成 ( $\ell= 6, n=13$ )	1.28E-13	40.9
E-合成 ( $\ell= 4, n=15$ )	9.24E-14	31.6

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.30 に示す。

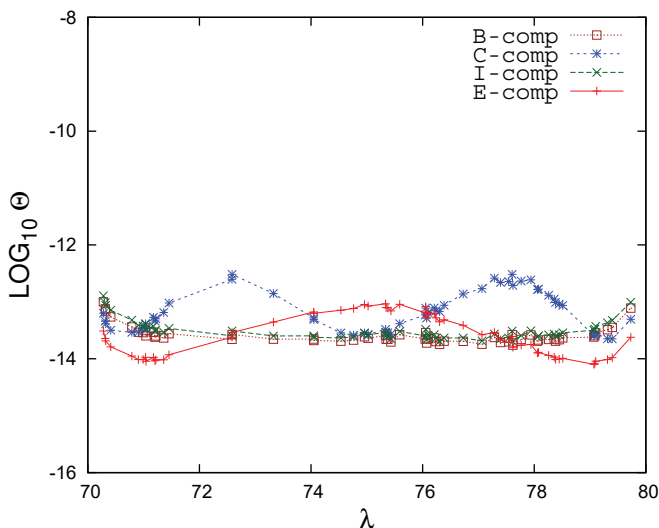


図 A.30 例題 3-2: 各合成による近似対の相対残差  $\Theta$  の分布

### A.3.3.3 例題 3-3

これは FEM の要素分割を  $(N_1, N_2, N_3) = (40, 50, 60)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 120,000$ ,  $w_L = 2,041$  である。固有値が区間  $[70, 80]$  にある対の真の数は 58 個である。フィルタを適用するベクトルの数は  $m = 100$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.17 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.17 例題 3-3: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=10, n=20$ )	2.16E-13	269.7
C-合成 ( $\ell= 6, n=13$ )	5.48E-13	120.3
I-合成 ( $\ell= 6, n=13$ )	2.93E-13	119.3
E-合成 ( $\ell= 4, n=15$ )	3.19E-13	92.6

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.31 に示す。

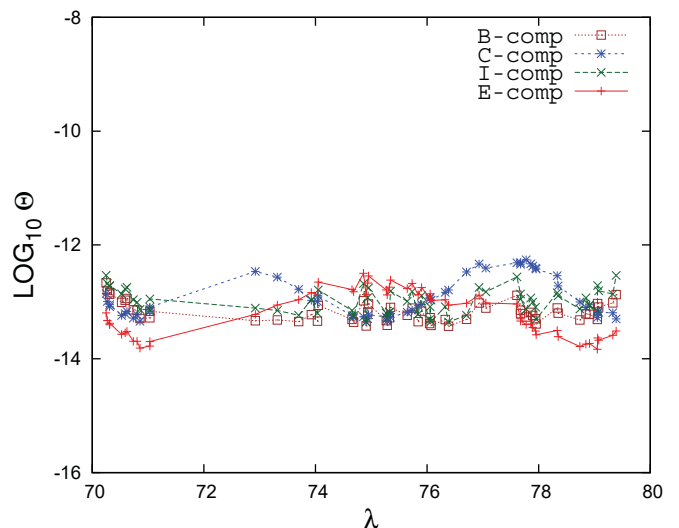


図 A.31 例題 3-3: 各合成による近似対の相対残差  $\Theta$  の分布



### A.3.3.4 例題 3-4

これは FEM の要素分割を  $(N_1, N_2, N_3) = (50, 60, 70)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 210,000$ ,  $w_L = 3,051$  である。固有値が区間  $[70, 80]$  にある対の真の数は 54 個である。フィルタを適用するベクトルの数は  $m = 100$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.18 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.18 例題 3-4: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=10, n=20$ )	3.09E-13	901.3
C-合成 ( $\ell= 6, n=13$ )	1.10E-12	438.0
I-合成 ( $\ell= 6, n=13$ )	4.13E-13	438.7
E-合成 ( $\ell= 4, n=15$ )	3.30E-13	321.4

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.32 に示す。

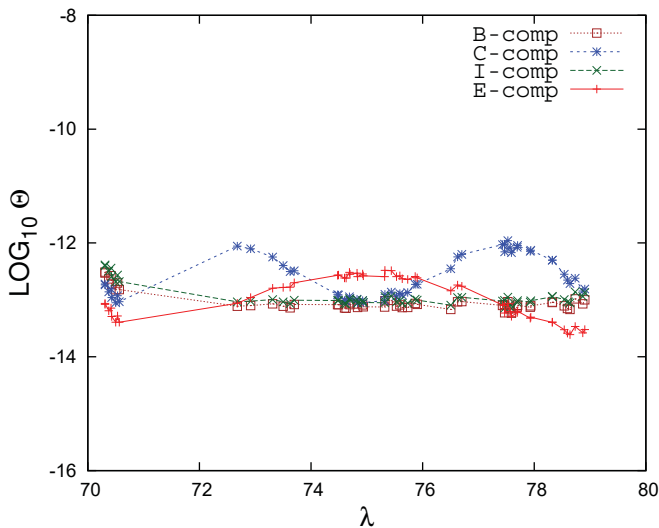


図 A.32 例題 3-4: 各合成による近似対の相対残差  $\Theta$  の分布

### A.3.3.5 例題 3-5

これは FEM の要素分割を  $(N_1, N_2, N_3) = (60, 70, 80)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 336,000$ ,  $w_L = 4,261$  である。固有値が区間  $[70, 80]$  にある対の真の数は 52 個である。フィルタを適用するベクトルの数は  $m = 110$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.19 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.19 例題 3-5: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=10, n=20$ )	8.08E-13	2,861.4
C-合成 ( $\ell= 6, n=13$ )	1.72E-13	1,490.6
I-合成 ( $\ell= 6, n=13$ )	1.77E-12	1,478.3
E-合成 ( $\ell= 4, n=15$ )	7.63E-13	1,047.8

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.33 に示す。

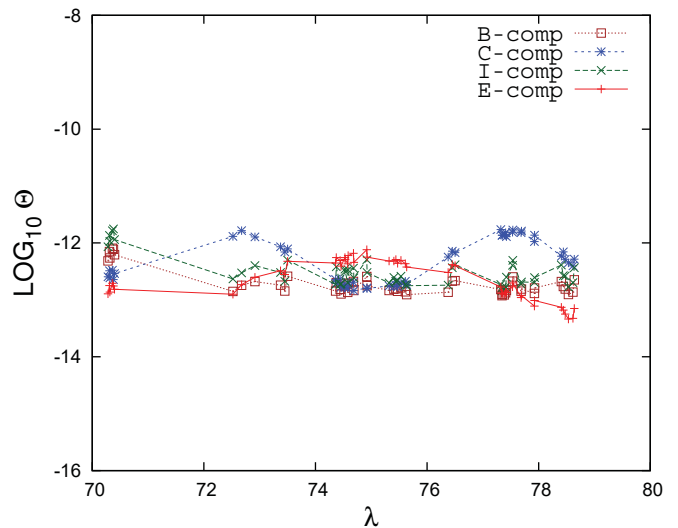


図 A.33 例題 3-5: 各合成による近似対の相対残差  $\Theta$  の分布

**A.3.3.6 例題 3-6**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (70, 80, 90)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 504,000$ ,  $w_L = 5,671$  である。固有値が区間  $[70, 80]$  にある対の真の数は 52 個である。フィルタを適用するベクトルの数は  $m = 100$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.20 に示す。E-合成以外の B-合成, C-合成, I-合成の場合については、システムが設定している使用可能な記憶量の制限値 168GiB を越えたので計算を実施できなかった。

表 A.20 例題 3-6: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=10, n=20$ )	(記憶量の制限を超過)	
C-合成 ( $\ell= 6, n=13$ )	(同上)	
I-合成 ( $\ell= 6, n=13$ )	(同上)	
E-合成 ( $\ell= 4, n=15$ )	2.13E-11	2,822.3

また、(計算が実施できた E-合成の場合だけについて) 合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.34 に示す。

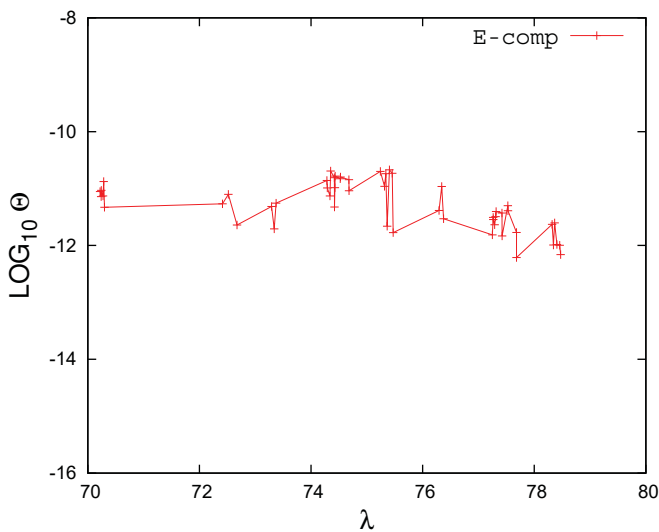


図 A.34 例題 3-6: 各合成による近似対の相対残差  $\theta$  の分布

**A.3.4 例題 4 (下端固有対を求めた例, その 1)**

例題 4 は有限要素法 (FEM) により要素分割を変えながら、離散化された立方体のラプラス固有値問題の固有値が区間  $[a, b] = [0, 30]$  にある下端固有対を求める問題である。使用したフィルタは、4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) のどれも伝達関数の 3 つの形状として  $g_p \geq 0.1$ ,  $g_s \leq 10^{-16}$ ,  $\xi = 1.1$  を要求した。合成用の有理関数の次数  $\ell$  は形状の要求を満たせる最小の整数を選んだ。なおチェビシェフ多項式の次数  $n$  は 50 以下に制限している。B-合成と I-合成は、どちらも次数  $\ell$  が偶数になったので、中間固有値を求めることもできるフィルタであり、固有値の区間  $[a, b]$  と規格化座標の区間  $t \in [-1, 1]$  を対応させた。C-合成と E-合成については、 $\ell$  の偶奇によらずに固有値の区間  $[a, b]$  と規格化座標の区間  $t \in [-1, 1]$  を対応させてよいので、そのようにした。

