

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

村上 弘^{1,a)}

受付日 2018年7月25日, 採録日 2018年11月2日

概要：実対称定値一般固有値問題をフィルタ対角化法を用いて解くためのフィルタの新しい設計法を提案した。本論文で扱うフィルタは単一のレゾルベントの多項式なので、レゾルベントを与える連立1次方程式を直接法を用いて解く場合には、複数のレゾルベントを用いるフィルタに比べて計算量が少ない。そのようなフィルタの伝達関数の形状を最適にするには多数のパラメタを決定する必要があるが、今回提案したフィルタではチェビシェフ多項式で表される準最適な形状の伝達関数を採用して簡易な設計を可能にした。提案するフィルタの構成法を、求める固有値を含む区間が固有値分布の端にある場合と任意の位置にある場合の両方に対して具体的に示した。

キーワード：フィルタ対角化, 固有値問題, レゾルベント

Simple Filters by a Chebyshev Polynomial of a Single Resolvent to Solve Real Symmetric-definite Generalized Eigenproblems

HIROSHI MURAKAMI^{1,a)}

Received: July 25, 2018, Accepted: November 2, 2018

Abstract: We proposed a new method to design a filter for the solutions of real symmetric-definite generalized eigenproblems. The filter we proposed is a polynomial of a single resolvent, therefore the amount of computation is less compared from those of current filters which use many resolvents when a system of linear equations which gives the action of the resolvent is solved by some direct method. Such filter has many parameters to be determined so to make the shape of the transfer function optimal. However for the present filter to make the simple design possible, we selected the transfer function represented by a Chebyshev polynomial whose shape is sub-optimal. We showed both constructions of present filters for the case when the interval of eigenvalues for solutions is located at the lower end of the eigenvalue distribution and also for the case when the interval is located at any place.

Keywords: filter diagonalization, eigenproblem, resolvent

1. はじめに

いま係数行列 A と B が実対称で B は正定値である一般固有値問題

$$A\mathbf{v} = \lambda B\mathbf{v} \quad (1)$$

の固有対でその固有値が指定された区間 $[a, b]$ にあるもの

¹ 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)} mrkmlhrsh@tmu.ac.jp

をフィルタ対角化法 [1] を用いて解くことにする。固有値が $[a, b]$ 近傍にある固有ベクトルは良く通過させるが、 $[a, b]$ から離れた固有ベクトルは強く減衰させる特性を持つ線形作用素をうまく構成してフィルタに用いる。

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上の積分の離散化近似からフィルタの構成を導出する方法がある (文献 [2], [3], [4], [5], [6], [7])。これらはフィルタの作用を積分の離散化点と対応するシフト ρ_i を持つ複数のレゾルベント $\mathcal{R}(\rho) \equiv (A - \rho B)^{-1}B$ を用いて実現している。

しかし線形作用素であるフィルタを用いて線形の固有値問題を解く場合には、各固有ベクトルに対するフィルタの伝達特性が本質的であり、線形フィルタ $\mathcal{F}$ の固有値が λ である固有ベクトル $\mathbf{v}$ に対する伝達関数 $f(\lambda)$ の特性が望ましいものでありさえすれば、フィルタは積分の離散化近似から導かれるものでなくても構わない。

いまフィルタをレゾルベントの線形結合 $\mathcal{F} = c_\infty I + \sum_j c_j \mathcal{R}(\rho_j)$ (あるいはその実部) とすると、その伝達関数 $f(\lambda)$ は固有値 λ の有理関数になることが分かる。いま有理関数 $f(\lambda)$ をフィルタの伝達関数として望ましい振る舞いを持つようにうまく決めてやり、それから逆に $\mathcal{F}$ を構成する、つまりシフト ρ_j の分布と線形結合の係数 c_j と c_∞ が決まる。このような手順により、固有値が指定区間 $[a, b]$ にある固有ベクトルを選択的に抽出するのに適した伝達特性を持つフィルタが設計できる (文献 [8], [10], [11], [14])。またその変種として、複素数演算の使用を避けるために、シフトを実数に限定したレゾルベントの線形結合をフィルタとして用いる方法も考えられる (文献 [12], [13], [18])。これらの方法では複数のレゾルベントを用いてフィルタを構成する。

いま、フィルタ $\mathcal{F}$ として「複数のレゾルベント」の作用の線形結合の代わりに「単一のレゾルベント」の作用の多項式を採用しても、やはりそれは線形作用素であり、フィルタの固有値 λ の固有ベクトルに対する伝達関数 $f(\lambda)$ は固有値 λ の有理関数になる。そこで有理関数 $f(\lambda)$ をうまく構成してやり、それがフィルタの伝達関数として欲しい固有値の固有ベクトルを選択的に抽出するのに適した特性を持つようにする。そうしてその $f(\lambda)$ からそれを伝達関数に持つフィルタ $\mathcal{F}$ が逆構成できる。「単一のレゾルベント」は、固有値を求める区間の位置を任意にするために虚数をシフトに用いる場合と、区間の位置を固有値分布の端に制限することで実数をシフトに用いる場合の2種類が考えられる (文献 [9], [15], [16], [17], [19], [20], [21], [22], [23])。

今回の論文では、フィルタは「単一のレゾルベント」の作用の (虚部の) 多項式とし、さらにその多項式はある1つのチェビシェフ多項式で表される簡易型の設計に限定した。またシフトが実数と虚数のそれぞれの場合について、その定式化と説明の記述がなるべく統一的なものとなるように努めた。

今回の単一のレゾルベントのチェビシェフ多項式を用いたフィルタでは、求めたい固有値の範囲が固有値分布の下端である特別な場合には、レゾルベントのシフトに実数 ρ を用いることができて、そのときフィルタは以下の式 (2) の形で与えられる。

$$\mathcal{F} = g_s T_n(2\gamma \mathcal{R}(\rho) - I). \quad (2)$$

求めたい固有値の範囲の位置に条件を付けないより一般的な場合には、レゾルベントのシフトに虚数 ρ' を用いて、

フィルタは以下の式 (3) の形で表される。

$$\mathcal{F} = g_s T_n(2\gamma' \text{Im } \mathcal{R}(\rho') - I). \quad (3)$$

ただしここで、 $T_n(x)$ は n 次の第1種チェビシェフ多項式を、 I は恒等演算子を、 Im は虚部を取り出す演算子を表し、 γ や γ' は実定数であり、また g_s も実定数でそれは阻止域での伝達率の大きさの上限値である。多項式をチェビシェフ多項式で表される形に限定しているのが得られるフィルタの特性は最適ではなくて準最適なものになるが、フィルタの設計は主に数式を用いて見通し良く簡易に行えて、フィルタの実装上も記述が簡単になるという利点がある。

単一のレゾルベントを用いるフィルタの別の構成法としては、すでに [9] において、求める固有値の範囲の位置は任意としてシフト ρ には適切に選んだ虚数を用いて、フィルタを以下の式 (4) で表されるレゾルベント $\mathcal{R}(\rho)$ の複素係数多項式の実部の形とする方法を示した。

$$\mathcal{F} = \text{Re} \sum_{j=1}^n \gamma_j \{\mathcal{R}(\rho)\}^j. \quad (4)$$

ただしこの多項式の係数 $\{\gamma_j\}$ は、フィルタの伝達関数の形状がなるべく良くなるように決めるものであり、たとえば (多少の式変形の後に) 最小2乗法に類似した方法を用いた最適化の計算により数値を並べた表として得られる。

2. フィルタ対角化法の概要

フィルタ対角化法の概要は以下のようになる。

- (1) まず固有値が区間 $[a, b]$ 近傍の固有ベクトルは良く通過させるが、 $[a, b]$ から離れた固有ベクトルは強く阻止する性質を持つように調整された線形作用素 $\mathcal{F}$ を用意してそれをフィルタとして用いる。
- (2) ランダムな N 次ベクトル m 個の組を生成し、それを元にして B -正規直交化により m 個の B -正規直交ベクトルの組 X ($N \times m$ 行列) を作る ($X^T B X = I$)。
- (3) 作成した X に線形作用素であるフィルタ $\mathcal{F}$ を作用させて、ベクトルの組 $Y \leftarrow \mathcal{F} X$ ($N \times m$ 行列) を作る。
- (4) 得られた Y の列の適切な線形結合の組により「 $[a, b]$ 近傍にある固有値すべてに対応する不変部分空間」を近似する空間の基底をうまく構成する (その構成には X, Y およびフィルタの伝達関数の特性を参照して決めると良い)。
- (5) 構成された「不変部分空間を近似する空間の基底」に対してレイリー・リッツ法を適用して得られたリッツ対を一般固有値問題の近似対とする。

3. 下端固有対を求める場合のフィルタ

本章では、固有値が固有値分布の下端にある固有対 (下端固有対) の近似を求めるためのフィルタについて扱う。

3.1 レゾルベントの多項式のフィルタとその伝達関数

「下端固有対」を求めるためのフィルタとして「シフトが実数のレゾルベント」の多項式を採用する．係数行列 A と B を持つ一般固有値問題に対応して，以下の式 (5) でシフトが ρ のレゾルベントを定義する．

$$\mathcal{R}(\rho) \equiv (A - \rho B)^{-1} B. \quad (5)$$

そうして実数のシフト ρ と実多項式 $\mathcal{P}$ をうまく選び，以下の式 (6) で与えられるレゾルベントの多項式をフィルタとする．

$$\mathcal{F} \equiv \mathcal{P}(\mathcal{R}(\rho)). \quad (6)$$

ここで $\mathcal{P}$ はある n 次の実多項式である．このフィルタはシフト ρ が最小固有値未満であれば有界な作用素になる．

いま扱っている固有値問題の固有対 $(\lambda, \mathbf{v})$ はすべて実にとれる．そうしてレゾルベントの固有ベクトルへの作用は以下の式 (7) となる．

$$\mathcal{R}(\rho) \mathbf{v} = \frac{1}{\lambda - \rho} \mathbf{v}. \quad (7)$$

すると固有ベクトル $\mathbf{v}$ に対するフィルタ $\mathcal{F} = \mathcal{P}(\mathcal{R}(\rho))$ の作用は以下の式 (8) となる．

$$\mathcal{F} \mathbf{v} = f(\lambda) \mathbf{v}. \quad (8)$$

ここでフィルタの伝達関数 $f(\lambda)$ は以下の式 (9) で与えられる実有理関数になる．

$$f(\lambda) = \mathcal{P}\left(\frac{1}{\lambda - \rho}\right). \quad (9)$$

下端固有対を求める場合には，その固有値が区間 $\lambda \in [a, b]$ に含まれるとすると，固有値 λ の正規化座標 t を区間 $\lambda \in [a, b]$ から標準区間 $t \in [0, 1]$ への 1 次変換 $\lambda = a + (b - a)t$ により定義する．その逆変換は $t = \frac{1}{b-a}(\lambda - a)$ になる．そうして伝達関数を正規化座標 t で表したものを $g(t) = f(\lambda)$ で定義する．そうしてパラメタ $\mu > 1$ を導入して $t \in [0, 1]$ を通過域， $t \in (1, \mu)$ を遷移域， $t \in [\mu, \infty)$ を阻止域とする (図 1 参照．緑の線で固有値の存在範囲を表している)．

パラメタ g_p と g_s が $1 > g_p \gg g_s > 0$ を満たすとき，伝達関数 $g(t)$ の形状 (図 2) は以下の 3 条件を満たすとする．

- (1) 阻止域では $|g(t)| \leq g_s$
- (2) 通過域では $g(t) \geq g_p$ で， t が非負の範囲では逆も成立
- (3) 通過域での $g(t)$ の最大値は 1 (と規格化)

固有値の座標 λ とその正規化座標 t の関係に従って $\lambda = \rho$ ($< a$) には $t = -\sigma$ (< 0) が対応しているとする．そのとき $f(\lambda)$ が n 次実多項式 $\mathcal{P}$ を用いて以下の式 (10) の形で表される実有理関数であれば， $g(t)$ はある n 次実多項式 $\mathcal{Q}$ を用いて以下の式 (11) の形で表される実有理関数である．

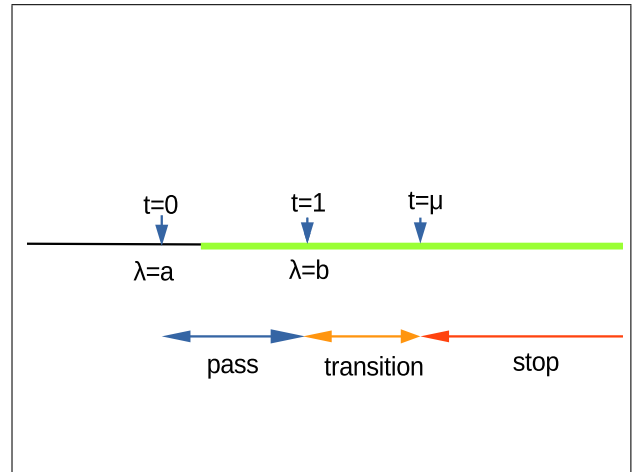


図 1 固有値 λ の区間 $[a, b]$ と正規化座標 t の関係通過域 $\lambda \in [a, b]$, $t \in [0, 1]$; 遷移域 $t \in (1, \mu)$; 阻止域 $[\mu, \infty)$

Fig. 1 Relation between the interval of eigenvalue $[a, b]$ and the normalized coordinate t . Passband $\lambda \in [a, b]$, $t \in [0, 1]$; Transitionband $t \in (1, \mu)$; Stopband $[\mu, \infty)$.

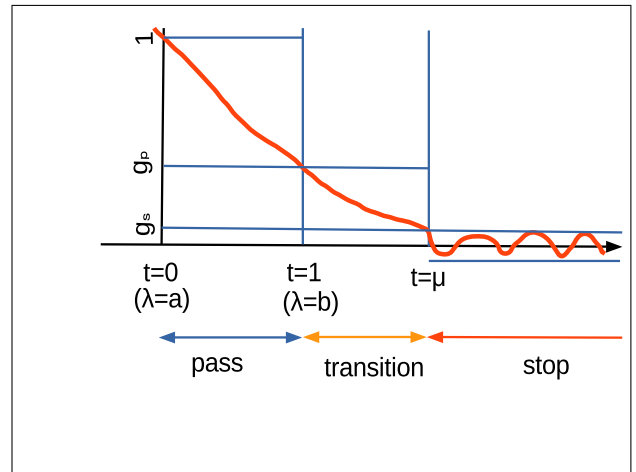


図 2 下端固有対用のフィルタの伝達関数 $g(t)$ の概形

Fig. 2 Conceptual figure of the transfer function $g(t)$ of a filter for lower exterior eigenpairs.

$$f(\lambda) = \mathcal{P}\left(\frac{1}{\lambda - \rho}\right). \quad (10)$$

$$g(t) = \mathcal{Q}\left(\frac{1}{t + \sigma}\right). \quad (11)$$

($f(\lambda) = g(t)$ の関係があるので， $f(\lambda)$ の式において λ を t に変換すると $\mathcal{P}$ から $\mathcal{Q}$ が決まる)．逆に $g(t)$ が式 (11) の形に決まればそれに対応して $f(\lambda)$ の表式が式 (10) の形に決まる．

3.2 伝達関数の 3 つの形状パラメタについて

伝達関数の 3 つの形状パラメタ μ , g_p , g_s について，まず前提として $\mu > 1$, $1 > g_p \gg g_s > 0$ であり，その条件のもとで形状の良さあるいは望ましさについては以下のように考える．

- μ は 1 に近いほど良い. なぜならば μ が大きくなると遷移域が広がるので, それにともなって遷移域に含まれる固有値の数が増えるとそれだけ (不変部分空間を近似する空間の基底を作るために) 多くのベクトルを濾過する必要が生じるからである.
- 阻止域における伝達率の大きさの上限値 g_s は微小であるほど良い. なぜならこの値が微小でなければ, 阻止域に固有値がある固有ベクトルが十分に排除されないでフィルタで濾過した結果に混入してしまうので, 不変部分空間の近似が悪化するからである.
- 通過域での伝達率の最小値である g_p は 1 に近いほど良い. なぜならばこの値が 1 に比べてかなり小さければ, 固有値が通過域にある固有対全体についての伝達率の最大最小比 $1/g_p$ が大きくなり, 伝達率が相対的に小さい近似対についてはそれにともなって有効精度が落ちる可能性があるからである.

たとえばいま次数 n を固定したときに式 (11) の伝達関数の極の位置 $t = -\sigma$ と n 次実多項式 Q を調整することで, $g(t)$ の 3 つの形状パラメタ μ , g_p , g_s をなるべく良いものにするを試みることになる. ただしこれら 3 つの形状パラメタはすべてを同時に良くすることはできない関係にあるので, 釣り合いを考えて調整をすることになる. 最適化の計算で Q を一般係数の実多項式にとり, 極の位置 $-\sigma$ もまた最適化のパラメタにとる場合には, 数値的手法を用いるので最適化で得られる最適な極の位置や多項式の係数は数値としてだけ得られる.

そこでチェビシェフ多項式の性質を利用して, 最適ではないが簡単な式計算だけで制約を満たす伝達関数 $g(t)$ が求められる設計法を導入する. その設計法によると, フィルタの阻止域での特性を良くするために $g(t)$ の阻止域に於ける大きさを微小にすることはチェビシェフ多項式の次数 n を上げることで容易に実現ができる. しかし通過域に於ける値の最大最小比を小さく抑えようとして μ の値をかなり大きくすると遷移域が広がり, それにともなってベクトルをフィルタで濾過した結果に含まれる不要な固有値を持つ固有ベクトルの数が (潜在的に) 増加するので, それに相応してフィルタで濾過するベクトルの数を増す必要が生じてしまい, 計算全体の効率が低下する. つまり今のチェビシェフ多項式を用いる簡易設計の場合には, 通過域に於ける最大最小比 $1/g_p$ をあまり小さくすることができない. 通過域での伝達率の最大最小比が大きければそれだけ, 得られる近似対の精度が均一ではなくなる可能性を持つ.

チェビシェフ多項式の定義

本論文では x についての n 次の (第 1 種) チェビシェフ多項式を $T_n(x)$ と表す. その定義は以下の漸化式 (12) で与えられる.

$$\begin{cases} T_0(x) = 1, T_1(x) = x, \\ T_n(x) = 2x T_{n-1}(x) - T_{n-2}(x) \quad (n \geq 2). \end{cases} \quad (12)$$

その最初の 5 次までのものは以下の式 (13) のようになる.

$$\begin{cases} T_0(x) = 1, \\ T_1(x) = x, \\ T_2(x) = 2x^2 - 1, \\ T_3(x) = 4x^3 - 3x, \\ T_4(x) = 8x^4 - 8x^2 + 1, \\ T_5(x) = 16x^5 - 20x^3 + 5x. \end{cases} \quad (13)$$

あるいはまた三角関数/双曲関数を用いて以下の式 (14) のようにも表せる.

$$T_n(x) = \begin{cases} \cos(n \cos^{-1} x) & (|x| \leq 1 \text{ のとき}), \\ \cosh(n \cosh^{-1} x) & (x \geq 1 \text{ のとき}). \end{cases} \quad (14)$$

そうして $T_n(x)$ は次数 n の偶奇に従って偶関数か奇関数である, つまり $T_n(-x) = (-1)^n T_n(x)$ を満たす.

3.3 チェビシェフ多項式で表された伝達関数

いま伝達関数 $g(t)$ を n 次のチェビシェフ多項式を用いた以下の形の式 (15) に制限する:

$$g(t) = g_s T_n(y), \quad y = 2x - 1, \quad x = \frac{\mu + \sigma}{t + \sigma}. \quad (15)$$

これはパラメタの 3 つ組 (n, μ, σ) から陽的に決定される (定数 g_s は規格化条件 $g(0) = 1$ から決まる). 阻止域 $t \in [\mu, \infty)$ には $y \in (-1, 1]$ の全域が対応する. チェビシェフ多項式の性質により, この $g(t)$ は阻止域では $|g(t)| \leq g_s$ を満たし, また阻止域以外の $t \in [0, \mu)$ において単調減少になるので, 残りの 2 つの条件 $g(0) = 1$, $g(1) = g_p$ から以下の式 (16) が得られる.

$$\begin{cases} \frac{1}{g_s} = T_n\left(1 + 2 \frac{\mu}{\sigma}\right), \\ \frac{g_p}{g_s} = T_n\left(1 + 2 \frac{\mu-1}{\sigma+1}\right). \end{cases} \quad (16)$$

あるいはこれを逆双曲線関数を用いて (実数 z についての公式 $\cosh^{-1}(1 + 2z^2) = 2 \sinh^{-1} |z|$ も用いて) 表すと以下の式 (17) になる.

$$\begin{cases} \cosh^{-1} \frac{1}{g_s} = 2n \sinh^{-1} \sqrt{\frac{\mu}{\sigma}}, \\ \cosh^{-1} \frac{g_p}{g_s} = 2n \sinh^{-1} \sqrt{\frac{\mu-1}{\sigma+1}}. \end{cases} \quad (17)$$

3.3.1 3 つ組 (n, μ, σ) からの g_p と g_s の計算

パラメタの 3 つ組 (n, μ, σ) が指定されたとき, それから g_p と g_s の値を求めるには, 式 (16) あるいは以下の式の組 (18) の中の各右辺をそれぞれ順に計算すればよい.

表 1 例: $g_p=10^{-7}$, $g_s=10^{-15}$ の場合の, n に対する μ と σ の値
Table 1 Example: Values of μ and σ for n for the case $g_p=10^{-7}$ and $g_s=10^{-15}$.

n	μ	σ
10	2.63	0.330
15	1.87	0.872
20	1.65	1.66
25	1.56	2.68
30	1.52	3.93
35	1.49	5.41
40	1.47	7.12
45	1.46	9.06
50	1.45	11.2

$$\begin{cases} \frac{1}{g_s} \leftarrow \cosh \left(2n \sinh^{-1} \sqrt{\frac{\mu}{\sigma}} \right), \\ \frac{g_p}{g_s} \leftarrow \cosh \left(2n \sinh^{-1} \sqrt{\frac{\mu-1}{\sigma+1}} \right). \end{cases} \quad (18)$$

3.3.2 3つ組として (n, g_p, g_s) を指定する場合

パラメタの3つ組として (n, g_p, g_s) が指定されたときに, それから σ と μ の値を求めるには, まず式 (17) を以下の式 (19) の形に変えて, その各右辺をそれぞれ計算したものを順に w_1 , w_2 とおく.

$$\begin{cases} \sqrt{\frac{\mu}{\sigma}} = \sinh \left(\frac{1}{2n} \cosh^{-1} \frac{1}{g_s} \right), \\ \sqrt{\frac{\mu-1}{\sigma+1}} = \sinh \left(\frac{1}{2n} \cosh^{-1} \frac{g_p}{g_s} \right). \end{cases} \quad (19)$$

すると以下の式 (20) の2つの関係が得られる.

$$\frac{\mu}{\sigma} = w_1^2, \quad \frac{\mu-1}{\sigma+1} = w_2^2. \quad (20)$$

それを解けば σ と μ が以下の式 (21) により簡単に求まる.

$$\sigma \leftarrow \frac{w_2^2 + 1}{(w_1 - w_2)(w_1 + w_2)}, \quad \mu \leftarrow \sigma w_1^2. \quad (21)$$

以上のようにして $g(t)$ の定義式 (15) が含むパラメタの3つ組 (n, μ, σ) が得られる.

例として $g_p = 10^{-7}$ と $g_s = 10^{-15}$ を指定したときに, 次数 n の値を 10 から 50 まで 5 刻みで変えた場合の μ と σ の値をそれぞれ表 1 に示す. 次数 n の増加にともなう μ の値の減少は非常に緩やかであることが分かる.

3.3.3 形状パラメタの3つ組 (μ, g_p, g_s) を指定する場合

形状パラメタの3つ組 (μ, g_p, g_s) が指定されたときに, それを (なるべく近く) 満たすように n と σ をうまく決めれば, 伝達関数を陽に決定できるパラメタの3つ組 (n, μ, σ) が揃うことになる. それにはまず2つの制約条件から以下の n を表す2通りの式 (22) が導かれる.

$$n = \frac{\cosh^{-1} \frac{1}{g_s}}{2 \sinh^{-1} \sqrt{\frac{\mu}{\sigma}}}, \quad n = \frac{\cosh^{-1} \frac{g_p}{g_s}}{2 \sinh^{-1} \sqrt{\frac{\mu-1}{\sigma+1}}}. \quad (22)$$

その2通りに表された n の値が一致する条件から, σ についての非線形方程式 $G(\sigma) = r$ が得られる. その左辺 $G(\sigma)$ と右辺 r は以下の (23) で与えられる.

$$G(\sigma) \equiv \frac{\sinh^{-1} \sqrt{\frac{\mu-1}{\sigma+1}}}{\sinh^{-1} \sqrt{\frac{\mu}{\sigma}}}, \quad r \equiv \frac{\cosh^{-1} \frac{g_p}{g_s}}{\cosh^{-1} \frac{1}{g_s}}. \quad (23)$$

この非線形方程式 $G(\sigma) = r$ に正の実数解があれば, それをたとえば2分法を用いて解くことで σ の値が得られる. そのとき次数 n を表す2つの式 (22) で計算される n の値はもちろん一致するが一般には整数にはならず, 次数が正の整数であるべきことに反する. そこで計算により得られた n に対する実数値を切り捨て (あるいは切り上げ) て整数化した値を次数 n として設定するという便法を用いる. するとすでに解いて得られた σ と整数化で得た次数 n , さらに最初に与えた μ をあわせた3つ組 (n, μ, σ) から, 前述の式の組 (18) を用いる方法により, 修正された g_p と g_s の値を求めることができる. このようにして得られた修正された g_p と g_s の値は最初に指定した元の値には一致しないが, その違いが許容できる程度であるならば, 修正されたそれらの値があたかも最初から指定されていたかのように扱うならば, 辻褄が合うことになる. このようにして求めたパラメタの3つ組 (n, μ, σ) を与えることでチェビシェフ多項式で表された式 (15) の $g(t)$ が決定する.

3.4 伝達関数 $g(t)$ からのフィルタ $\mathcal{F}$ の構成

いま式 (15) の形を持つ $g(t)$ が決まったときに, それを伝達関数とするフィルタ $\mathcal{F}$ を以下のようにして構成できる.