この例題 4 で各合成の場合に実現されたフィルタのパラメタを表 A.21 に示す。いまの場合には次数  $\ell$  が最小になるフィルタの構成を選ぶことにしているため、下端用のフィルタでは  $\ell$  が奇数でもよくて、E-合成では実験例 2 の場合の  $\ell = 6$  とは異なり  $\ell = 5$  になっていて、それに伴い  $n$  が 10 から 17 に増えている。

表 A.21 例題 4: 使用したフィルタの構成パラメタ

フィルタ	$\ell$	$n$	実現した $g_p$ の値
B-合成	24	36	0.1007
C-合成	8	48	0.1003
I-合成	8	48	0.1003
E-合成	5	17	0.1131

また、各合成の場合のフィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$  を図 A.35 に示す。

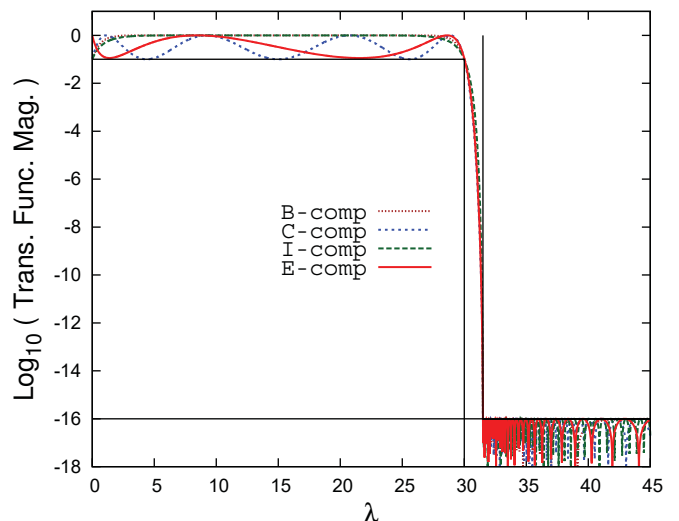


図 A.35 例題 4: フィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$

**A.3.4.1 例題 4-1**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (20, 30, 40)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 24,000$ ,  $w_L = 621$  である。固有値が区間  $[0, 30]$  にある対の真の数は 54 個である。フィルタを適用するベクトルの数は  $m = 80$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.22 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.22 例題 4-1: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	3.55E-13	84.0
C-合成 ( $\ell= 8, n=48$ )	1.82E-12	37.9
I-合成 ( $\ell= 8, n=48$ )	2.22E-13	38.0
E-合成 ( $\ell= 5, n=17$ )	1.44E-12	10.7

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.36 に示す。

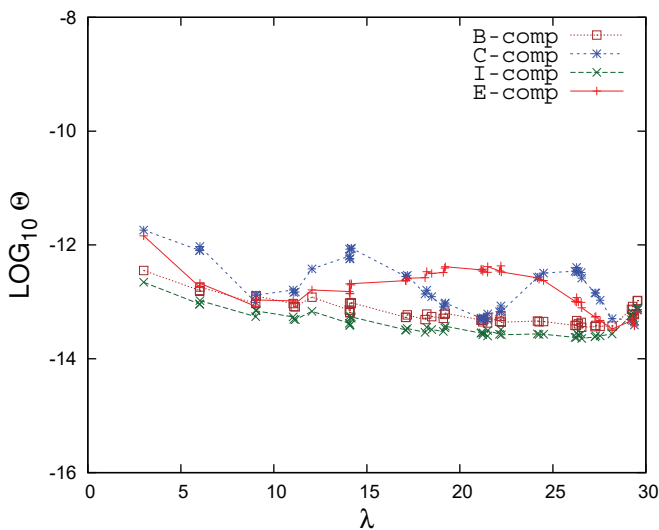


図 A.36 例題 4-1: 各合成による近似対の相対残差  $\theta$  の分布

**A.3.4.2 例題 4-2**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (30, 40, 50)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 60,000$ ,  $w_L = 1,231$  である。固有値が区間  $[0, 30]$  にある対の真の数は 54 個である。フィルタを適用するベクトルの数は  $m = 80$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.23 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.23 例題 4-2: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	8.43E-13	341.2
C-合成 ( $\ell= 8, n=48$ )	4.35E-12	153.5
I-合成 ( $\ell= 8, n=48$ )	5.60E-13	154.9
E-合成 ( $\ell= 5, n=17$ )	3.53E-12	43.4

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.37 に示す。

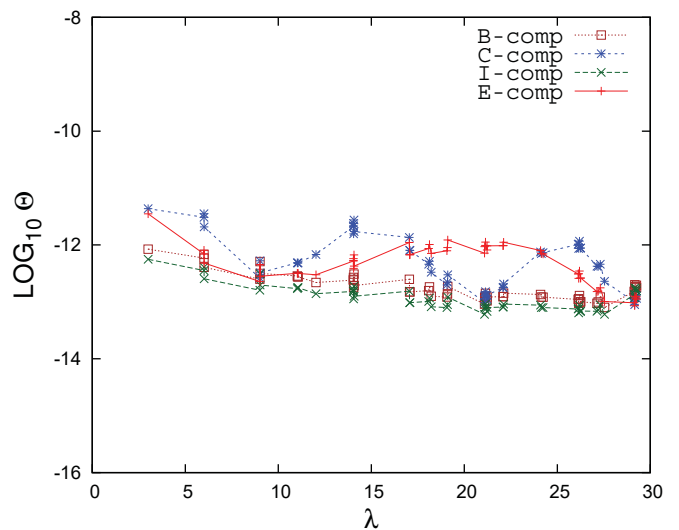


図 A.37 例題 4-2: 各合成による近似対の相対残差  $\theta$  の分布



### A.3.4.3 例題 4-3

これは FEM の要素分割を  $(N_1, N_2, N_3) = (40, 50, 60)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 120,000$ ,  $w_L = 2,041$  である。固有値が区間  $[0, 30]$  にある対の真の数は 54 個である。フィルタを適用するベクトルの数は  $m = 80$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.24 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.24 例題 4-3: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	1.77E-12	1,041.0
C-合成 ( $\ell= 8, n=48$ )	9.14E-12	456.3
I-合成 ( $\ell= 8, n=48$ )	1.45E-12	457.0
E-合成 ( $\ell= 5, n=17$ )	7.39E-12	125.0

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.38 に示す。

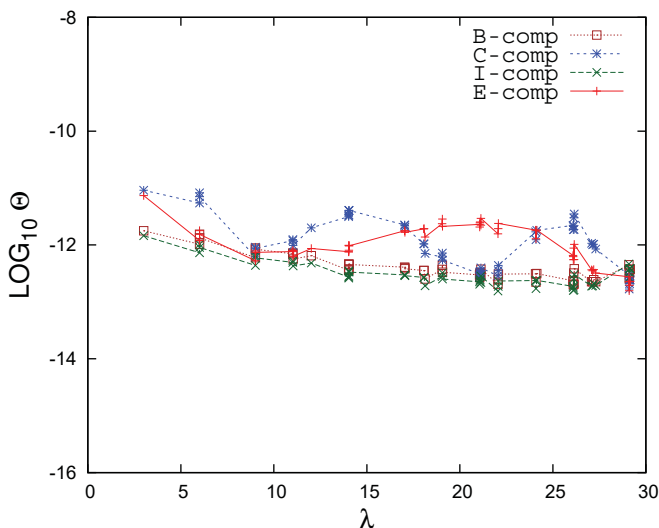


図 A.38 例題 4-3: 各合成による近似対の相対残差  $\Theta$  の分布

### A.3.4.4 例題 4-4

これは FEM の要素分割を  $(N_1, N_2, N_3) = (50, 60, 70)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 210,000$ ,  $w_L = 3,051$  である。固有値が区間  $[0, 30]$  にある対の真の数は 54 個である。フィルタを適用するベクトルの数は  $m = 80$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.25 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.25 例題 4-4: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	5.14E-12	3,103.2
C-合成 ( $\ell= 8, n=48$ )	2.75E-11	1,307.7
I-合成 ( $\ell= 8, n=48$ )	4.35E-12	1,321.3
E-合成 ( $\ell= 5, n=17$ )	2.19E-11	405.6

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.39 に示す。

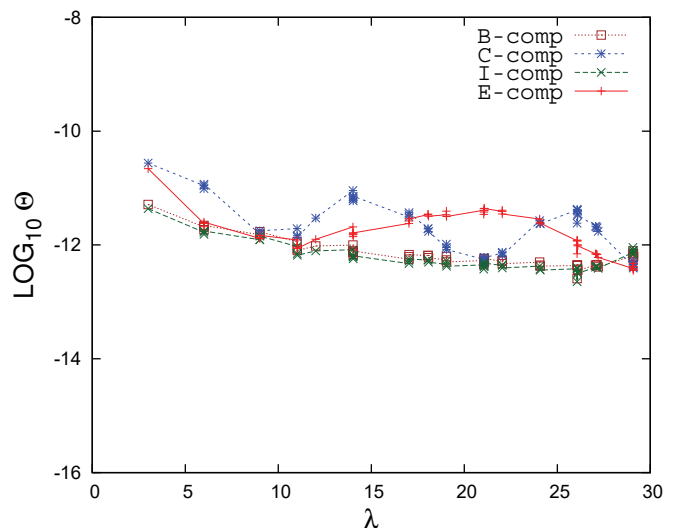


図 A.39 例題 4-4: 各合成による近似対の相対残差  $\Theta$  の分布

**A.3.4.5 例題 4-5**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (60, 70, 80)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 336,000$ ,  $w_L = 4,261$  である。固有値が区間  $[0, 30]$  にある対の真の数は 54 個である。フィルタを適用するベクトルの数は  $m = 80$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.26 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。B-合成の場合については、システムが設定している使用可能な記憶量の制限値 168GiB を越えたため計算を実施できなかった。

表 A.26 例題 4-5: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	(記憶量の制限を超過)	
C-合成 ( $\ell= 8, n=48$ )	4.88E-11	3,531.2
I-合成 ( $\ell= 8, n=48$ )	8.34E-12	3,550.1
E-合成 ( $\ell= 5, n=17$ )	3.86E-11	1,311.7

また、(計算が実施できなかった B-合成以外の) 各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.40 に示す。