まず正規化座標 t と固有値 λ の間の対応関係 $t = (\lambda - a)/(b - a)$ を用いることで, x を λ で表した以下の式 (24) が得られる. ただしここで $\gamma \equiv (b - a)(\sigma + \mu)$, $\rho \equiv a - (b - a)\sigma$ である ($\sigma > 0$ であることにより $\rho < a$ である).

$$x = \frac{\mu + \sigma}{t + \sigma} = \frac{\gamma}{\lambda - \rho}. \quad (24)$$

伝達関数 $g(t) = g_s T_n(y)$ のチェビシェフ多項式の引数 $y = 2x - 1$ を λ で表したものは以下の式 (25) になる.

$$y = \frac{2\gamma}{\lambda - \rho} - 1. \quad (25)$$

すると正規化座標 t による伝達関数である $g(t)$ を固有値の座標 λ で表した伝達関数 $f(\lambda)$ は以下の式 (26) になる.

$$f(\lambda) = g_s T_n \left(\frac{2\gamma}{\lambda - \rho} - 1 \right). \quad (26)$$

よって $x = \gamma/(\lambda - \rho)$ を作用素 $\gamma \mathcal{R}(\rho)$ に, 定数 1 を恒等作用素 I にそれぞれ置き換えると, 伝達関数 $f(\lambda)$ に対応するフィルタの作用素 $\mathcal{F}$ が以下の式 (27) で得られる.

$$\mathcal{F} = g_s T_n (2\gamma \mathcal{R}(\rho) - I). \quad (27)$$

```

3 組  $(n, \mu, \sigma)$  と区間  $[a, b]$  を与えて
 $\rho \leftarrow a - (b - a)\sigma$  ;    $\gamma \leftarrow (b - a)(\sigma + \mu)$  ;

 $Z \leftarrow \mathcal{R}(\rho)X$  ;
 $V \leftarrow X$  ;
 $W \leftarrow 2\gamma Z - X$  ;
for  $k := 2$  to  $n$  do begin
     $Z \leftarrow \mathcal{R}(\rho)W$  ;
     $Y \leftarrow 4\gamma Z - 2W - V$  ;
     $V \leftarrow W$  ;
     $W \leftarrow Y$ 
end ;
 $Y \leftarrow g_s Y$ 

```

図 3 算法：下端固有対用フィルタのベクトルの組への作用 $Y \leftarrow \mathcal{F}X$

Fig. 3 Algorithm: An application of the filter to a set of vectors $Y \leftarrow \mathcal{F}X$ for lower exterior eigenvalues.

3.5 下端固有対用のフィルタの作用の実装法

レゾルベント $\mathcal{R}(\rho) = (A - \rho B)^{-1}B$ をベクトルの組 W に作用させる処理 $Z \leftarrow \mathcal{R}(\rho)W$ は、まず W から右辺ベクトルの組 BW を作り、次に係数 $C \equiv A - \rho B$ の連立 1 次方程式の組 $CZ = BW$ を Z について解くことで実現する。

行列 A と B はともに実対称で B は正定値であり、シフト ρ は実数で最小固有値未満であることから、行列 C は実対称で正定値である。係数行列が実対称で正定値の連立 1 次方程式は実数演算だけを用いてピボット交換を行わずに（改訂）コレスキ法で数値的に安定に解くことができる。さらにもしも A と B がともに帯行列であれば C もまた帯行列となるので、帯行列用の効率的な解法が使用できる。ピボット交換を行わない帯行列用の（改訂）コレスキ法では行列分解で帯幅は拡がらない。

レゾルベントの n 次多項式を作用させる処理の中では、同じ行列 C を係数に持つ連立 1 次方程式の組を解く処理が n 回現れる。連立 1 次方程式の組を右辺を変えて n 回解く処理では最初に行った行列分解の結果を再利用すると、2 回目以降は行列分解の計算を省ける。

いま X と Y が N 次ベクトル m 個の組 ($N \times m$ 行列) であるときに、フィルタ $\mathcal{F}$ を X に作用して Y を作る処理 $Y \leftarrow \mathcal{F}X$ の算法を図 3 に示す。この計算はチェビシェフ多項式の持つ 3 項漸化式を利用している (Z, V, W は作業用の $N \times m$ 行列である)。

3.6 下端固有対用のフィルタの伝達関数の設計例

下端固有対用のフィルタの伝達関数 $g(t)$ で単一のレゾルベントのチェビシェフ多項式で表されるものを設計した例を示す。正規化座標 t での伝達関数は $g(t) = g_s T_n(2x - 1)$ 、ただし $x \equiv (\mu + \sigma)/(t + \sigma)$ である。通過域は $t \in [0, 1]$ で

表 2 3 組として (n, μ, σ) を指定した場合の伝達関数の設計例
Table 2 Examples of designs of transfer functions when (n, μ, σ) is specified as a triplet.

n	μ	σ	g_p	g_s	グラフ
18	2.0	1.8	3.10×10^{-6}	8.53×10^{-15}	図 4
24	1.5	3.0	3.15×10^{-7}	3.75×10^{-14}	図 5
32	2.0	6.11	1.13×10^{-5}	1.45×10^{-15}	図 6

表 3 3 組として (μ, g_p, g_s) を指定した場合の伝達関数の設計例
Table 3 Examples of designs of transfer functions when (μ, g_p, g_s) is specified as a triplet.

μ	g_p	g_s	σ	n	修正後の g_p	修正後の g_s	グラフ
2.0	10^{-4}	3×10^{-13}	11.5	36	1.12×10^{-4}	4.27×10^{-13}	図 7
2.0	10^{-5}	10^{-15}	6.11	32	1.13×10^{-5}	1.45×10^{-15}	図 8
1.5	3×10^{-6}	10^{-12}	6.52	30	3.85×10^{-6}	1.74×10^{-12}	図 9

阻止域は $t \in [\mu, \infty)$ である。3 組 (n, μ, σ) を指定すると、それから簡単な式計算により g_p と g_s が求まる。3 組として (n, μ, σ) の値を指定して伝達関数を設計した例を表 2 に示す。また 3 組として (μ, g_p, g_s) の値を指定してそれから伝達関数を設計した例を表 3 に示す。

3.7 下端固有対に対するフィルタについてのまとめ

実対称定値一般固有値問題で、いま a の値を最小固有値 $\lambda_{\min}$ 以下とすると、フィルタ対角化法を用いて固有値が区間 $[a, b]$ にある下端固有対を求めるものとする。それに用いるフィルタを、シフトが実数である単一のレゾルベントの多項式として構成する方法について考察を行った。

いま a が最小固有値以下であるとき、レゾルベントのシフト ρ は最小固有値 $\lambda_{\min}$ よりも小さい実数になるので、レゾルベント $\mathcal{R}(\rho)$ の作用を実現する連立 1 次方程式の係数行列 $C = A - \rho B$ は実対称正定値となる。連立 1 次方程式を行列分解を用いて解く場合には、係数行列 C の分解を 1 回だけ行ってその分解の結果が保持できるのであれば、係数 C が不変で右辺だけが異なる連立 1 次方程式は C を新たに分解せずにすでに求めた行列分解の結果を再利用して解くことができる。そのため単一のレゾルベントの多項式を作用させる処理の中では多項式の次数 n と同じ回数だけ同一のレゾルベントを作用させるので、係数 C をピボット選択なしの（改訂）コレスキ法で 1 度だけ対称分解すれば良く、複数のレゾルベントを用いる場合に比べて行列分解に費やす計算量が少ない利点がある。

レゾルベントのチェビシェフ多項式である簡易型のフィルタは簡単な式を用いて設計できる。しかしその簡単さと引き換えに、得られるフィルタの特性は通過域に於ける伝達率の最大最小比が小さくない（つまり 1 に比べて g_s がある程度小さい値になる）ので、それだけ求める固有対の近似精度が不均一となる傾向を持つ。

実数をシフトとする単一のレゾルベントのチェビシェフ

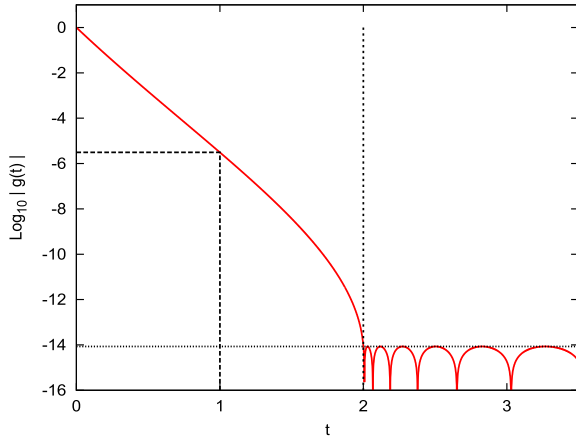


図 4 伝達関数の大きさ $|g(t)|$ ($n = 18$, $\mu = 2.0$, $\sigma = 1.8$), $g_p = 3.10 \times 10^{-6}$, $g_s = 8.53 \times 10^{-15}$

Fig. 4 Magnitude of transfer function $|g(t)|$ ($n = 18$, $\mu = 2.0, \sigma = 1.8$), $g_p = 3.10 \times 10^{-6}$, $g_s = 8.53 \times 10^{-15}$.

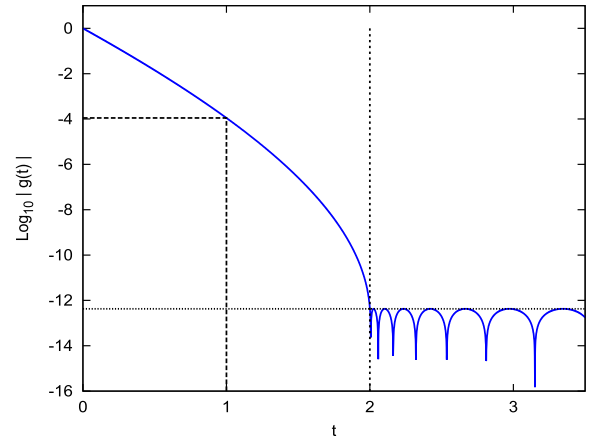


図 7 伝達関数の大きさ $|g(t)|$ ($\mu = 2.0$, $g_p = 1.12 \times 10^{-4}$, $g_s = 4.27 \times 10^{-13}$), $\sigma = 11.54$, $n = 36$

Fig. 7 Magnitude of transfer function $|g(t)|$ ($\mu = 2.0$, $g_p = 1.12 \times 10^{-4}$, $g_s = 4.27 \times 10^{-13}$), $\sigma = 11.54$, $n = 36$.

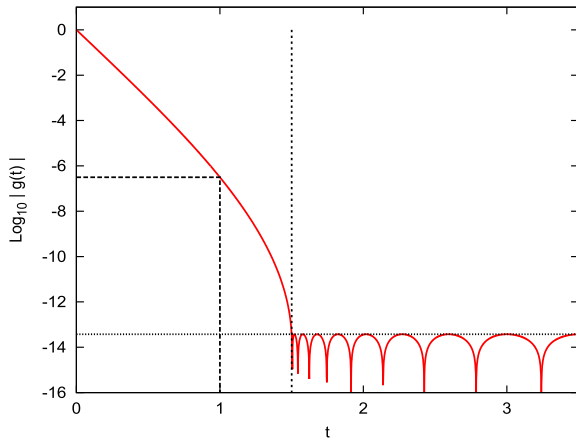


図 5 伝達関数の大きさ $|g(t)|$ ($n = 24$, $\mu = 1.5$, $\sigma = 3.0$), $g_p = 3.15 \times 10^{-7}$, $g_s = 3.75 \times 10^{-14}$

Fig. 5 Magnitude of transfer function $|g(t)|$ ($n = 24$, $\mu = 1.5$, $\sigma = 3.0$), $g_p = 3.15 \times 10^{-7}$, $g_s = 3.75 \times 10^{-14}$.

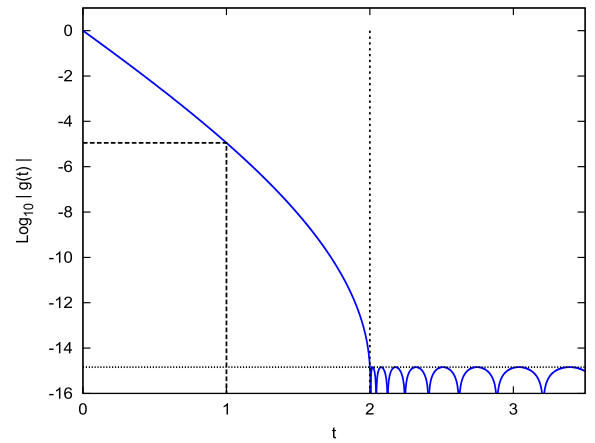


図 8 伝達関数の大きさ $|g(t)|$ ($\mu = 2.0$, $g_p = 1.13 \times 10^{-5}$, $g_s = 1.45 \times 10^{-15}$), $\sigma = 6.11$, $n = 32$

Fig. 8 Magnitude of transfer function $|g(t)|$ ($\mu = 2.0$, $g_p = 1.13 \times 10^{-5}$, $g_s = 1.45 \times 10^{-15}$), $\sigma = 6.11$, $n = 32$.

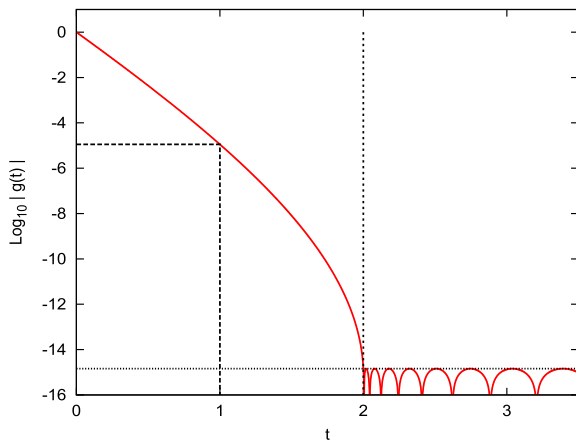


図 6 伝達関数の大きさ $|g(t)|$ ($n = 32$, $\mu = 2.0$, $\sigma = 6.11$), $g_p = 1.13 \times 10^{-5}$, $g_s = 1.45 \times 10^{-15}$

Fig. 6 Magnitude of transfer function $|g(t)|$ ($n = 32$, $\mu = 2.0, \sigma = 6.11$), $g_p = 1.13 \times 10^{-5}$, $g_s = 1.45 \times 10^{-15}$.

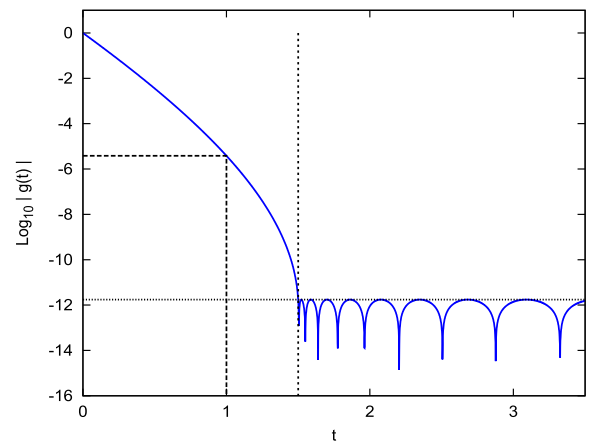


図 9 伝達関数の大きさ $|g(t)|$ ($\mu = 1.5$, $g_p = 3.85 \times 10^{-6}$, $g_s = 1.74 \times 10^{-12}$), $\sigma = 6.524$, $n = 30$

Fig. 9 Magnitude of transfer function $|g(t)|$ ($\mu = 1.5$, $g_p = 3.85 \times 10^{-6}$, $g_s = 1.74 \times 10^{-12}$), $\sigma = 6.524$, $n = 30$.

多項式として構成された簡易型のフィルタは、それを用いた対角化の計算を実験してみることで、ある程度うまく働くことを確認した (6 章)。

4. 中間固有対を求める場合のフィルタ

本章では、固有値が固有値分布の任意の位置にある固有対 (中間固有対) の近似を求めるためのフィルタを扱う。

いま行列 A と B が実対称で B は正定値の実対称定値一般固有値問題を $A\mathbf{v} = \lambda B\mathbf{v}$ とする。この問題の「中間固有対」で固有値が任意の区間 $[a, b]$ にあるものをフィルタ対角化法を用いて解くことにする。この場合はシフトとして虚数 ρ' を採用することで、レゾルベント $\mathcal{R}(\rho') = (A - \rho' B)^{-1} B$ は有界な作用素になることが (固有値分布によらずに) 保証される。さらにフィルタは実作用素 (実ベクトルから実ベクトルへの作用素) であるとし、「単一のレゾルベント」 $\mathcal{R}(\rho')$ だけを用いてその作用の (虚部の) 多項式として構成することにする。フィルタをその多項式の次数 n を決めて構成する場合には、レゾルベントのシフト ρ' と n 次多項式 $\mathcal{P}$ をうまく決めてフィルタの伝達関数ができるだけ良い特性を持つものとなるように設計することになる。

いま設計を簡単化するために伝達関数をチェビシェフ多項式で表される形に限定すれば、阻止域における伝達率の大きさの上限である g_s を微小にすることはチェビシェフ多項式の次数 n を上げれば容易に実現できるが、通過域における伝達率の最大最小比 $1/g_p$ を小さくすることは容易ではない (あるいは遷移域の幅 $\mu - 1$ を小さくすることは容易ではない)。ただしシフトが虚数である単一のレゾルベントを使う中間固有対用のフィルタの場合は、シフトが実数である単一のレゾルベントを使う下端固有対用のフィルタの場合に比べて、通過域における伝達関数の最大最小比を小さくできる。

また、レゾルベントの作用を連立 1 次方程式を解いて与える計算は、シフトが実数の場合には実数の演算だけでできるが、シフトが虚数の場合には複素数の演算を行う必要がある。

4.1 中間固有対用のフィルタ作用素と伝達関数

いま指定区間 $[a, b]$ に固有値 λ がある対を解くものとする。この中間固有対を求める問題の場合は、固有値の座標 λ に対する正規化座標 t を $\lambda \in [a, b]$ から標準区間 $t \in [-1, 1]$ への 1 次変換 $\lambda = \frac{a+b}{2} + \left(\frac{b-a}{2}\right)t$ を用いて定義する。いまフィルタ $\mathcal{F}'$ の伝達関数が $f'(\lambda)$ であるときに、正規化座標 t を引数とする伝達関数を $g'(t) \equiv f'(\lambda)$ と定義して、ある実数 $\mu' > 1$ により、 $t \in [-1, 1]$ は通過域、 $|t| \in [\mu', \infty)$ は阻止域、その中間の $|t| \in (1, \mu')$ は遷移域とする。この伝達関数 $g'(t)$ に課す制約条件は、1) 最大値が 1 であること、2) 通過域における最小値は g_p で、逆に $g'(t) \geq g_p$ となるのは通過域においてだけであること、3) さらに阻止域

における $|g'(t)|$ の上限値が g_s であること、とする。特に理由がなければ $g'(t)$ を偶関数にするのが便利なので、レゾルベントのシフト ρ' に虚数を用いる場合には互いに複素共役である $g'(t)$ の 1 対の極を純虚数にとる。

中間固有対用の場合も下端固有対用のフィルタの場合と同様に、伝達関数の 3 つの形状パラメタ μ' , g_p , g_s の値については以下のように考える。

- 遷移域と通過域の幅の形状比である μ' の値は 1 より大きいことが望ましい。 μ' が大きいと遷移域は広く、多くの固有値が遷移域に入る傾向が増す。固有値が遷移域にある固有ベクトルの数が増すと、それだけ多くのベクトルをフィルタで濾過することが必要となり、計算量や必要な記憶の量が増える。
- 阻止域における伝達率の大きさの上限 g_s は微小な正数であることが望ましい。そうでなければ固有値が阻止域にある不要な固有ベクトルが十分に除去されずにフィルタで濾過された結果に混入し、それで張られる不変部分空間の近似を悪くする。
- 通過域での伝達率の最大最小比の逆数 g_p は 1 以下の正数で、それが 1 に近いほど良い。もしも g_p の値が小さければ、求めた (固有値が通過域にある) 近似対の精度はそれだけ均一ではなくなる可能性がある。

ただしこれら 3 つの形状パラメタの値 μ' , g_p , g_s の改良の要求は互いに相反するので、釣り合いを考えて調整することになる。

最良の特性を持つフィルタの設計は、次数 n を固定した場合、複素数のシフト ρ' の値とフィルタを与える多項式を調整する数値最適化の問題になり、それを解くことは容易ではなく、しかも最適解が得られても単に数値を並べて示すしかないので、方法の説明や実装や普及には不便である。そこで以下では、多項式をチェビシェフ多項式の形に限定することでこの数値最適化を省いた「簡易な設計法」を示す。これは限定により伝達関数 $g'(t)$ の選択の範囲を狭めているので、得られる伝達関数の形状は最良なものではないが、形状を指定すると伝達関数が簡単な式の計算で決定できるという利点がある。

4.2 チェビシェフ多項式で表された伝達関数の設計

以前の 3 章で扱った下端固有対を求めるためのチェビシェフ多項式で表されたフィルタの伝達関数 $g(t)$ の表式 (15) から、関係 $g'(t) = g(t^2)$ により新たな関数 $g'(t)$ を定義する。すなわち n , $\mu > 1$, $\sigma > 0$ が与えられたときに $g'(t)$ は以下の式 (28) で与えられる (定数 g_s は $g'(t)$ の最大値を 1 に規格化することから決まる)。

$$g'(t) = g_s T_n(y), \quad y = 2x - 1, \quad x = \frac{\mu + \sigma}{t^2 + \sigma}. \quad (28)$$

すると $g(t)$ は $t \geq 0$ で定義される n 次の実有理関数であるので、この新しい関数 $g'(t)$ は偶関数で実軸全体で定義さ

れた t の $2n$ 次の実有理関数である。

いま $\mu' = \sqrt{\mu}$ において、元の伝達関数 $g(t)$ の通過域が $t \in [0, 1]$ 、遷移域が $t \in (1, \mu)$ 、阻止域が $t \in [\mu, \infty)$ であることに対応して、新しい伝達関数 $g'(t)$ の通過域を $|t| \in [0, 1]$ 、遷移域を $|t| \in (1, \mu')$ 、阻止域を $|t| \in [\mu', \infty)$ とする。そのとき関係 $g'(t) = g(t^2)$ により、これらの伝達関数 $g(t)$ と $g'(t)$ はそれぞれの通過域で最大値が 1 であり、それぞれの通過域で最小値は g_p であり、それぞれの阻止域で大きさの上限値が g_s であることが容易に分かる。つまり、伝達関数 $g(t)$ の 3 つの形状パラメタの組を (μ, g_p, g_s) とするとき、伝達関数 $g'(t)$ の 3 つの形状パラメタの組は (μ', g_p, g_s) になる、ただしここで $\mu' = \sqrt{\mu}$ である。つまり、伝達率の閾値 g_p と g_s は共通で、遷移域の幅に関するパラメタが μ から $\mu' = \sqrt{\mu}$ になっている。

そうして $g'(t)$ の極は 1 対の複素共役な純虚数 $t = \pm i\sqrt{\sigma}$ だけであり、それぞれ n 位の極である。ここで i は虚数単位 $\sqrt{-1}$ である。またチェビシェフ多項式の性質から、 $g'(t)$ の阻止域 $|t| \geq \mu'$ における制約条件 $|g'(t)| \leq g_s$ は最初から満たされ、阻止域の外部 $t \in [0, \mu']$ では単調減少になる。すると残りの 2 つの制約条件 $g'(0)=1$ と $g'(1)=g_p$ はそれぞれ以前と同様の式の組 (16) あるいは式の組 (17) になる。

4.2.1 3 つ組 (n, μ', σ) を指定する場合

いま 3 つ組 (n, μ', σ) が指定されたら、関係 $\mu = (\mu')^2$ により下端固有対を求める場合の伝達関数 $g(t)$ で 3 つ組 (n, μ, σ) の値が指定された場合へ帰着させる。つまり式の組 (18) の各右辺の値をそれぞれ計算すれば、閾値 g_p と g_s の値がただちに求まる。

求めたそれらの値 g_p と g_s が望ましいものでなければ、最初に与えた 3 つ組 n, μ', σ の値を変えて再度試行する。

4.2.2 3 つ組として (n, g_p, g_s) を指定する場合

いま (n, g_p, g_s) が指定された場合には、同様に下端固有対を求める場合の伝達関数 $g(t)$ で (n, g_p, g_s) の値が指定された場合に帰着させる。つまり式 (19), (20), (21) により、 σ と μ が求まるので、それから σ と $\mu' = \sqrt{\mu}$ の値を求めることができる。

4.2.3 形状パラメタの 3 つ組 (μ', g_p, g_s) を指定する場合

形状パラメタの 3 つ組 (μ', g_p, g_s) が指定されたときにそれを実現する σ と n の値を決定する場合についてもまず $\mu = (\mu')^2$ において、下端固有対を求めるためのフィルタの伝達関数 $g(t)$ に対して (μ, g_p, g_s) が指定された場合の処理 (すでに 3.3.3 項で記述) に帰着させる。

つまり、非線形方程式 $G(\sigma) = r$ (ただしその両辺は式の組 (23) の各式) が正の実数解 σ を持つ場合には、それを用いると 2 つの式 (22) の n を表す値は一致するが、その値は一般には整数値にはならないのでそれを適切に整数化して n として、 (n, μ, σ) の組を作り、それから逆に式の組 (18) の右辺の計算により新しい g_p と g_s の数値を求めて、最初に指定した g_p と g_s の値を修正することで辻褄を合わせる

方法をとる。修正の程度が許容できればそれで構わない。

4.3 伝達関数に対応するフィルタの構成法

チェビシェフ多項式を用いて表された式 (28) の関数 $g'(t)$ を伝達関数として持つフィルタ $\mathcal{F}'$ を、単一のレゾルベントを用いて構成する方法を示す。まず σ は正の実数で、固有値の正規化座標 t は実数であるから、以下の式 (29) が成り立つ。

$$\frac{1}{t^2 + \sigma} = \frac{1}{\sqrt{\sigma}} \operatorname{Im} \frac{1}{t - i\sqrt{\sigma}}. \quad (29)$$

ここで Im は複素数の虚部をとる作用を表す。さらに正規化座標 t と固有値の座標 λ の間の対応関係である $t = \frac{2}{b-a}(\lambda - a) - 1$ を用いると、 $\rho' \equiv \frac{a+b}{2} + i\left(\frac{b-a}{2}\right)\sqrt{\sigma}$ とおけば、以下の式 (30) が成り立つ。

$$\frac{1}{t - i\sqrt{\sigma}} = \left(\frac{b-a}{2}\right) \frac{1}{\lambda - \rho'}. \quad (30)$$

さらに表記を簡単化するために、 γ' を以下の式 (31) とおく (この γ' は実数である)。

$$\gamma' \equiv \frac{\mu + \sigma}{\sqrt{\sigma}} \left(\frac{b-a}{2}\right). \quad (31)$$

すると t を用いて表されているチェビシェフ多項式の引数の式を λ を用いて表せば以下の式 (32) になる。

$$2 \frac{\mu + \sigma}{t^2 + \sigma} - 1 = 2\gamma' \operatorname{Im} \frac{1}{\lambda - \rho'} - 1. \quad (32)$$

これに対応する線形作用素は $2\gamma' \operatorname{Im} \mathcal{R}(\rho') - I$ であるから、以下の式 (33) の伝達関数：

$$\begin{cases} g'(t) = g_s T_n \left(2 \frac{\mu + \sigma}{t^2 + \sigma} - 1 \right), \\ f'(\lambda) = g_s T_n \left(2\gamma' \operatorname{Im} \frac{1}{\lambda - \rho'} - 1 \right) \end{cases} \quad (33)$$

に対応するフィルタ $\mathcal{F}'$ は以下の式 (34) で与えられることが分かる。

$$\mathcal{F}' = g_s T_n (2\gamma' \operatorname{Im} \mathcal{R}(\rho') - I). \quad (34)$$

ここで $\mathcal{R}(\rho')$ はシフトが虚数 ρ' のレゾルベントであり、 I は恒等作用素を表す。

4.4 中間固有対用のフィルタの作用を与える算法

上述のように構成された伝達関数を持つ中間固有対用のフィルタ $\mathcal{F}'$ を、与えられた実の N 次列ベクトルを m 個並べた組 X ($N \times m$ 実行列) に作用させて、別の実の N 次列ベクトルを m 個並べた組 Y (これも $N \times m$ 実行列) を作る処理 $Y \leftarrow \mathcal{F}' X$ の算法は、たとえば図 10 のようにすれば実現できる。ただし V, W はそれぞれ作業用の実の列ベクトルの組 ($N \times m$ 実行列) であり、 Z は作業用の複素列ベクトルの組 ($N \times m$ 複素実行列) である。またレゾルベ

3 組 (n, μ', σ) と区間 $[a, b]$ を与えて

$$\rho' \leftarrow \frac{a+b}{2} + i \left(\frac{b-a}{2} \right) \sqrt{\sigma} ;$$

$$\gamma' \leftarrow \frac{(\mu')^2 + \sigma}{\sqrt{\sigma}} \left(\frac{b-a}{2} \right) ;$$

$$Z \leftarrow \mathcal{R}(\rho') X ;$$

$$V \leftarrow X ;$$

$$W \leftarrow 2 \gamma' \text{Im} Z - X ;$$

for $k := 2$ to n do begin

$$Z \leftarrow \mathcal{R}(\rho') W ;$$

$$Y \leftarrow 4 \gamma' \text{Im} Z - 2W - V ;$$

$$V \leftarrow W ;$$

$$W \leftarrow Y$$

end ;

$$Y \leftarrow g_s Y$$
図 10 中間固有対用フィルタの作用 $Y \leftarrow \mathcal{F}' X$ の算法Fig. 10 Algorithm of an application of the filter $Y \leftarrow \mathcal{F}' X$ for interior eigenvalues.