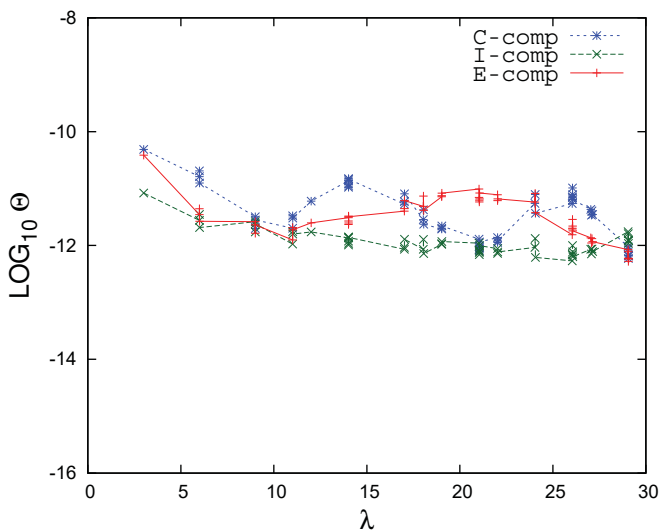


図 A.40 例題 4-5: 各合成による近似対の相対残差  $\theta$  の分布

**A.3.4.6 例題 4-6**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (70, 80, 90)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 504,000$ ,  $w_L = 5,671$  である。固有値が区間  $[0, 30]$  にある対の真の数は 54 個である。フィルタを適用するベクトルの数は  $m = 80$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.27 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。E-合成以外の B-合成, C-合成, I-合成については、システムが設定している使用可能な記憶量の制限値 168GiB を越えたため計算を実施できなかった。

表 A.27 例題 4-6: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=24, n=36$ )	(記憶量の制限を超過)	
C-合成 ( $\ell= 8, n=48$ )	(同上)	
I-合成 ( $\ell= 8, n=48$ )	(同上)	
E-合成 ( $\ell= 5, n=17$ )	2.89E-11	4,257.9

また、(計算が実施できた E-合成の場合だけについて) 合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.41 に示す。

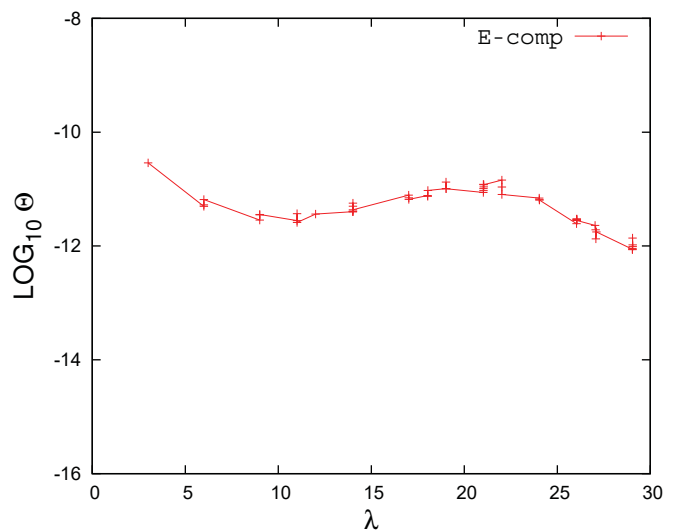


図 A.41 例題 4-6: 各合成による近似対の相対残差  $\theta$  の分布

### A.3.5 例題 5 (下端固有対を求めた例, その 2)

例題 5 は有限要素法 (FEM) により要素分割を変えながら, 離散化された立方体のラプラス固有値問題の固有値が区間  $[a, b] = [0, 20]$  にある中間固有対を求める問題である. 使用したフィルタは, 4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) のどれも伝達関数の伝達率の閾値に対する条件として  $g_p = 0.1$ ,  $g_s \leq 10^{-16}$  を課す. そうして B-合成と I-合成には次数  $l$  として奇数を許し,  $l$  が奇数になれば (この例の場合には実際そうである)  $\xi = 1.3$  と指定し, 固有値の区間  $[a, b]$  に対して規格化座標  $t$  の区間  $[0, 1]$  を対応させることにする. 他方で C-合成と E-合成では次数  $l$  が奇数の場合でも常に  $t$  の区間  $[-1, 1]$  を対応させることにして, そのかわりに固有値の座標  $\lambda$  で表したフィルタの伝達関数の阻止域の位置を B-合成や I-合成の場合のものと同揃えるために C-合成と E-合成では  $\xi = 1.6$  を指定した.

この例題 5 で採用された各合成の場合のフィルタのパラメタを表 A.28 に示す.

表 A.28 例題 5: 使用したフィルタの構成パラメタ

フィルタ	$l$	$\xi$	$n$	実現した $g_s$ の値
B-合成	9	1.3	30	7.48E-17
C-合成	4	1.6	17	3.51E-17
I-合成	5	1.3	26	6.04E-17
E-合成	3	1.6	24	6.71E-17

また, 各合成の場合のフィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$  を図 A.42 に示す.

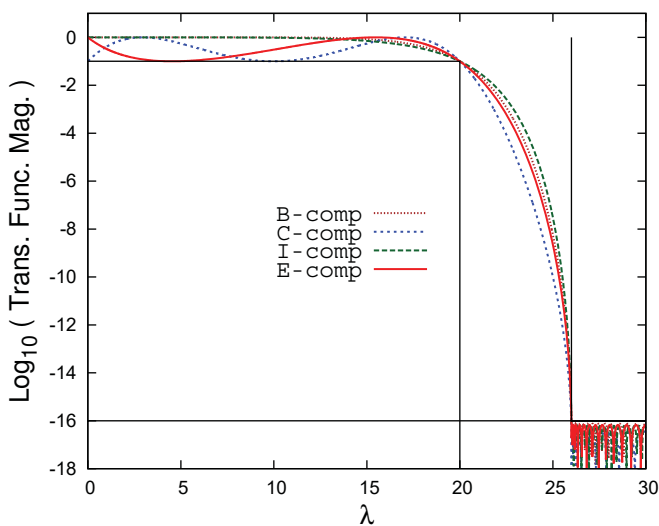


図 A.42 例題 5: フィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$

### A.3.5.1 例題 5-1

これは FEM の要素分割を  $(N_1, N_2, N_3) = (20, 30, 40)$  とした場合で, 固有値問題の行列の次数と下帯幅は  $N = 24,000$ ,  $w_L = 621$  である. 固有値が区間  $[0, 20]$  にある対の真の数は 26 個である. フィルタを適用するベクトルの数は  $m = 50$  とした. 各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.29 に示す. 表からは C-合成の場合の経過時間が最も短いことがわかるが, E-合成の場合も経過時間はそれほど違いはない.

表 A.29 例題 5-1: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	3.12E-13	25.9
C-合成 ( $l=4, n=17$ )	6.17E-13	7.5
I-合成 ( $l=5, n=26$ )	1.80E-13	13.4
E-合成 ( $l=3, n=24$ )	2.02E-13	8.4

また, 各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.43 に示す.

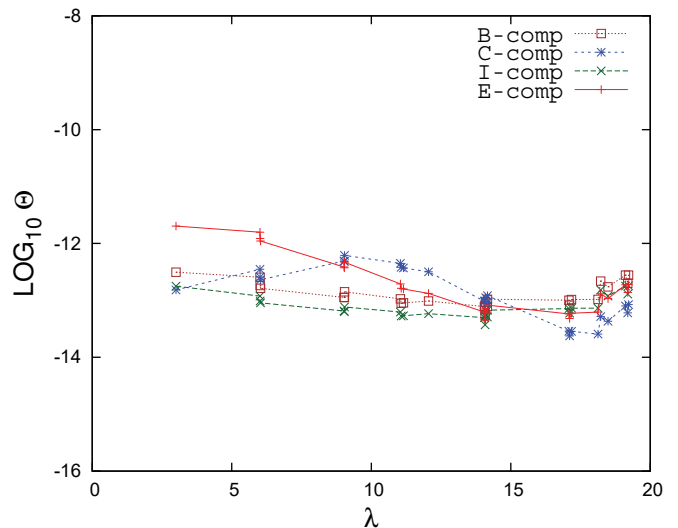


図 A.43 例題 5-1: 各合成による近似対の相対残差  $\theta$  の分布



**A.3.5.2 例題 5-2**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (30, 40, 50)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 60,000$ ,  $w_L = 1,231$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 50$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.30 に示す。表からは C-合成の場合の経過時間が最も短いことがわかるが、E-合成の場合も経過時間はそれほど違いはない。

表 A.30 例題 5-2: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=9, n=30$ )	9.67E-13	98.7
C-合成 ( $\ell=4, n=17$ )	1.76E-12	29.6
I-合成 ( $\ell=5, n=26$ )	6.88E-13	52.9
E-合成 ( $\ell=3, n=24$ )	6.62E-12	31.1

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.44 に示す。

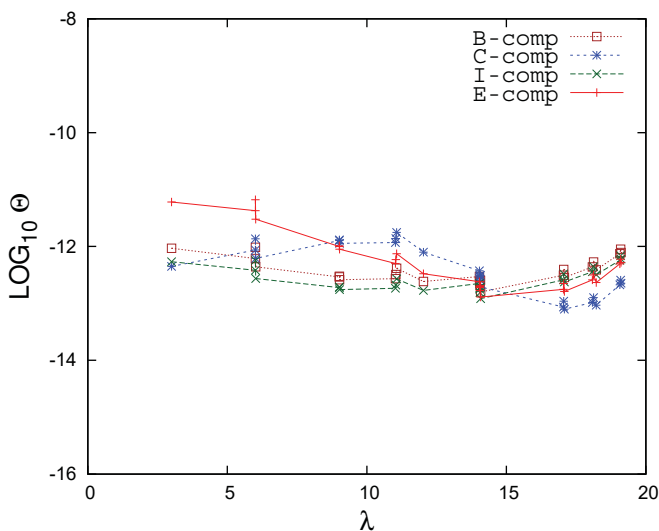


図 A.44 例題 5-2: 各合成による近似対の相対残差  $\theta$  の分布

**A.3.5.3 例題 5-3**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (40, 50, 60)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 120,000$ ,  $w_L = 2,041$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 50$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.31 に示す。表からは C-合成の場合の経過時間が最も短いことがわかるが、E-合成の場合も経過時間はそれほど違いはない。

表 A.31 例題 5-3: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $\ell=9, n=30$ )	2.59E-12	291.6
C-合成 ( $\ell=4, n=17$ )	3.73E-12	85.2
I-合成 ( $\ell=5, n=26$ )	1.63E-12	150.8
E-合成 ( $\ell=3, n=24$ )	1.68E-11	90.9