ント $\mathcal{R}(\rho')$ を作用させる処理 $Z \leftarrow \mathcal{R}(\rho') W$ は実際には与えられたベクトルの組 W からまず右辺ベクトルの組 BW を作り、次に行列 $C \equiv A - \rho' B$ を係数とする連立 1 次方程式の組 $CZ = BW$ を Z について解くことにより実現する。行列 A, B が実対称でシフト ρ' は複素数なので C は複素対称行列 ($C^T = C$) である。さらに、もしも A と B が帯行列の場合には C も帯行列となり、連立 1 次方程式は複素数版の帯行列用の (改訂) コレスキ法を用いて能率よく解ける (ピボット選択をしない場合に精度の面で支障がないかについては、今後の検討を要する)。連立 1 次方程式の右辺の組 BW は実であるが、係数行列 C や解の組 Z は複素である。 n 次の多項式を用いたフィルタの実現には、行列 C を係数とする連立 1 次方程式の組を解く処理が全部で n 回含まれているが、行列分解を用いて連立 1 次方程式を解く場合には、最初に C の行列分解を 1 度だけ行ってその結果を保持できると、その後は連立 1 次方程式の組を分解計算は行わずに前進消去と後退代入だけで少ない演算量により解くことができる。

4.5 チェビシェフ多項式で表された伝達関数の例

チェビシェフ多項式で表される形の伝達関数 $g(t)$ を実際にパラメータを指定して設計し構成した例をいくつか示す。通過域は $t \in [-1, 1]$ で、阻止域は $|t| \in [\mu', \infty)$ である。伝達関数の極の位置は純虚数 $t = \pm i\sqrt{\sigma}$ である。各図のグラフでは伝達率の大きさ $|g(t)|$ の常用対数の値を赤い曲線で、 t の偶関数であるから右半分 $t \geq 0$ だけをプロットした。黒い補助線は、通過域の端 $t = 1$ とそのときの伝達率の値 g_p 、それと阻止域の端 $t = \mu$ とそのときの伝達率の値

表 4 3 組として (n, μ', σ) を指定した場合の伝達関数の設計例
Table 4 Examples of designs of transfer functions when (n, μ', σ) is specified as a triplet.

n	μ'	σ	g_p	g_s	グラフ
10	2.0	1.0	2.64×10^{-4}	5.78×10^{-13}	図 11
15	2.0	2.25	6.38×10^{-4}	9.71×10^{-15}	図 12
20	1.5	2.25	7.41×10^{-6}	9.77×10^{-16}	図 13
20	1.5	4.0	2.08×10^{-4}	1.82×10^{-12}	図 14
30	1.5	9.0	3.10×10^{-4}	5.78×10^{-13}	図 15
40	2.0	25.0	1.08×10^{-2}	5.62×10^{-14}	図 16

表 5 3 組として (μ', g_p, g_s) を指定した場合の伝達関数の設計例
Table 5 Examples of designs of transfer functions when (μ', g_p, g_s) is specified as a triplet.

μ'	g_p	g_s	σ	n	修正後 g_p	修正後 g_s	グラフ
1.5	10^{-4}	3×10^{-13}	3.48	20	1.03×10^{-4}	3.32×10^{-13}	図 17
2.0	10^{-2}	3×10^{-13}	12.0	26	1.17×10^{-2}	8.23×10^{-13}	図 18

g_s を示している。

パラメタの 3 組として (n, μ', σ) を指定して伝達関数を設計した例を表 4 に示す。またパラメタの 3 組として (μ', g_p, g_s) を指定した場合の例を表 5 に示す (この場合はまず方程式を解いて σ を得て、次に n を表す式を用いて求めた実数値を整数値に切り捨てて次数 n にして、そうして得られた (n, μ', σ) の値からそれと整合する g_p, g_s の値を計算しなおすことにより最初に指定されたそれらの値を修正している)。

4.6 中間固有対に対するフィルタについてのまとめ

実対称定値一般固有値問題の固有対で (実軸上の任意の位置にある) 区間 $[a, b]$ に固有値があるものをフィルタ対角化法を用いて求める方法について考察した。それに使用するフィルタを、シフトが虚数である単一のレゾルベントの虚部の多項式として構成する方法について考察を行った。

レゾルベントのシフト ρ' が虚数なので、レゾルベントの作用を実現する連立 1 次方程式の係数行列 $C = A - \rho' B$ は A と B が実対称であるから正則な複素対称行列になる。そうして A と B がともに帯行列の場合には C も帯行列になる。その C を係数行列とする連立 1 次方程式を行列分解を用いて解く場合は、行列分解の結果を保持できるのであれば、分解を最初に 1 回だけ行ってそれを再利用すれば、係数 C が共通の連立 1 次方程式を素早く解くことができる。そのため「単一のレゾルベント $\mathcal{R}(\rho')$ の虚部」の多項式を作用させる処理の中では多項式の次数と等しい回数だけ毎回同一のレゾルベントを作用させるので、複素版の (改訂) コレスキ分解を用いて正則な複素対称行列を 1 回だけ対称分解すれば良いので、複数のレゾルベントを用いる場合に比べて行列分解に費やす計算量が少ないという利点がある。

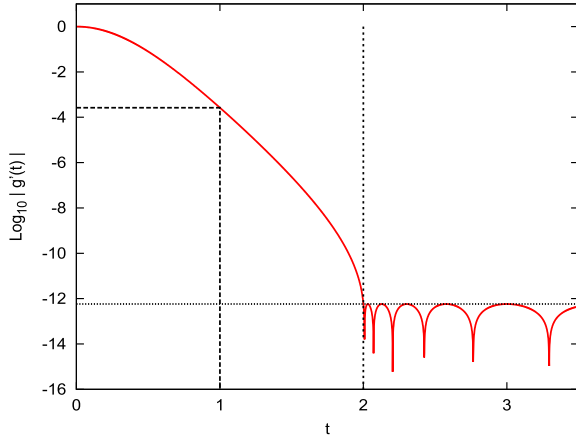


図 11 伝達関数の大きさ $|g'(t)|$ (右半分) ($n = 10$, $\mu' = 2.0$, $\sigma = 1.0$); $g_p = 2.64 \times 10^{-4}$, $g_s = 5.78 \times 10^{-13}$

Fig. 11 Magnitude of transfer function $|g'(t)|$ (right-half) ($n = 10$, $\mu' = 2.0$, $\sigma = 1.0$); $g_p = 2.64 \times 10^{-4}$, $g_s = 5.78 \times 10^{-13}$.

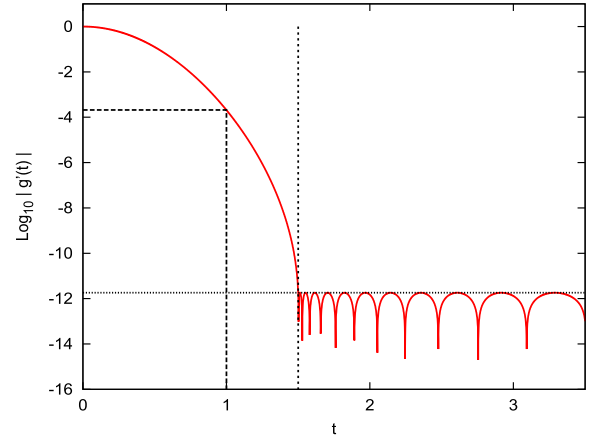


図 14 伝達関数の大きさ $|g'(t)|$ (右半分) ($n = 20$, $\mu' = 1.5$, $\sigma = 4.0$); $g_p = 2.08 \times 10^{-4}$, $g_s = 1.82 \times 10^{-12}$

Fig. 14 Magnitude of transfer function $|g'(t)|$ (right-half) ($n = 20$, $\mu' = 1.5$, $\sigma = 4.0$); $g_p = 2.08 \times 10^{-4}$, $g_s = 1.82 \times 10^{-12}$.

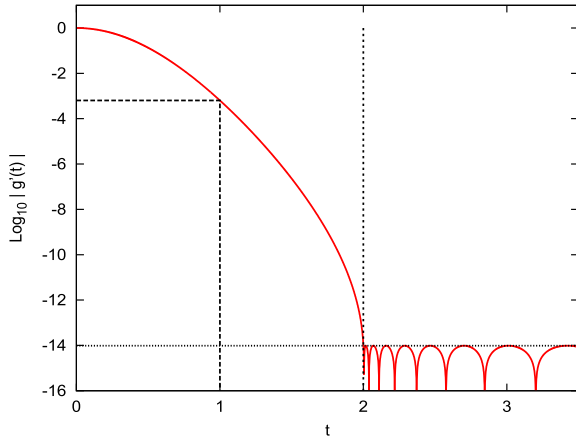


図 12 伝達関数の大きさ $|g'(t)|$ (右半分) ($n = 15$, $\mu' = 2.0$, $\sigma = 2.25$); $g_p = 6.38 \times 10^{-4}$, $g_s = 9.71 \times 10^{-15}$

Fig. 12 Magnitude of transfer function $|g'(t)|$ (right-half) ($n = 15$, $\mu' = 2.0$, $\sigma = 2.25$); $g_p = 6.38 \times 10^{-4}$, $g_s = 9.71 \times 10^{-15}$.

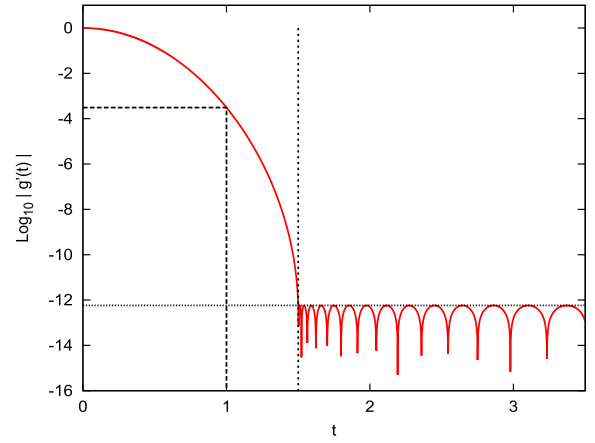


図 15 伝達関数の大きさ $|g'(t)|$ (右半分) ($n = 30$, $\mu' = 1.5$, $\sigma = 9.0$); $g_p = 3.10 \times 10^{-4}$, $g_s = 5.78 \times 10^{-13}$

Fig. 15 Magnitude of transfer function $|g'(t)|$ (right-half) ($n = 30$, $\mu' = 1.5$, $\sigma = 9.0$); $g_p = 3.10 \times 10^{-4}$, $g_s = 5.78 \times 10^{-13}$.

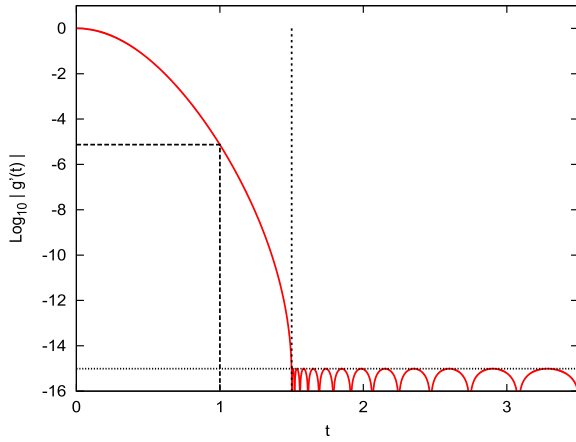


図 13 伝達関数の大きさ $|g'(t)|$ (右半分) ($n = 20$, $\mu' = 1.5$, $\sigma = 2.25$); $g_p = 7.41 \times 10^{-6}$, $g_s = 9.77 \times 10^{-16}$

Fig. 13 Magnitude of transfer function $|g'(t)|$ (right-half) ($n = 20$, $\mu' = 1.5$, $\sigma = 2.25$); $g_p = 7.41 \times 10^{-6}$, $g_s = 9.77 \times 10^{-16}$.