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.45 に示す。

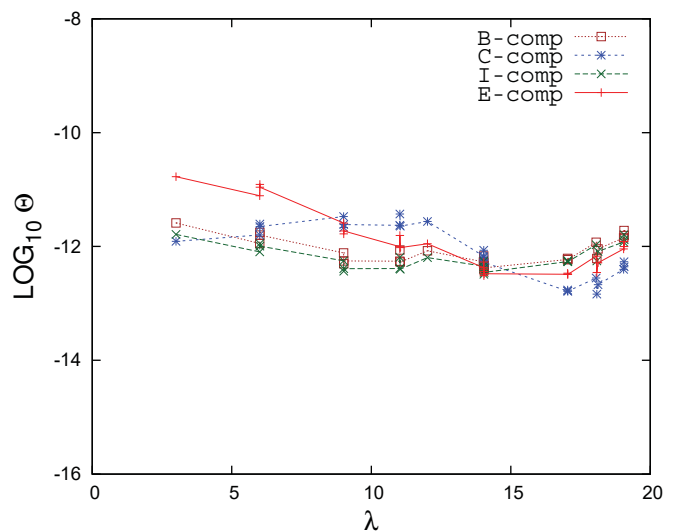


図 A.45 例題 5-3: 各合成による近似対の相対残差  $\theta$  の分布

**A.3.5.4 例題 5-4**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (50, 60, 70)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 210,000$ ,  $w_L = 3,051$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 50$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A-32 に示す。表からは E-合成の場合の経過時間が最も短いことがわかるが、C-合成の場合も経過時間はそれほど大きな違いはない。

表 A-32 例題 5-4: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	3.80E-12	891.7
C-合成 ( $l=4, n=17$ )	5.91E-12	299.3
I-合成 ( $l=5, n=26$ )	3.30E-12	457.7
E-合成 ( $l=3, n=24$ )	2.31E-11	264.3

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A-46 に示す。

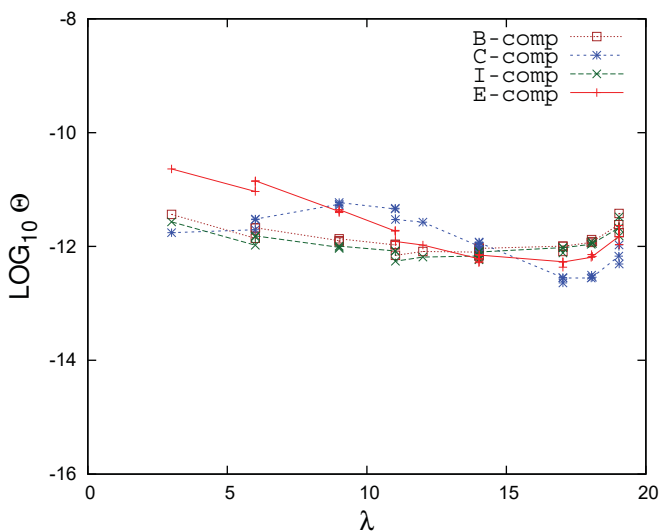


図 A-46 例題 5-4: 各合成による近似対の相対残差  $\theta$  の分布

**A.3.5.5 例題 5-5**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (60, 70, 80)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 336,000$ ,  $w_L = 4,261$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 50$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A-33 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A-33 例題 5-5: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	1.13E-11	2,714.8
C-合成 ( $l=4, n=17$ )	1.08E-11	1,003.8
I-合成 ( $l=5, n=26$ )	8.14E-12	1,421.0
E-合成 ( $l=3, n=24$ )	6.91E-11	822.0

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A-47 に示す。

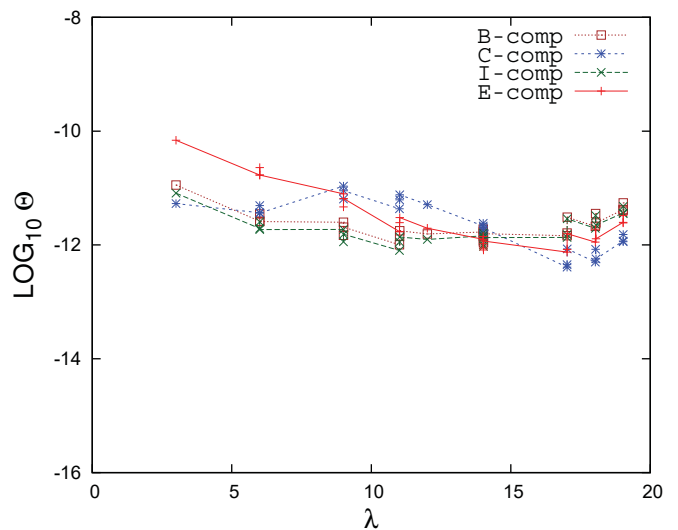


図 A-47 例題 5-5: 各合成による近似対の相対残差  $\theta$  の分布

A.3.5.6 例題 5-6

これは FEM の要素分割を  $(N_1, N_2, N_3) = (70, 80, 90)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 504,000$ ,  $w_L = 5,671$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 50$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.34 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。B-合成については、システムで設定している使用可能な記憶量の制限値 168GiB を越えたために計算を実施できなかった。

表 A.34 例題 5-6: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	(記憶量の制限を超過)	
C-合成 ( $l=4, n=17$ )	1.64E-11	2,722.9
I-合成 ( $l=5, n=26$ )	8.14E-12	4,859.3
E-合成 ( $l=3, n=24$ )	6.91E-11	2,243.0

また、(計算を実施できなかった B-合成を除く) 各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.48 に示す。

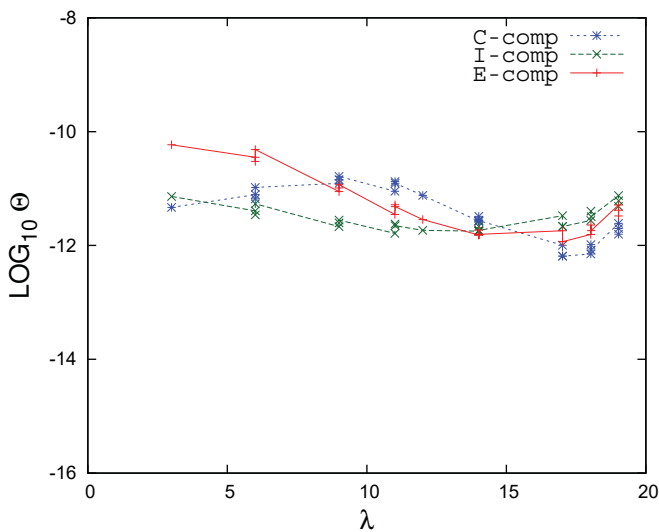


図 A.48 例題 5-6: 各合成による近似対の相対残差  $\Theta$  の分布

A.3.6 例題 6 (下端固有対を求めた例, その 3)

例題 6 は有限要素法 (FEM) により要素分割を変えながら、離散化された立方体のラプラス固有値問題で固有値が区間  $[a, b] = [0, 20]$  にある中間固有対を求める問題である。

使用したフィルタには、4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) のどれにも伝達関数の 3 つの形状に関する条件  $g_p = 0.1$ ,  $g_s \leq 10^{-16}$ ,  $\xi = 1.3$  を課した。

得られた伝達関数は、どれも規格化座標  $t$  の区間  $[-1, 1]$  において極がないので、4 種類の合成のどれについても固有値の区間  $[a, b]$  に対して規格化座標の区間  $[-1, 1]$  を対応させてフィルタを構成した。このようにすると、B-合成と I-合成では次数  $l$  が奇数の場合には (この例題 6 ではどちらも  $l$  は奇数である) 伝達関数は原点  $t = 0$  で高次の平坦性を持つが、 $t \in [-1, 1]$  で単調減少なので、 $t$  が下端  $t = -1$  に向かうとき伝達率は 1 を越える。そのため本来ならば  $l$  が奇数の場合には B-合成と I-合成についての伝達関数の定義式や閾値の設定の意味を見直すべきだが、通過域において伝達率の値が  $g_p$  以上という性質は保たれているので、伝達率の  $t = -1$  における最大値が 1 に比べてあまり大きくなければそのまま用いてもそれほど無駄な精度低下を起さずに適用できるから、そのような実験を試してみた。

この例題 6 で使用した各合成の場合のフィルタのパラメータを表 A.35 に示す。

表 A.35 例題 6: 使用したフィルタの構成パラメータ

フィルタ	$l$	$n$	実現した $g_s$ の値
B-合成	9	30	7.48E-17
C-合成	5	26	6.04E-17
I-合成	5	26	6.04E-17
E-合成	4	15	2.40E-17

また、各合成によるフィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$  のグラフを図 A.49 に示す (B-合成と I-合成の場合の伝達関数の  $t = -1$  に対応する  $\lambda = 0$  付近での最大値が 10 程度であることがわかる)。

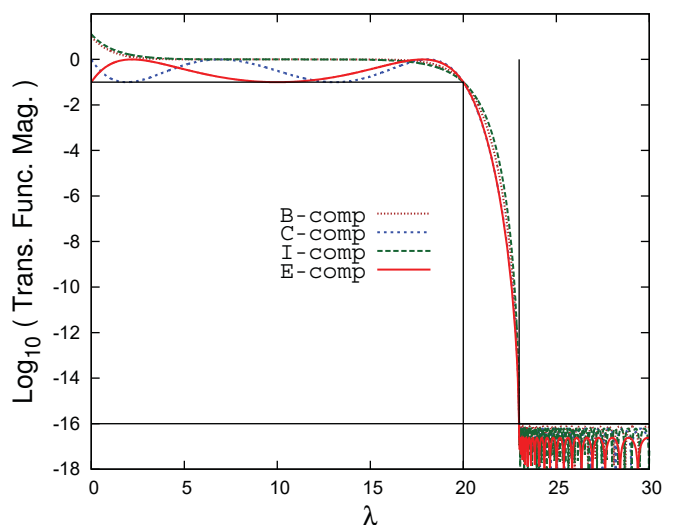


図 A.49 例題 6: フィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$

### A.3.6.1 例題 6-1

これは FEM の要素分割を  $(N_1, N_2, N_3) = (20, 30, 40)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 24,000$ ,  $w_L = 621$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 40$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.36 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.36 例題 6-1: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	1.00E-12	24.5
C-合成 ( $l=5, n=26$ )	5.95E-12	13.4
I-合成 ( $l=5, n=26$ )	3.50E-13	13.0
E-合成 ( $l=4, n=15$ )	6.22E-13	6.8

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.50 に示す。

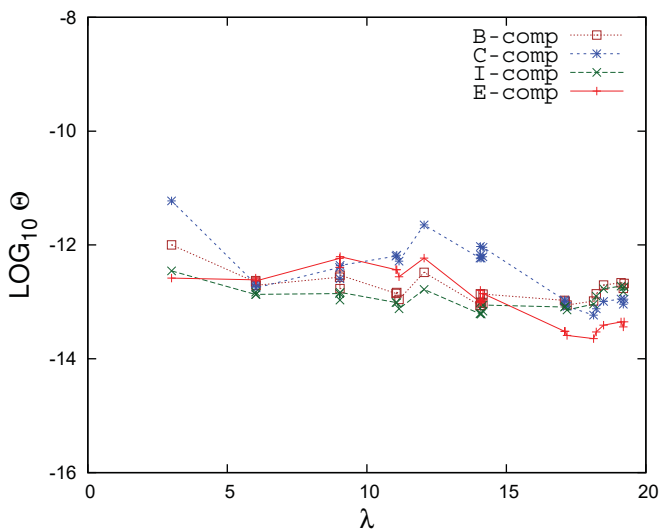


図 A.50 例題 6-1: 各合成による近似対の相対残差  $\Theta$  の分布

### A.3.6.2 例題 6-2

これは FEM の要素分割を  $(N_1, N_2, N_3) = (30, 40, 50)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 60,000$ ,  $w_L = 1,231$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 40$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.37 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.37 例題 6-2: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	2.32E-12	97.5
C-合成 ( $l=5, n=26$ )	7.93E-12	50.0
I-合成 ( $l=5, n=26$ )	1.35E-12	51.2
E-合成 ( $l=4, n=15$ )	2.76E-12	26.1

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.51 に示す。

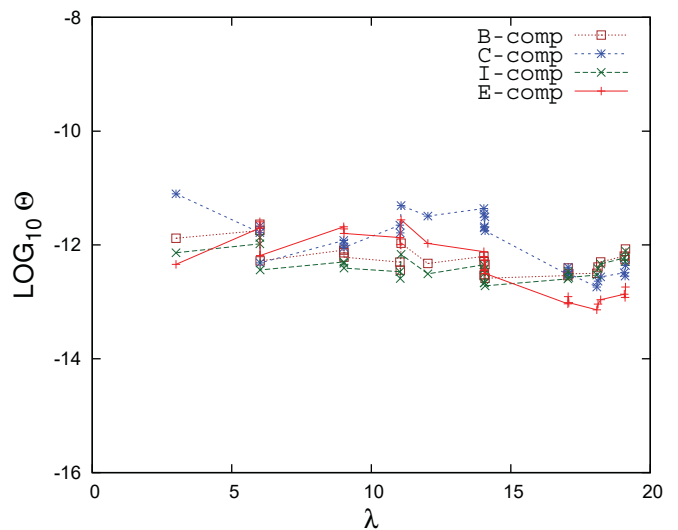


図 A.51 例題 6-2: 各合成による近似対の相対残差  $\Theta$  の分布



### A.3.6.3 例題 6-3

これは FEM の要素分割を  $(N_1, N_2, N_3) = (40, 50, 60)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 120,000$ ,  $w_L = 2,041$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 40$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.38 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.38 例題 6-3: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	3.92E-12	280.8
C-合成 ( $l=5, n=26$ )	2.07E-11	142.9
I-合成 ( $l=5, n=26$ )	2.50E-12	143.8
E-合成 ( $l=4, n=15$ )	7.15E-12	75.0

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.52 に示す。

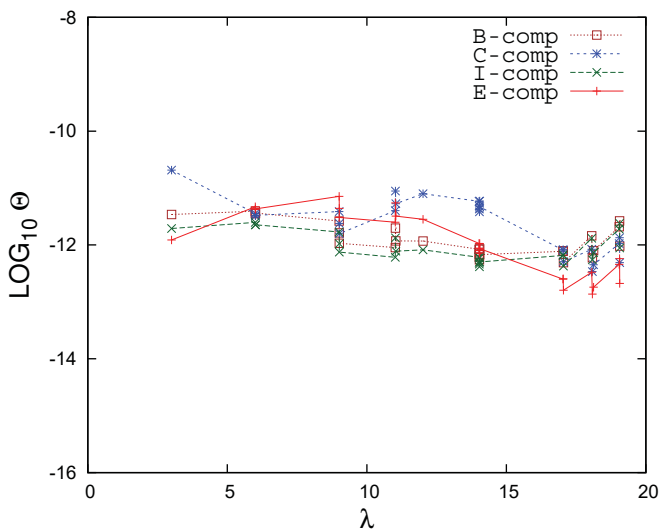


図 A.52 例題 6-3: 各合成による近似対の相対残差  $\theta$  の分布

### A.3.6.4 例題 6-4

これは FEM の要素分割を  $(N_1, N_2, N_3) = (50, 60, 70)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 210,000$ ,  $w_L = 3,051$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 40$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.39 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.39 例題 6-4: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	1.95E-11	856.1
C-合成 ( $l=5, n=26$ )	1.18E-10	439.3
I-合成 ( $l=5, n=26$ )	1.09E-11	440.5
E-合成 ( $l=4, n=15$ )	1.06E-11	273.8

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.53 に示す。

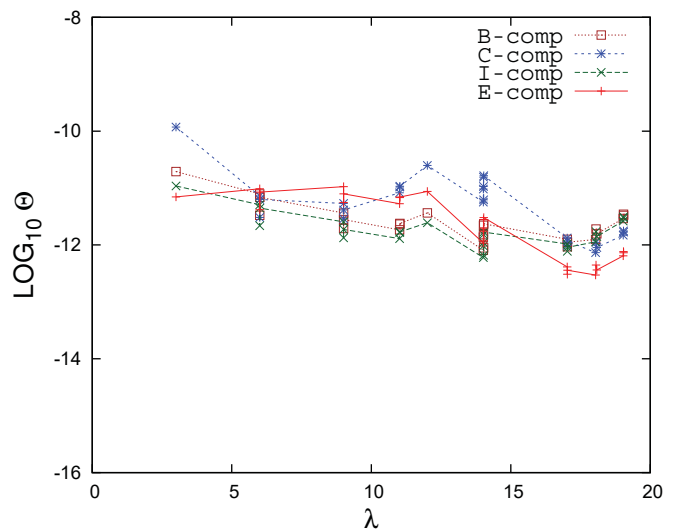


図 A.53 例題 6-4: 各合成による近似対の相対残差  $\theta$  の分布

### A.3.6.5 例題 6-5

これは FEM の要素分割を  $(N_1, N_2, N_3) = (60, 70, 80)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 336,000$ ,  $w_L = 4,261$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 40$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.40 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.40 例題 6-5: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	1.01E-11	2,639.9
C-合成 ( $l=5, n=26$ )	6.12E-11	1,372.4
I-合成 ( $l=5, n=26$ )	6.31E-12	1,378.5
E-合成 ( $l=4, n=15$ )	1.27E-11	952.5

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.54 に示す。

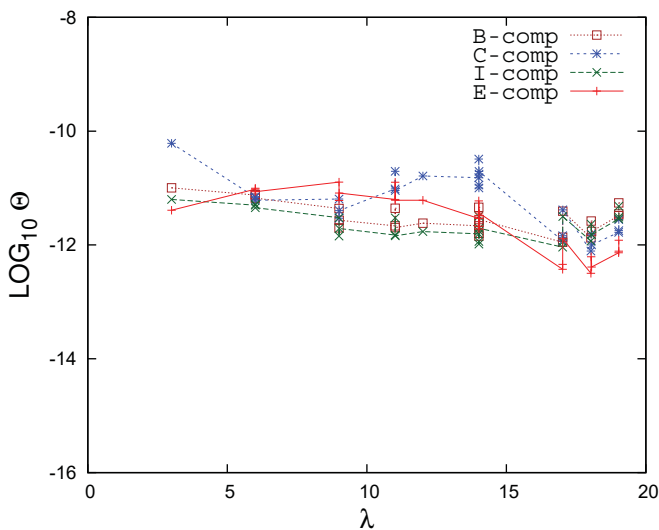


図 A.54 例題 6-5: 各合成による近似対の相対残差  $\theta$  の分布

### A.3.6.6 例題 6-6

これは FEM の要素分割を  $(N_1, N_2, N_3) = (70, 80, 90)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 504,000$ ,  $w_L = 5,671$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 40$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.41 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。B-合成については、システムが設定している使用可能な記憶量の制限値 168GiB を越えたため計算を実施できなかった。

表 A.41 例題 6-6: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B-合成 ( $l=9, n=30$ )	(記憶量の制限を超過)	
C-合成 ( $l=5, n=26$ )	1.46E-10	5,308.2
I-合成 ( $l=5, n=26$ )	1.44E-11	5,037.8
E-合成 ( $l=4, n=15$ )	2.66E-11	2,596.1

また、(計算を実施できなかった B-合成を除いた) 各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.55 に示す。

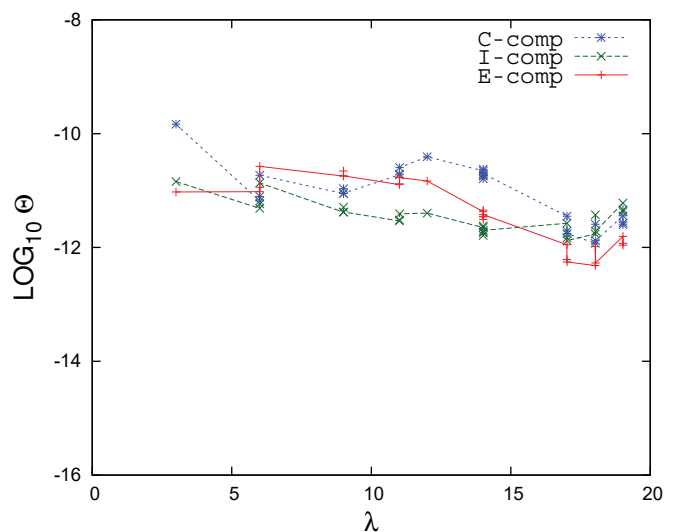


図 A.55 例題 6-6: 各合成による近似対の相対残差  $\theta$  の分布

A.3.7 例題 7 (下端固有対を求めた例, その 4)