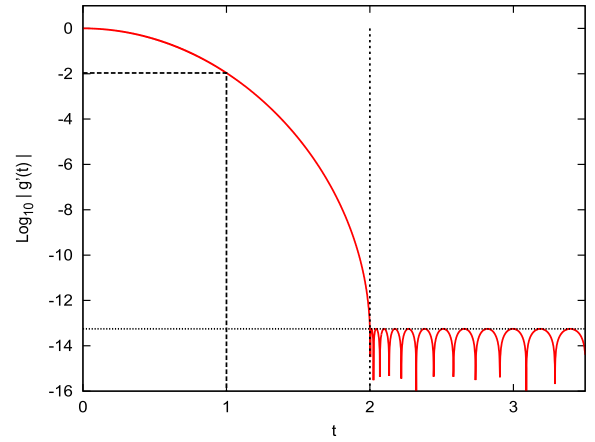


図 16 伝達関数の大きさ $|g'(t)|$ (右半分) ($n = 40$, $\mu' = 2.0$, $\sigma = 25.0$); $g_p = 1.08 \times 10^{-2}$, $g_s = 5.62 \times 10^{-14}$

Fig. 16 Magnitude of transfer function $|g'(t)|$ (right-half) ($n = 40$, $\mu' = 2.0$, $\sigma = 25.0$); $g_p = 1.08 \times 10^{-2}$, $g_s = 5.62 \times 10^{-14}$.

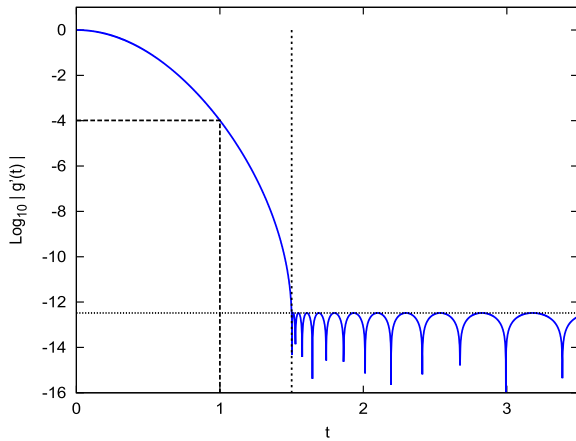


図 17 伝達関数の大きさ $|g'(t)|$ (右半分) ($\mu' = 1.5$, $g_p = 1.03 \times 10^{-4}$, $g_s = 3.32 \times 10^{-13}$; $\sigma = 3.48$, $n = 20$)

Fig. 17 Magnitude of transfer function $|g'(t)|$ (right-half) ($\mu' = 1.5$, $g_p = 1.03 \times 10^{-4}$, $g_s = 3.32 \times 10^{-13}$; $\sigma = 3.48$, $n = 20$).

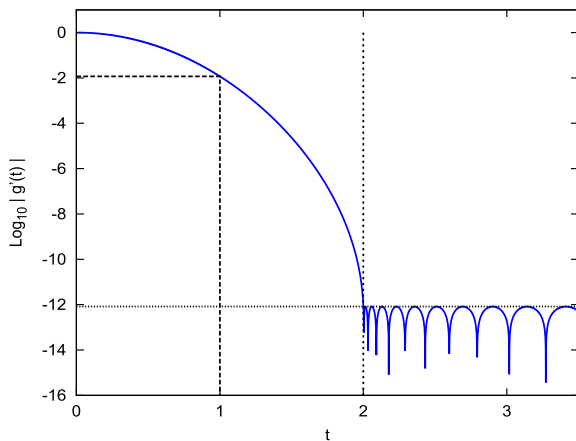


図 18 伝達関数の大きさ $|g'(t)|$ (右半分) ($\mu' = 2.0$, $g_p = 1.17 \times 10^{-2}$, $g_s = 8.23 \times 10^{-13}$; $\sigma = 12.04$, $n = 26$)

Fig. 18 Magnitude of transfer function $|g'(t)|$ (right-half) ($\mu' = 2.0$, $g_p = 1.17 \times 10^{-2}$, $g_s = 8.23 \times 10^{-13}$; $\sigma = 12.04$, $n = 26$).

レゾルベントの虚部のチェビシェフ多項式として構成されるフィルタは簡単な式を用いて設計できる。しかし得られるフィルタの特性はその簡単さと引き換えに、通過域における伝達率の最大最小比が小さくないため、得られる近似対の精度がそれだけ均一ではなくなる傾向を持つ（ただし下端固有対用のフィルタの場合に比べると、通過域における伝達率の最大最小比を小さく抑えることができる。そのことは、中間固有対用の伝達関数 $g'(t)$ でその形状パラメタの 1 つ μ' を指定したときに、下端固有対用の伝達関数 $g(t)$ で $\mu = (\mu')^2$ を形状パラメタとして持つものと考え、 μ の値が大きいため、通過域における伝達率の最大最小比を小さくできる（つまり g_p をより 1 に近づけることができる）。もちろん μ が大きければ $g(t)$ は遷移域の幅 $\mu - 1$ が広いので実用性の低いフィルタとなるが、対応する $g'(t)$ については遷移域の幅は $\mu' - 1 = \sqrt{\mu} - 1$ である）。

虚数をシフトとする単一のレゾルベントの虚部のチェビシェフ多項式として構成される簡易型のフィルタは、それを用いて対角化の計算を実験してみることで、ある程度うまく働くことを確認した（7 章）。

5. 実験について

5.1 計算に用いたシステム

CPU は Intel Core i7-5960X（8 コア、クロック 3 GHz、L3 キャッシュ 20 MiB、AVX2 命令）で、Hyper-Thread 機能と Turbo-Boost 機能は BIOS のメニューの設定でオフにした。メモリは DDR4-2133 MHz（PC4-17000）でクアドチャネル、8 GiB のモジュール 8 個で主記憶容量は 64 GiB である。OS は CentOS 7 for x86_64（64 bit 版）である。

5.2 使用したプログラム

作成したプログラムは Fortran90 を用いてコーディングしたものに OpenMP のディレクティブを挿入した。コンパイラは Intel Fortran v15.0.0 for x86_64 であり、コンパイラのオプションとして `"-fast -openmp"` を指定して OpenMP による 8 スレッド並列で実行した。

計算の主要部分は、行列 X と Y が実あるいは複素の N 次の縦ベクトルを列として m 個並べた $N \times m$ 行列とするとき、与えられた X に対してシフトが ρ のレゾルベント $\mathcal{R}(\rho)$ を適用して $Y \leftarrow \mathcal{R}(\rho) X$ を求めるところである。いま N 次の実対称行列 A と B に対して、シフト ρ のレゾルベントは $\mathcal{R}(\rho) \equiv (A - \rho B)^{-1} B$ なので、シフト行列を $C \equiv A - \rho B$ とすると、 $Y \leftarrow \mathcal{R}(\rho) X$ の計算は、 X を与えて Y についての連立 1 次方程式 $CY = BX$ を解くことに帰着する。係数行列 C はシフト ρ が実数であるかまたは虚数であるかによって実対称行列あるいは複素対称行列になるが、さらに以下の例題のように A と B が帯行列であれば C もまた帯行列となる。

今回のプログラムでは、実対称あるいは複素対称の帯行列を係数とする連立 1 次方程式の解法として、領域分割法やブロック化手法などは用いずに、古典的な BLAS-2 レベルの帯行列用の改訂コレスキ法を用いている。それを Fortran90 で記述したものに OpenMP の指示を挿入している。これは現段階ではフィルタ対角化の手法の数値的な性質の検証を目的としていて、使用した計算機システムの持つ性能を高度に絞り出すことまでは目標にはしていないからである。今回の実験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。与えられた計算機システムのうえで実対称あるいは複素対称の帯行列を係数とする連立 1 次方程式を高効率で解く方法については、また別に研究すべき課題になるであろう。

連立 1 次方程式 $CY = BX$ を解くためには、まず X に帯行列 B を乗じて右辺を作る処理 $Z \leftarrow BX$ が必要。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 $C \rightarrow LDL^T$

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って Y を求める処理 $Y \leftarrow L^{-T}(D^{-1}L^{-1}Z)$ が要る。複素対称版の改訂コレスキ法は通常の実対称版に現れる数値と四則演算をそのまま複素数のものに置き換えて得られるものである（複素エルミート行列用のものとは異なる）。

注 1：シフト ρ が実数で C が実対称正定値の場合には、ピボット選択をしない改訂コレスキ法の計算は数値安定性が保証されるが、シフトが虚数で C が複素対称の場合にはそのような保証はない。そのため C が複素対称の場合には、行列の対称性を用いずに片側ピボット選択付きの帯行列用の LU 分解法を採用すると計算の数値安定性が保証できる。しかしそのようにすると行列分解結果である L と U の記憶量の合計は（行選択により上帯幅が 2 倍まで拡がり得るので）改訂コレスキ法の場合に比べて 3 倍になり、また計算に用いる演算量もかなり増えるので、今回の実験例では係数行列 C が複素対称の場合には連立 1 次方程式をピボット選択なしの帯行列用の複素対称版の改訂コレスキ法を用いて解いている。

注 2：今回の実験では Y に対する連立 1 次方程式 $CY = BX$ を解く際に、改訂コレスキ法による帯行列の対称分解 $C \rightarrow LDL^T$ を終えた後の前進消去と後退代入を行う処理では、 m 個のベクトルの組をまとめて扱うことにより、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 の対称分解により得た帯行列 L の記憶に対する参照量を節約している。

5.3 テストに用いた一般固有値問題の例題

テストに用いた例題の実対称定値一般固有値問題 $A\mathbf{v} = \lambda B\mathbf{v}$ は、1 辺の長さが π の 3 次元立方体領域における零-ディリクレ境界条件での（逆符号の）ラプラシアン $-\nabla^2$ の固有値問題を有限要素法を用いて離散化近似をすることで得られたものである。用いた要素分割は立方体領域を各辺の方向にそれぞれ個数 $N_1 + 1$, $N_2 + 1$, $N_3 + 1$ の小区間に等間隔に分割する（図 19）。

そうして各要素内での展開基底には各辺方向の 3 重線形関数を用いる。この有限要素法の離散化で得られる行列の A と B の次数は $N = N_1 N_2 N_3$ で、行列の帯幅がなるべく小さくなるように（ $N_1 \leq N_2 \leq N_3$ として）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 $w_L = 1 + N_1 + N_1 N_2$ となる。

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（ $1 \leq i_k \leq N_k$, $k=1,2,3$ ）。すると 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。そうして 3 重線形関数に対してその

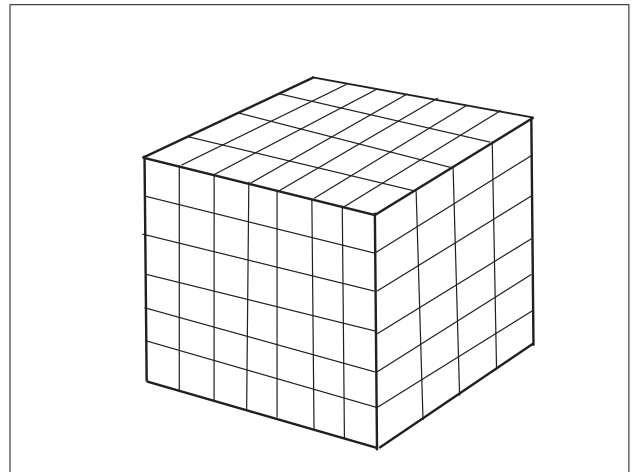


図 19 立方体の FEM 分割の概念図。分割例 $(N_1, N_2, N_3) = (3, 5, 6)$

Fig. 19 Conceptual figure of subdivision of a cube into finite elements. For the case of subdivision $(N_1, N_2, N_3) = (3, 5, 6)$.