例題 7 は有限要素法 (FEM) により要素分割を変えながら、離散化された立方体のラプラス固有値問題の固有値が区間  $[a, b] = [0, 20]$  にある中間固有対を求める問題である。使用したフィルタは、4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) に対してはいずれも伝達関数の 3 つの形状に対する条件として  $g_p = 0.1$ ,  $g_s \leq 10^{-16}$ ,  $\xi = 1.3$  を課した。合成用の有理関数の次数  $\ell$  は形状の要求を満たせるなるべく小さい整数としたが、例題 6 とは異なり、これは下端対を求める例であるが、B-合成と I-合成の次数  $\ell$  を偶数に強制した。なおチェビシェフ多項式の次数  $n$  は 50 以下で最小のものに制限している。

伝達関数はどれも  $t \in [-1, 1]$  には極が無いので、固有値の区間  $\lambda \in [0, 20]$  に  $t \in [-1, 1]$  を対応させてフィルタを構成した。

この例題 7 で各合成の場合に実現されたフィルタのパラメタを表 A.42 に示す。

合成の種類	次数 $\ell$	次数 $n$	$g_s$ の実現値
B	10	20	7.0E-17
C	5	26	6.0E-17
I	6	13	8.4E-17
E	4	15	2.4E-17

また、各合成によるフィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$  のグラフを図 A.56 に示す。

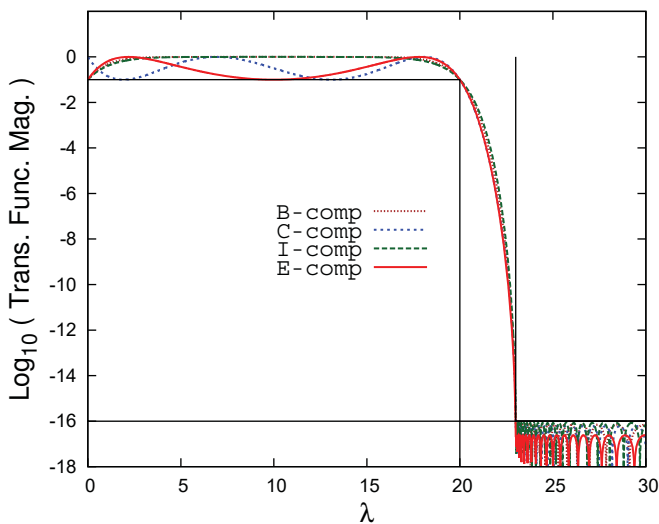


図 A.56 例題 7: フィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$

A.3.7.1 例題 7-1

これは FEM の要素分割を  $(N_1, N_2, N_3) = (60, 70, 80)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 336,000$ ,  $w_L = 4,261$  である。固有値が指定区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 50$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.43 に示す。表からは E-合成の場合の経過時間が最も短いことがわかる。

表 A.43 例題 7-1: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B ( $\ell=10, n=20$ )	8.3E-12	2,603.8
C ( $\ell= 5, n=26$ )	4.7E-11	1,433.8
I ( $\ell= 6, n=13$ )	1.0E-11	1,388.9
E ( $\ell= 4, n=15$ )	6.4E-12	987.1

また、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.57 に示す。

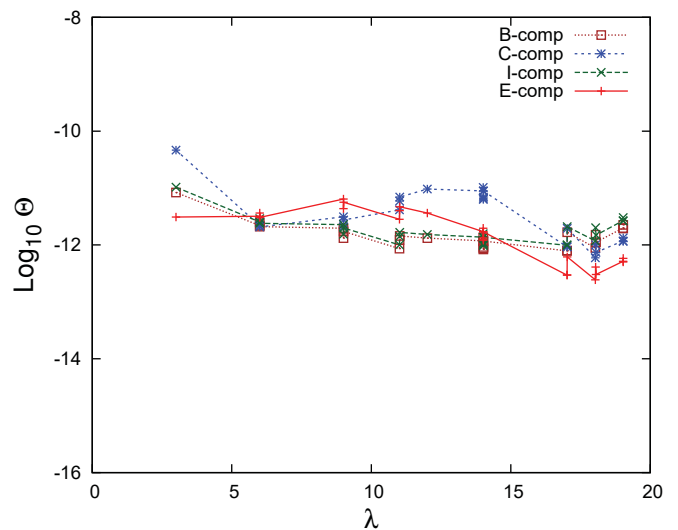


図 A.57 例題 7-1: 各合成による近似対の相対残差  $\Theta$  の分布

### A.3.7.2 例題 7-2

これは FEM の要素分割を  $(N_1, N_2, N_3) = (70, 80, 90)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 504,000$ ,  $w_L = 5,671$  である。固有値が区間  $[0, 20]$  にある対の真の数は 26 個である。フィルタを適用するベクトルの数は  $m = 50$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.44 に示す。B-合成と I-合成では記憶量の制限 168GiB を越えたため計算を実施できなかった。表からは C-合成よりも E-合成の場合の方が経過時間が短いことがわかる。

表 A.44 例題 7-2: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B ( $\ell=10, n=20$ )	(記憶量の制限を超過)	
C ( $\ell= 5, n=26$ )	5.0E-11	4,092.9
I ( $\ell= 6, n=13$ )	(記憶量の制限を超過)	
E ( $\ell= 4, n=15$ )	1.1E-11	2,651.6

また、計算が実施できた C-合成と E-合成の場合について、それらのフィルタを用いて得られた近似対の相対残差のグラフを図 A.58 に示す。

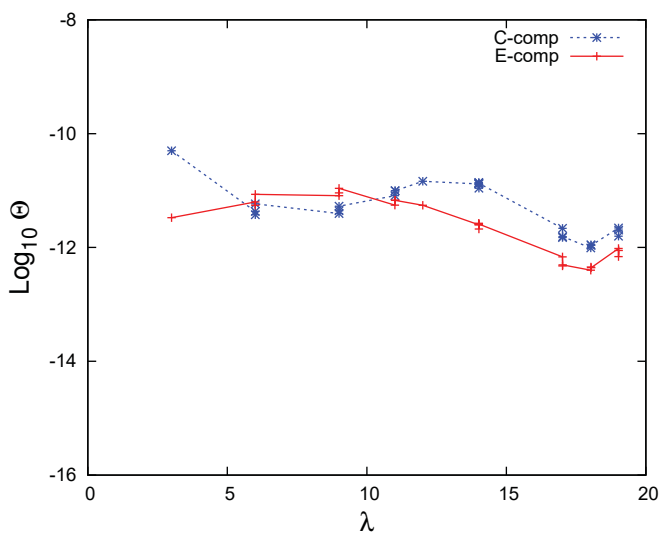


図 A.58 例題 7-2: 各合成による近似対の相対残差  $\theta$  の分布

### A.3.8 例題 8 (中間固有対を求めた例, その 4)

例題 8 は有限要素法 (FEM) により要素分割を変えて、離散化された立方体のラプラシアン固有値問題の固有値が区間  $[a, b] = [600, 650]$  にある中間固有対を求める問題である。

使用したフィルタは、4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) に対してはいずれも伝達関数の 3 つの形状に対する条件として  $g_p = 0.1$ ,  $g_s \leq 10^{-16}$ ,  $\xi = 1.1$  を課した。合成用の有理関数の次数  $\ell$  は形状の要求を満たせる最小の偶数とした。チェビシェフ多項式の次数  $n$  は条件を満たせる最小値とした。

この例題で各合成の場合に実現されたフィルタのパラメータを表 A.45 に示す。

表 A.45 例題 8: 使用したフィルタの構成パラメータ

合成の種類	次数 $\ell$	次数 $n$	$g_s$ の実現値
B	24	36	9.2E-17
C	8	48	9.6E-17
I	8	48	9.6E-17
E	6	10	1.5E-17

また、各合成によるフィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$  のグラフを図 A.59 に示す。

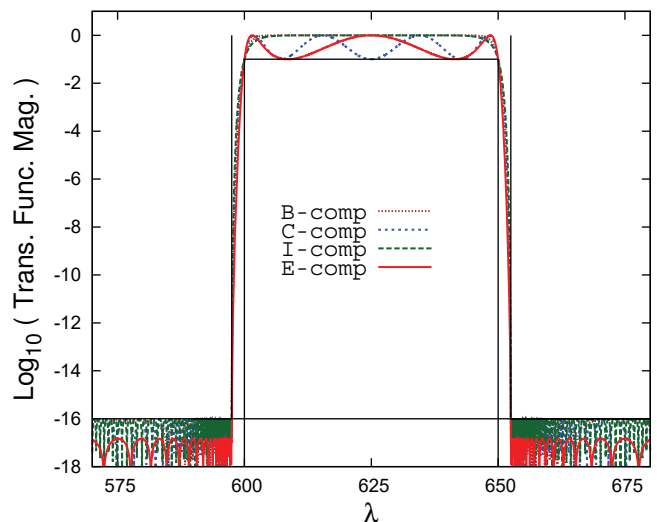


図 A.59 例題 8: フィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$



**A.3.8.1 例題 8-1**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (50, 60, 70)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 210,000$ ,  $w_L = 3,051$  である。固有値が区間  $[600, 650]$  にある対の真の数は 755 個である。フィルタを適用するベクトルの数は  $m = 900$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.46 に示す。表からは E-合成の経過時間が最も短いことがわかる。

表 A.46 例題 8-1: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B ( $\ell=24, n=36$ )	3.1E-12	10,471.6
C ( $\ell= 8, n=48$ )	1.7E-12	4,856.2
I ( $\ell= 8, n=48$ )	5.8E-12	4,796.4
E ( $\ell= 6, n=10$ )	4.9E-13	995.8