3 重添字から決まる $i_1 + N_1(i_2 - 1) + N_1 N_2(i_3 - 1)$ の値により番号付けを行い、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、フィルタ対角化法を適用して（固有値分布の下端側である）区間 $[a, b]$ に固有値 λ が含まれる固有対の近似を求めた。この一般固有値問題のテスト例題では、固有値の厳密値は簡単な数式の計算で求めることができる（次節 5.4）。またそのことから固有値が区間 $[a, b]$ にある固有対の個数も、厳密な固有値を列挙してその値を大小順に並べて区間に入るものを数えることで容易に求められる。

5.4 テスト例題の固有値の厳密値を与える式

1 次元の場合の区間 $[0, \pi]$ における零-ディリクレ境界条件の 1 次元ラプラシアン（ただし符号が逆のもの $-\nabla^2$ ）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、有限要素法の基底関数を区間 $[0, \pi]$ を $N + 1$ 等分した幅 $h = \pi / (N + 1)$ の小区間内の区分線形関数とする場合には、 $\theta_k \equiv hk$ とおくと以下の式 (35) で与えられる。

$$E(N, k) = \frac{k^2 \left(\frac{\sin \theta_k}{\theta_k} \right)^2}{\frac{(1 + \cos \theta_k)(2 + \cos \theta_k)}{6}}, \quad k=1, 2, \dots, N. \quad (35)$$

そうして 3 次元立方体領域 $[0, \pi]^3$ で各方向の区間をそれぞれ $N_1 + 1$, $N_2 + 1$, $N_3 + 1$ に等分割して、1 次元の場合の各方向に対する基底関数の直積（3 重線形関数）を 3 次元の基底関数として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 次元ラプラシアン（ $-\nabla^2$ ）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

固有値の和の形として式 $E(N_1, k_1) + E(N_2, k_2) + E(N_3, k_3)$ により与えられる。

5.5 用いたフィルタ対角化法の概要

実験に用いたフィルタ対角化法では、まず最初に乱数でランダムな m 個のベクトルの組 ($N \times m$ 行列) を生成して、それを元にして B -正規直交化された m 個のベクトルの組 X を作る ($X^T B X = I_m$)。そうしてその X に区間 $[a, b]$ を通過域とするフィルタ $\mathcal{F}$ を作用させることで 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 [11]。その「不変部分空間」は、固有値が区間 $[a, b]$ の近傍にあるすべての固有ベクトル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レイリー・リッツ法を適用して得られたリッツ対を求めてそれ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改良の後に）、近似固有値が指定区間 $[a, b]$ にないものは棄却する。

5.6 近似対の品質の評価法

もしも近似対の品質を確認するために「逆反復」を用いて近似対の改良を行い、改良によるその変化の大きさから固有値あるいは固有ベクトルの誤差を推定を試みるものとすれば、各近似対の固有値 λ_i ごとに行列 $A - \lambda_i B$ を係数とする連立 1 次方程式を解く必要があり、近似対の数が多ければ大規模問題の場合には必要な演算の量が多く計算には長い時間が掛かり実用的ではない（さらにまた固有値に近接や縮重がある場合にはそれに対する配慮も要る）。

そこで各近似対ごとに元の固有値方程式に対する残差のノルム、あるいは残差のノルムの相対比である相対残差の値を計算することにする。

5.6.1 近似固有対の相対残差 Θ

今回の各実験例では得られた近似固有対 $(\lambda, \mathbf{v})$ の品質を評価する方法として、以下の式 (36) で表された相対残差 Θ を用いた。

$$\Theta \equiv \frac{\|A\mathbf{v} - \lambda B\mathbf{v}\|}{\|\lambda B\mathbf{v}\|} = \frac{\|\mathbf{r}\|}{\|\lambda B\mathbf{v}\|}. \quad (36)$$

この式の中で 2 カ所に現れているベクトルのノルム $\|\cdot\|$ にはたとえば 2-ノルム（ユークリッドノルム）のように容易に計算できるものを採用する。この Θ の値はベクトル $\mathbf{v}$ の規格化には依らず、また行列 A と B に共通の定数が乗じられた場合にも変化しない。

ベクトルに 2-ノルムを用いた場合の相対残差 Θ の幾何学的な意味は、いま N 次元空間内の 2 つのベクトル $A\mathbf{v}$ と $\lambda B\mathbf{v}$ の挟む角を θ とするとき、 $\Theta = 2 \sin \frac{\theta}{2}$ であり、 θ が微小角の場合には $\Theta \approx \theta$ である。

この Θ の値を求める計算の主要部は、ベクトル $\mathbf{v}$ に対して行列 A や B を乗じる処理であり、 A と B の疎性が高いほど計算が楽になり、さらに複数のベクトルをまとめて A や B に乗じれば A と B の記憶全体に対する参照はそれぞれ 1 回ずつになるので、きわめて効率的に計算できる。

5.7 各計算実験の結果のグラフについて

各計算実験の結果については、それぞれ 4 種類のグラフを示している。

- 「各 m に対する β の固有値の分布」のグラフは、上述の X と Y から作った m 次の (対称) 行列 $\beta = X^T B Y$ の固有値の分布を減少順に並べて横軸にその順位をとり、縦軸に β の固有値の大きさ（絶対値）を常用対数でプロットしたものである。そうして阻止域でのフィルタの伝達率の最大値を g_s とし、倍精度計算のマシンイプシロンを $\epsilon_{\text{MAC}} \approx 2.22 \times 10^{-16}$ とするとき、 g_s の 10 倍の値と ϵ_{MAC} の 100 倍の値の大きい方の値を β の固有値に対する閾値として設定したので、グラフ中にその位置に水平に黒い点線 (THRESHOLD) を描いている。

- 「各 m に対する $|\phi|$ の値の分布」のグラフは、それぞれの m の場合について、 m 次の実対称行列 $\alpha = Y^T B Y$ により、 m 次の一般固有値問題 $\alpha \mathbf{u} = \phi \beta \mathbf{u}$ を考えて、その m 次の一般固有値問題に対して、 β の固有空間を固有値の絶対値の小さい部分を切断して解いて得られた固有値 ϕ の分布を減少順に、横軸にその順位をとり、縦軸に大きさ $|\phi|$ を常用対数でとってプロットしたものである。

この ϕ の値は Y で張られる空間に含まれる元の固有値問題の固有対 $(\lambda, \mathbf{v})$ のベクトル $\mathbf{v}$ に対するフィルタの伝達率 $f(\lambda)$ の近似値である。グラフに描いた黒い水平線は、通過域における伝達率の最小値 g_p を示すが、この g_p から減少方向に探索を始めて ϕ の値分布の中に見つかる比較的広い空隙のある最初の場所の値を ϕ の閾値として採用している。

- 「各 m に対する近似対の相対残差 Θ 」のグラフは、それぞれの m の場合について、横軸には近似対の固有値をとり、縦軸には相対残差 Θ の値を常用対数でとってプロットしたものである。
- 「各 m に対する近似固有値の絶対誤差」のグラフは、それぞれの m の場合について、横軸には近似対の固有値をとり、縦軸には近似固有値の厳密値との差の大きさを常用対数でとってプロットしたものである。

6. 下端固有対を求める実験の例

3 章で記述した方法により、フィルタに用いる多項式としてチェビシェフ多項式を採用する簡易構成法を用いて、下端固有対に対するフィルタをシフトが最小固有値未満の

実数である単一のレゾルベントの多項式として具体的に構成した．それを用いて対角化の実験を行い，固有値が固有値分布の下端にある固有対の近似がうまく求まるかについて調べた．以下にその実験結果の例を示す．

6.1 例題 L (次数 50 万 4 千の大規模問題)

この例題 L で扱う一般固有値問題は，有限要素法の離散

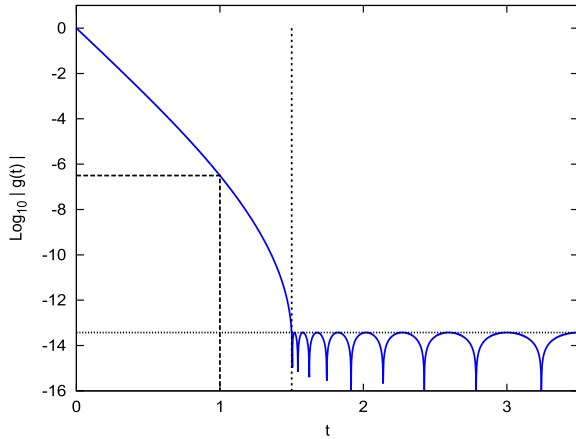


図 20 (例題 L) フィルタの伝達関数の大きさ $|g(t)|$ ($n = 24$, $\mu = 1.5$, $\sigma = 3.0$) $g_p = 3.15 \times 10^{-7}$, $g_s = 3.75 \times 10^{-14}$

Fig. 20 (Exam L) Magnitude of transfer function $|g(t)|$ of the filter ($n = 24$, $\mu = 1.5$, $\sigma = 3.0$) $g_p = 3.15 \times 10^{-7}$, $g_s = 3.75 \times 10^{-14}$.

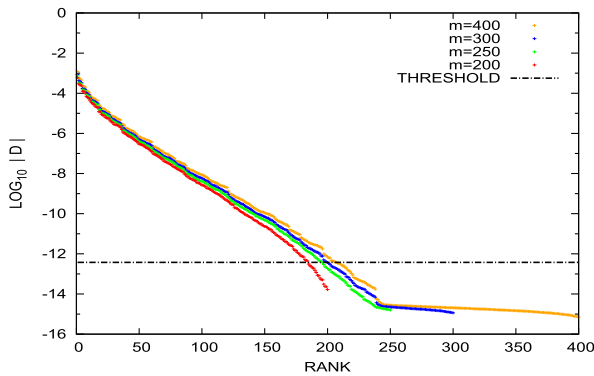


図 21 (例題 L) 各 m に対する β の固有値の分布

Fig. 21 (Exam L) Eigenvalue distribution of β for each m .

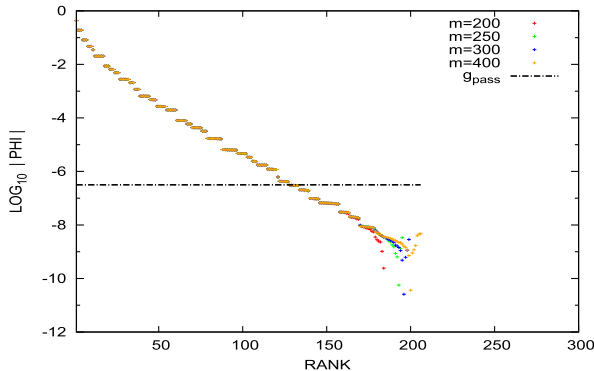


図 22 (例題 L) 各 m に対する $|\phi|$ の値の分布

Fig. 22 (Exam L) Distribution of values of $|\phi|$ for each m .

化の要素分割数を $(N_1, N_2, N_3) = (70, 80, 90)$ としたものである．行列 A と B の次数は $N = 504,000$ であり，下帯幅は $w_L = 5,671$ である．求める固有対の固有値の区間は $[a, b] = [0, 50]$ とした．この区間に固有値がある真の固有対の数は 127 である．

この例題 L で用いるフィルタの伝達関数 $g(t)$ のパラメタの 3 つ組の値は $(n = 24, \mu = 1.5, \sigma = 3.0)$ と設定した．それから決まる閾値は $g_p = 3.14759 \times 10^{-7}$ と $g_s = 3.75222 \times 10^{-14}$ である．伝達関数 $g(t)$ の大きさのグラフを図 20 に示す．

固有値の区間 $[a, b] = [0, 50]$ を通過域とするフィルタは，シフト ρ の値を -150.0 とし，係数 γ の値を 225.0 とするとき， $\mathcal{F} = g_s T_n(2\gamma \mathcal{R}(\rho) - I)$ である．

フィルタ対角化法で各近似対を得た後に，結果の品質の評価用に各近似対についての残差のノルムと近似固有値の絶対誤差を求めている．

フィルタを適用するベクトルの数をそれぞれ $m=200$, $m=250$, $m=300$, $m=400$ とした場合の各結果をまとめてグラフに示す (図 21, 図 22, 図 23, 図 24)．

6.1.1 (例題 L) 各 m に対する β の固有値の分布

図 21 は，各 m の場合について， β の固有値分布を減少

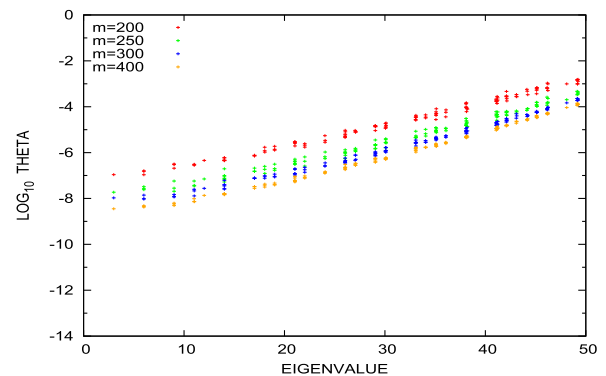


図 23 (例題 L) 各 m に対する近似対の相対残差のノルム Θ

Fig. 23 (Exam L) Values of Θ , norms of relative residuals of approximate pairs for each m .

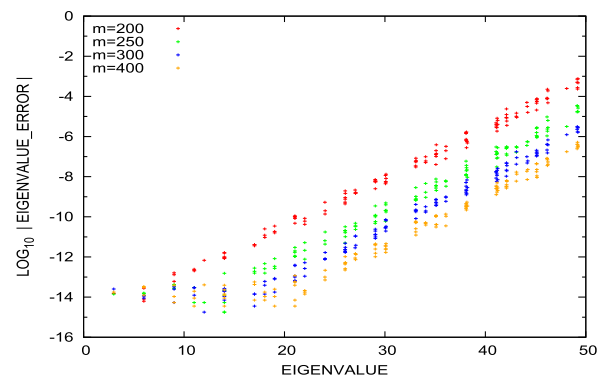


図 24 (例題 L) 各 m に対する近似固有値の絶対誤差

Fig. 24 (Exam L) Absolute errors of approximate eigenvalues for each m .