また各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.60 に示す。

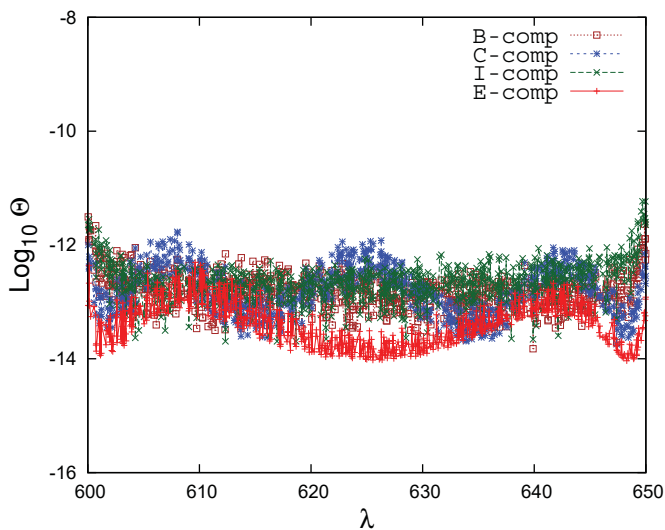


図 A.60 例題 8-1: 各合成による近似対の相対残差  $\Theta$  の分布

**A.3.8.2 例題 8-2**

これは FEM の要素分割を  $(N_1, N_2, N_3) = (60, 70, 80)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 336,000$ ,  $w_L = 4,261$  である。固有値が区間  $[600, 650]$  にある対の真の数は 795 個である。フィルタを適用するベクトルの数は  $m = 1,000$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.47 に示す。ただし B-合成の場合には記憶量の制限 168GiB を越えたために計算を実施できなかった。表からは E-合成の経過時間が最も短いことがわかる。

表 A.47 例題 8-2: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B ( $\ell=24, n=36$ )	(記憶量の制限を超過)	
C ( $\ell= 8, n=48$ )	3.2E-13	14,812.0
I ( $\ell= 8, n=48$ )	7.5E-13	14,659.8
E ( $\ell= 6, n=10$ )	4.0E-13	3,192.9

また、計算の実行できた C-合成, I-合成, E-合成のそれぞれについて、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.61 に示す。

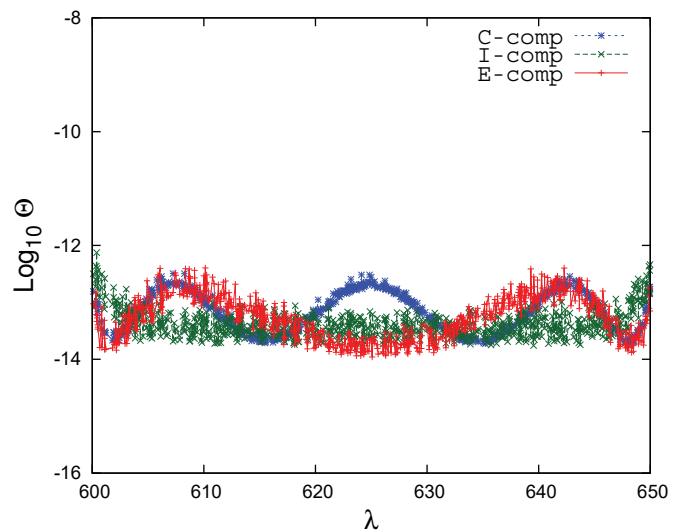


図 A.61 例題 8-2: 各合成による近似対の相対残差  $\Theta$  の分布

### A.3.9 例題 9 (下端固有対を求めた例, その 5)

例題 9 は有限要素法 (FEM) により要素分割を変えて、離散化された立方体のラプラシアン固有値問題の固有値が区間  $[a, b] = [0, 100]$  にある下端固有対を求める問題である。

使用したフィルタは、4 種類の各合成 (B-合成, C-合成, I-合成, E-合成) に対してはいずれも伝達関数の 3 つの形状に対する条件として  $g_p \leq 0.1$ ,  $g_s = 10^{-16}$ ,  $\xi = 1.1$  を課した。合成用の有理関数の次数  $\ell$  は形状の要求を満たせる最小の整数とした。チェビシェフ多項式の次数  $n$  は条件を満たせる最小値である。

この例題で各合成の場合に実現されたフィルタのパラメータを表 A.48 に示す。

合成の種類	次数 $\ell$	次数 $n$	$g_p$ の実現値
B	24	36	0.10068
C	8	48	0.10032
I	8	48	0.10032
E	5	17	0.11305

また、各合成によるフィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$  のグラフを図 A.62 に示す。

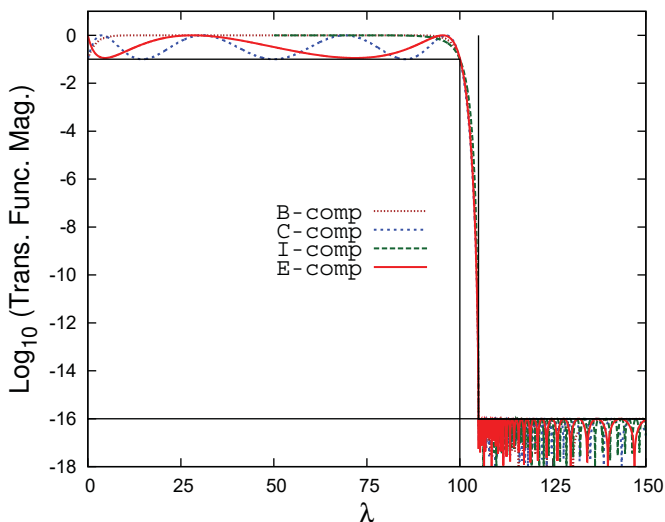


図 A.62 例題 9 : フィルタの伝達関数の大きさ  $|\hat{f}(\lambda)|$

### A.3.9.1 例題 9-1

これは FEM の要素分割を  $(N_1, N_2, N_3) = (50, 60, 70)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 210,000$ ,  $w_L = 3,051$  である。固有値が区間  $[0, 100]$  にある対の真の数は 402 個である。フィルタを適用するベクトルの数は  $m = 550$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A.49 に示す。表からは E-合成の経過時間が最も短いことがわかる。

表 A.49 例題 9-1 : 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B ( $\ell=24, n=36$ )	2.8E-12	7,239.0
C ( $\ell= 8, n=48$ )	5.0E-12	3,176.4
I ( $\ell= 8, n=48$ )	2.4E-12	3,212.7
E ( $\ell= 5, n=17$ )	1.1E-11	860.8

また各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A.63 に示す。

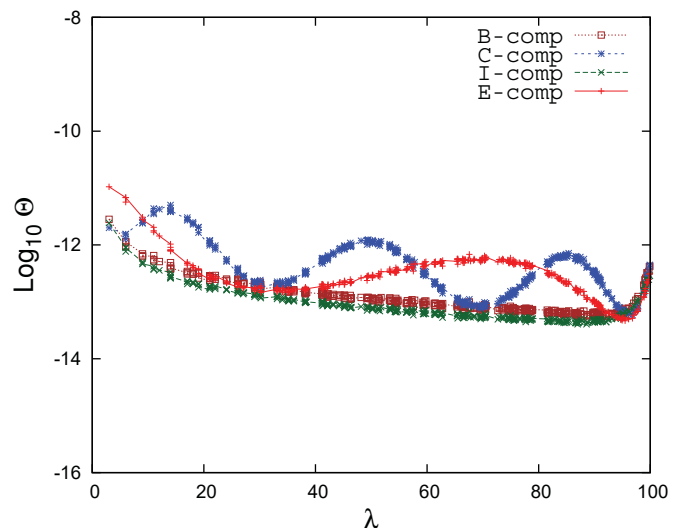


図 A.63 例題 9-1 : 各合成による近似対の相対残差  $\Theta$  の分布

A.3.9.2 例題 9 の 2

これは FEM の要素分割を  $(N_1, N_2, N_3) = (60, 70, 80)$  とした場合で、固有値問題の行列の次数と下帯幅は  $N = 336,000$ ,  $w_L = 4,261$  である。固有値が区間  $[0, 100]$  にある対の真の数は 408 個である。フィルタを適用するベクトルの数は  $m = 600$  とした。各合成によるフィルタを用いて得られた近似対の相対残差の最大値および近似対を得るまでの経過時間を表 A-50 に示す。ただし B-合成の場合には記憶量の制限 168GiB を越えたために計算を実施できなかった。表からは E-合成の経過時間が最も短いことがわかる。

表 A-50 例題 9-2: 各合成による相対残差の最大値と経過時間 (秒)

フィルタ	相対残差の最大値	経過時間
B ( $\ell=24, n=36$ )	(記憶量の制限を超過)	
C ( $\ell= 8, n=48$ )	1.8E-11	8,166.4
I ( $\ell= 8, n=48$ )	5.8E-11	8,216.9
E ( $\ell= 5, n=17$ )	3.3E-11	2,397.6

また、計算の実行できた C-合成, I-合成, E-合成のそれぞれについて、各合成によるフィルタを用いて得られた近似対の相対残差のグラフを図 A-64 に示す。

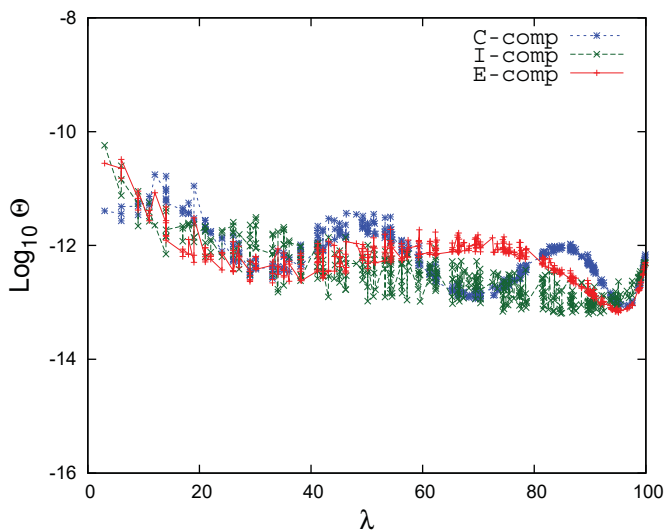


図 A-64 例題 9-2: 各合成による近似対の相対残差  $\Theta$  の分布

A.3.10 各例題で用いた, E-合成の部分分数分解の数値表

例題 1 から例題 9 のそれぞれで用いた E-合成の場合の部分分数分解  $\hat{x}(t) = x(h(t)) = c_\infty + \sum_{j=1}^{\ell} c_j/(t-t_j)$  について、定数項の係数  $c_\infty$  (実数) および虚部が非負の極  $t_j$  とその係数  $c_j$  の倍精度計算による数値の表を示す。(表 A-51, 表 A-52, 表 A-53, 表 A-54, 表 A-55). ただし、例題 6 と例題 7 で用いたものは例題 3 で用いたものと同じであり、例題 8 で用いたものは例題 1 で用いたものと同じであり、例題 9 で用いたものは例題 4 で用いたものと同じである。

次数  $\ell$  が奇数の場合には、実数の極が 1 つ現れるのでそれを  $t_R$ , そうしてその極に対する実数の係数を  $c_R$  と記している。次数  $\ell$  が偶数の場合には、 $\hat{x}(t)$  は偶関数なので、符号を反対にした極を持つことなども考慮に入れている。

数値は IEEE-754 の倍精度浮動小数点 (2 進 64bit) による計算で得たものである。(これらは実際に実験で用いた値で、より正確な値はたとえば 4 倍精度計算を用いて作ることができる)。

IEEE-754 の倍精度の真の有効精度は 10 進で 15.95 桁であるのに、表中では数値を 10 進 17 桁の精度で掲載している理由は、IEEE-754 の 2 進表現の 64bit 倍精度の値はそれを 10 進表現の精度 17 桁の数値に変換してから再び IEEE-754 の 2 進表現の倍精度に変換すると元の値に戻るからである。(ただし今回の実験では数値を表から読み込むのではなく、計算して得た値を直接使っている)。

表 A-51 例題 1, 例題 8 で用いた E-合成の部分分数分解 ( $\ell = 6$ )

値	実部	虚部
$t_1$	1.0183741988631465E+00	9.8314833085967862E-02
$t_2$	0	8.3274449632028524E-01
$t_3$	-1.0183741988631465E+00	9.8314833085967862E-02
$c_1$	-5.7673926346438742E-01	-1.7941921352872067E-01
$c_2$	0	-4.6337422188923432E+00
$c_3$	5.7673926346438742E-01	-1.7941921352872067E-01
$c_\infty$	0	

表 A-52 例題 2 で用いた E-合成の部分分数分解 ( $\ell = 6$ )

値	実部	虚部
$t_1$	1.0271978792948515E+00	1.0071437647109674E-01
$t_2$	0	9.0488236260493948E-01
$t_3$	-1.0271978792948515E+00	1.0071437647109674E-01
$c_1$	-4.9959045495499710E-01	-1.1673752735992586E-01
$c_2$	0	-4.1632752766212917E+00
$c_3$	4.9959045495499710E-01	-1.1673752735992586E-01
$c_\infty$	0	

表 A-53 例題 3, 例題 6, 例題 7 で用いた E-合成の部分分数分解 ( $\ell = 4$ )

値	実部	虚部
$t_1$	1.1552396197007031E+00	4.0897771272137828E-01
$t_2$	-1.1552396197007031E+00	4.0897771272137828E-01
$c_1$	-7.3704751451400419E-01	-1.2273002275978494E-01
$c_2$	7.3704751451400419E-01	-1.2273002275978494E-01
$c_\infty$	1	

表 A-54 例題 4, 例題 9 で用いた E-合成の部分分数分解 ( $\ell = 5$ )

値	実部	虚部
$t_1$	1.0946512591586728E+00	1.3998460308860974E-01
$t_2$	-9.9242878792622491E-01	9.2097096311897797E-01
$t_R$	-1.0666626241682344E+00	0
$c_1$	-1.8267067418533467E-01	3.9500805513145824E-02
$c_2$	1.0347753904109060E+00	-5.4670156890649124E-01
$c_R$	5.6396101623216781E-02	0
$c_\infty$	6.1585829119002800E-01	

表 A-55 例題 5 で用いた E-合成の部分分数分解 ( $\ell = 3$ )

値	実部	虚部
$t_1$	1.6081270689290255E+00	1.1428106752166010E+00
$t_R$	-1.4570474685157588E+00	0
$c_1$	-9.7349353519661719E-01	-1.2680166120629213E-01
$c_R$	1.8156855388212986E-01	0
$c_\infty$	7.0659077095890510E-01	

参考文献

[1] Abramowitz, Milton, Stegun, Irene A., (Eds): *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York (1972).

[2] Richard W. Daniels: *Approximation Methods for Electronic Filter Design*, McGraw-Hill, 1974.

[3] Miroslav D. Lutovac, Dejan V. Tošić, Brian L. Evans: *Filter Design for Signal Processing*, §12.8, Prentice Hall, 2001.

[4] William J. Thompson: *Atlas for Computing Mathematical Functions*, Wiley-Interscience, 1997.

[5] Sivan Toledo and Eran Rabani: Very Large Electronic Structure Calculations Using an Out-of-Core Filter-Diagonalization Method, *J. Comput. Phys.*, Vol.180, No.1, pp.256–269 (2002).

[6] Tetsuya Sakurai and Hiroshi Sugiura: A Projection Method for Generalized Eigenvalue Problems Using Numerical Integration, *J. Comput. Appl. Math.*, Vol.159, pp.119–128 (2003).

[7] Tetsuya Sakurai and Hiroto Tadano: CIRR: a Rayleigh-Ritz Type Method with Contour Integral for Generalized Eigenvalue Problems, *Hokkaido Math. J.*, Vol.36, No.4, pp.745–757 (2007).

[8] Eric Polizzi: A Density Matrix-based Algorithm for Solving Eigenvalue Problems, *Phys. Rev. B*, Vol.79, No.1, pp.115112–115117 (2009).

[9] Tsutomu Ikegami, Tetsuya Sakurai and Umpei Nagashima: A Filter Diagonalization for Generalized Eigenvalue Problems Based on the Sakurai-Sugiura Projection Method, *J. Comput. Appl. Math.*, Vol.233, No.8, pp.1927–1936 (2010).

[10] Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, Charles W. Clark (Eds): *NIST Handbook of Mathematical Functions*, Cambridge University Press (2010).

[11] Martin Galgon, Lukas Krämer and Brunno Lang: The FEAST Algorithm for Large Eigenvalue Problems, *PAMM- Proc. Appl. Math. Mech.*, Vol.11, pp.747–748 (2011).

[12] Anthony P. Austin and Lloyd N. Trefethen: Computing Eigenvalues of Real Symmetric Matrices with Rational Filters in Real Arithmetic, *SIAM J. Sci. Comput.*, Vol.37, No.3, pp.A1365–A1387 (2015).

[13] Stefan Güttel, Eric Polizzi, Ping Tak Peter Tang and Gautier Viaud: Zolotarev Quadrature Rules and Load

Balancing for the FEAST Eigensolver, *SIAM J. Sci. Comput.*, Vol.37, No.4, pp.A2100–A2122 (2015).

[14] Hiroshi Murakami: Filter Diagonalization Method by Using a Polynomial of a Resolvent as the Filter for a Real Symmetric-Definite Generalized Eigenproblem, in *proceedings of EPASA2015*, Springer, LNCSE-117, pp.205–232 (2018).

[15] Hiroshi Murakami: Filters Consist of a Few Resolvents to Solve Real Symmetric-Definite Generalized Eigenproblems, *Japan J. Indust. Appl. Math.*, Vol.36, No.2, pp.579–618 (July, 2019).  
url=(<https://doi.org/10.1007/s13160-019-00355-5>).

[16] 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, pp.1–21 (2010).

[17] 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, pp.1–14 (2011).

[18] 村上 弘: レゾルベントを用いたフィルタによる固有値問題の解法について, 情報処理学会研究報告, Vol.2012-HPC-133, No.22, pp.1–8 (2012).

[19] 村上 弘: 実対称定値一般固有値問題の最小側固有値を持つ固有対に対する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解法, 先進的計算基盤システムシンポジウム論文集 2012, pp.81–82 (2012).

[20] 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, pp.1–8 (2012).

[21] 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, pp.57–72 (2014).

[22] 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, pp.1–4 (2014).

[23] 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, pp.1–10 (2014).

[24] 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, pp.329–330 (2014 年 8 月).

[25] 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, pp.38–51 (2015).

[26] 村上 弘: 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, 情報処理学会研究報告, Vol.2015-HPC-149, No.7, pp.1–16 (2015).

[27] 村上 弘: 実数シフトのレゾルベントの多項式をフィルタに用いた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, 日本応用数学会 2015 年度年会予稿集 (統合版), pp.442–443 (2015).

[28] 村上 弘: 固有値問題の解法に用いるレゾルベントの多項式型のフィルタの設計, 情報処理学会研究報告, Vol.2016-HPC-153, No.38, pp.1–13 (2016 年 3 月).

[29] 村上 弘: 虚数シフトのレゾルベントの多項式の実部をフィルタに用いた実対称定値一般固有値問題の中間固有対の解法, HPCS2016 シンポジウム論文集 (ポスタ発表論文), p.49 (2016 年 6 月).

[30] 村上 弘: 実対称定値一般固有値問題の最小側固有対を解くための実数シフトのレゾルベントの多項式によるフィルタの簡易な設計法, 情報処理学会研究報告集, Vol.2016-HPC-155, No.44, pp.1–27 (2016).



- [31] 村上 弘：レゾルベントの多項式をフィルタに用いた対称定値一般固有値問題のフィルタ対角化法，日本応用数理学会 2016 年度年会予稿集，2 頁分 (2016 年 9 月).
- [32] 村上 弘：レゾルベントの多項式によるフィルタの伝達特性の調整，数理解析研究所講究録，No.2054, pp.168–181 (2017 年 10 月).
- [33] 村上 弘：少数のレゾルベントから構成されたフィルタを用いた対称定値一般固有値問題の解法，情報処理学会研究報告，Vol.2017-HPC-162, No.21, pp.1–34 (2017).
- [34] 村上 弘：少数のレゾルベントの多項式型フィルタを用いた一般固有値問題の解法，情報処理学会研究報告，Vol.2018-HPC-165, No.15, pp.1–21 (2018).
- [35] 村上 弘：フィルタにレゾルベントの線形結合の多項式を用いた複素エルミート定値一般固有値問題の解法，情報処理学会研究報告，Vol.2018-HPC-166, No.10, pp.1–17 (2018).
- [36] 村上 弘：単一のレゾルベントのチェビシェフ多項式による実対称定値一般固有値問題の解法用の簡易型フィルタ，情報処理学会論文誌：コンピューティングシステム (ACS64)，Vol.12, No.2, pp.1–26 (2019).
- [37] 村上 弘：フィルタの反復適用による実対称定値一般固有値問題の近似対の改良，情報処理学会論文誌：コンピューティングシステム (ACS65)，Vol.12, No.3, pp.14–33 (July, 2019).