順に並べてその大きさの常用対数の値をプロットしたグラフである。

$m = 200$ の場合には、 β の固有値の分布は設定した閾値 ($10g_s = 3.752 \times 10^{-13}$) により切断を受けて階数は 16 だけ低下して 184 となったのでベクトルの数は $m = 200$ ですでに十分であることを示している。

同様に $m = 250$ では階数は 55 だけ低下して 195 に、 $m = 300$ では階数は 101 だけ低下して 199 に、 $m = 400$ では階数は 194 だけ低下して 206 にそれぞれなった。

6.1.2 (例題 L) 各 m に対する $|\phi|$ の値の分布

図 22 は、各 m の場合について、 $|\phi|$ の値の分布を減少順に並べてその常用対数の値をプロットしたグラフである。各 m の場合についてのプロット点が非常によく重なって見えていく。

$m = 200$ の場合には、 $|\phi|$ の値の分布に基づいて決めた閾値より上側にある ϕ の値の数として決まる不変部分空間の近似空間の基底ベクトルの数は 127 であった。それから求めた 127 個の近似対の固有値はどれも区間 $[0, 50]$ にあり、真の固有対の数 127 と同数の近似対が得られた。

そうして $m = 250$, $m = 300$, $m = 400$ の各場合にも、閾値より上側にある ϕ の値の数は同じ 127 であり、区間 $[0, 50]$ に固有値のある近似対が真の固有対の数と同数 127 個得られた。

6.1.3 (例題 L) 各 m に対する近似対の相対残差 Θ

図 23 は、各 m の場合について、横軸に求めた近似対の固有値をとり、縦軸に近似対の相対残差 Θ の常用対数の値をプロットしたグラフである。

$m = 200$ の場合には、固有値の区間 $[0, 50]$ のうち左端では相対残差の値 Θ は 10^{-7} 程度であるが、右端では値は 3×10^{-3} 程度であり、4 桁半程度の開きがある。この偏りは用いた伝達関数の通過域における値の非一様さが左端と右端で 6.5 桁もあることに由来するのであろう。

$m = 250$ の場合には、区間の左端では相対残差の値は 10^{-8} 程度であるが、右端では 3×10^{-4} 程度であり、4 桁半程度の開きがある。

$m = 300$ の場合には、区間の左端では相対残差の値は 10^{-8} 程度であるが、右端では 3×10^{-4} 程度であり、4 桁半程度の開きがある。

$m = 400$ の場合には、区間の左端では相対残差の値は 3×10^{-9} 程度であるが、右端では 10^{-4} 程度であり、4 桁半程度の開きがある。

6.1.4 (例題 L) 各 m に対する近似固有値の絶対誤差

図 24 は、各 m の場合について、横軸に求めた近似対の固有値をとり、縦軸に近似固有値の真の値からの絶対誤差の常用対数の値をとってプロットしたグラフである。

$m = 200$ の場合、固有値は区間 $[0, 50]$ の左端では絶対誤差が 10^{-14} 程度で相対精度は 14 桁程度であるが、右端では絶対誤差は 10^{-3} 程度で相対精度は 4 桁半程度であるこ

表 6 (例題 L) 経過時間の内訳 (単位: 秒)

Table 6 (Exam L) Breakdown of elapsed time (in second).

処理ステップ	$m=200$	$m=250$	$m=300$	$m=400$
フィルタ対角化全体	3,947.8	4,645.1	5,205.4	6,623.9
・乱数入力ベクトル生成	0.4	0.6	0.6	0.8
・入力ベクトル B-正規直交化	36.9	47.1	56.3	75.6
・フィルタ作用適用	3,619.0	4,270.6	4,789.5	6,138.8
・不変部分空間の基底作成	184.4	208.6	230.9	148.0
・レイリー・リッツ法適用	107.2	118.4	128.1	260.8

とが分かる。

$m = 250$ の場合、固有値は区間 $[0, 50]$ の左端では絶対誤差が 10^{-14} 程度で相対精度は 14 桁程度であるが、右端では絶対誤差は 3×10^{-5} 程度で相対精度は 5 桁半程度であることが分かる。

$m = 300$ の場合、固有値は区間 $[0, 50]$ の左端では絶対誤差が 10^{-14} 程度で相対精度は 14 桁程度であるが、右端では絶対誤差は 3×10^{-6} 程度で相対精度は 6 桁半程度であることが分かる。

$m = 400$ の場合、固有値は区間 $[0, 50]$ の左端では絶対誤差が 10^{-14} 程度で相対精度は 14 桁程度であるが、右端では絶対誤差は 3×10^{-7} 程度で相対精度が 7 桁半程度であることが分かる。

経過時間

フィルタを適用するベクトルの数を $m=200$, $m=250$, $m=300$, $m=400$ としたそれぞれの場合について、処理ステップごとの経過時間の例を表 6 に示す。

7. 中間固有対を求める実験の例

4 章で記述した方法により、フィルタに用いる多項式としてチェビシェフ多項式を採用する簡易構成法を用いて、中間固有対に対するフィルタをシフトが虚数である単一のレゾルベントの作用の虚部の多項式として具体的に構成した。それを用いて対角化の数値実験を行い、固有値が固有値分布の中央 (一般の位置に) にある中間固有対の近似がうまく求まるかについて調べた。以下にその実験結果の例を示す。

中間固有対用のフィルタ

フィルタの通過域 $\lambda \in [a, b]$ が $t \in [-1, 1]$ に対応する正規化座標 t で表したフィルタの伝達関数を、 $g'(t) = g_s T_n(y)$, ただし $y = 2x' - 1$ で $x' = (\mu + \sigma)/(t^2 + \sigma)$ とする。そのとき阻止域は $|t| \geq \mu' (= \sqrt{\mu})$ であり、 $g'(t)$ の極は複素共役な虚数の対 $t = \pm i\sqrt{\sigma}$ である。伝達関数を決める 3 つ組の値として n , $\mu' = \sqrt{\mu}$, σ を指定すると、伝達関数の閾値 g_p と g_s は式の計算により決まる。

レゾルベントのシフトは虚数 $\rho' = \frac{a+b}{2} + i\left(\frac{b-a}{2}\right)\sqrt{\sigma}$ で、係数を $\gamma' = \frac{\mu+\sigma}{\sqrt{\sigma}}\left(\frac{b-a}{2}\right)$ とすると、フィルタ作用素は $\mathcal{F}' = g_s T_n(2\gamma' \text{Im } \mathcal{R}(\rho') - I)$ となる。

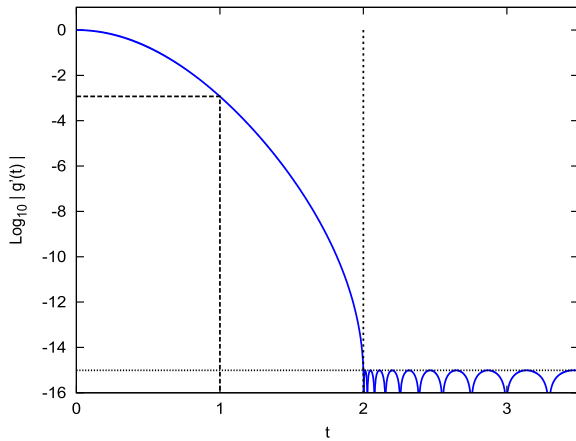


図 25 例題 M: 伝達関数の大きさ $|g'(t)|$ (右半分) ($n = 20$, $\mu' = 2.0$, $\sigma = 4.0$) $g_p = 1.18 \times 10^{-3}$, $g_s = 9.77 \times 10^{-16}$

Fig. 25 (Exam M) Magnitude of transfer function $|g'(t)|$ (right-half) ($n = 20$, $\mu' = 2.0$, $\sigma = 4.0$) $g_p = 1.18 \times 10^{-3}$, $g_s = 9.77 \times 10^{-16}$.

7.1 例題 M (次数 33 万 6 千の大規模問題)

この例題 M の有限要素法の要素分割は $(N_1, N_2, N_3) = (60, 70, 80)$ とした. すると行列 A と B の次数は $N = 336,000$ で下帯幅は $w_L = 4,261$ である. そうして求める固有対の固有値の区間は $[a, b] = [500, 510]$ にした (この場合の真の固有対の数は 153 である).

この例題 M で用いたフィルタの伝達関数のパラメタの 3 つ組の値には $(n = 20, \mu' = 2.0, \sigma = 4.0)$ を指定した. それから決まる伝達関数の閾値はそれぞれ $g_p = 1.18 \times 10^{-3}$ と $g_s = 9.77 \times 10^{-16}$ である. 伝達関数 $g'(t)$ の大きさのグラフを (偶関数であるから右半分 $t \geq 0$ について) 図 25 に示す.

固有値の区間 $[a, b] = [500, 510]$ を通過域とするフィルタは, レゾルベントのシフトを虚数 $\rho' = 505 + 10i$ とし, 係数を $\gamma' = 20$ とするとき, $\mathcal{F}' = g_s T_n(2\gamma' \text{Im } \mathcal{R}(\rho') - I)$ である.

フィルタを適用するベクトルの数をそれぞれ $m = 200$, $m = 250$, $m = 300$, $m = 400$ とした場合の各結果をまとめてグラフに示す (図 26, 図 27, 図 28, 図 29).

7.1.1 (例題 M) 各 m に対する β の固有値の分布

図 26 は各 m の場合について, β の固有値分布を減少順に並べてその大きさの常用対数の値をプロットしたグラフである. $m = 200$ の場合には, β の固有値の分布は設定された閾値 (この場合はマシンプシロンの 100 倍 $= 2.220 \times 10^{-14}$) よりもかなり大きくて切断を受けておらず, これはまだベクトルの数 m が十分ではないことを意味する. $m = 250$ の場合にもまだ β の固有値分布は閾値よりも少し上にあって切断を受けていないので, ベクトルの数はまだ少し足りないことを意味する. $m = 300$ の場合には, β の固有値の最小値は 10^{-15} 程度であり, 閾値による切断により階数が 34 だけ低下して 266 となったので, こ

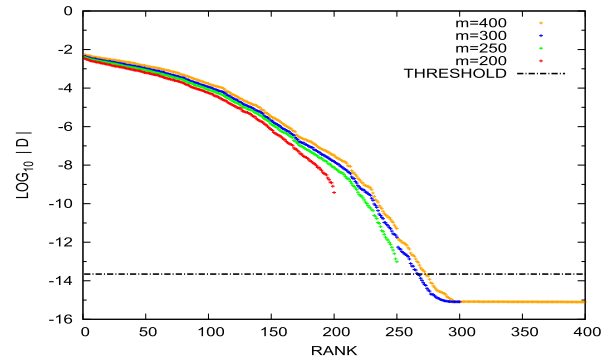


図 26 (例題 M) 各 m に対する β の固有値の分布

Fig. 26 (Exam M) Eigenvalue distribution of β for each m .

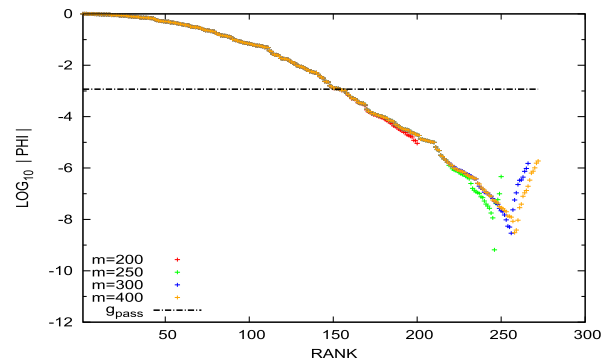


図 27 (例題 M) 各 m に対する $|\phi|$ の値の分布

Fig. 27 (Exam M) Distribution of values of $|\phi|$ for each m .

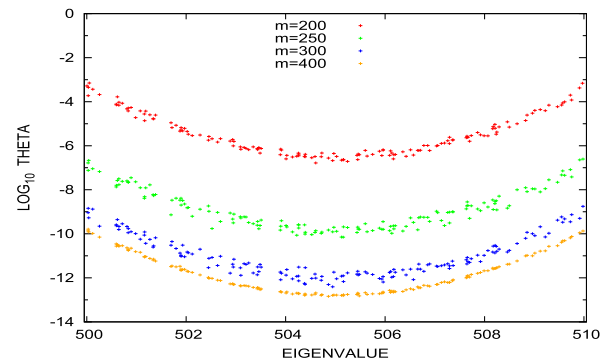


図 28 (例題 M) 各 m に対する近似対の相対残差 Θ

Fig. 28 (Exam M) Values of Θ , norms of relative residuals of approximate pairs for each m .

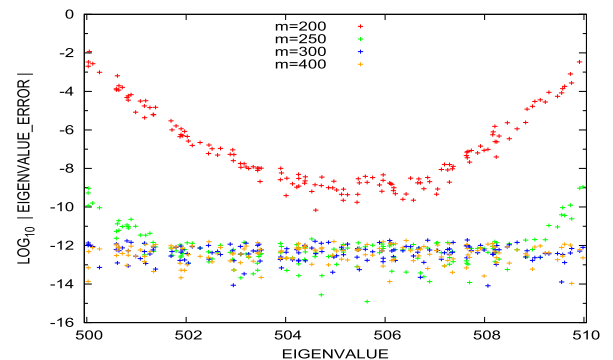


図 29 (例題 M) 各 m に対する近似固有値の絶対誤差

Fig. 29 (Exam M) Absolute errors of approximate eigenvalues for each m .

の場合にはベクトルの数はすでに十分である。 $m = 400$ の場合にも、多くの階数低下を起こしているため、ベクトルの数は十分である。

7.1.2 (例題 M) 各 m に対する $|\phi|$ の値の分布

図 27 は各 m の場合について、 $|\phi|$ の値の分布を減少順に並べてその常用対数の値をプロットしたグラフである。

$m = 200$ や $m = 250$ の場合には、もともとベクトルの数 m が不足しているため、 ϕ の値は固有ベクトルに対する伝達率の近似としては良くない。

それでも $m = 250$ の場合に、 $|\phi|$ の値の分布に基づいて決めた閾値より上側にある ϕ の値の数から決まる不変部分空間の近似空間の基底ベクトルの数は 157 であった。それから求めた 157 個の近似対で固有値が $[500, 510]$ にあるものだけを残すと、真の固有対の数 153 と同数の近似対が得られた。

$m = 300$ と $m = 400$ の場合にも、決定した閾値から決まる不変部分空間の近似の基底ベクトルの数は 157 であり、そのうち固有値が $[500, 510]$ にあるものを残すと、真の固有対の数 153 と同数の近似対が得られた。

7.1.3 (例題 M) 各 m に対する近似対の相対残差 Θ

図 28 は各 m の場合について、横軸に求めた近似対の固有値をとり、縦軸に相対残差 Θ の常用対数の値をとってプロットしたグラフである。

$m = 200$ のときはベクトルの数がかなり不足しているため近似対の精度はきわめて悪く、固有値の区間 $[500, 510]$ の中央付近では相対残差は 3×10^{-7} 程度で、両端付近では 10^{-3} 程度となっている。その値の違いは約 3 桁半である。

$m = 250$ のとき、区間の中央付近では相対残差は 10^{-10} 程度で、両端付近では値は 10^{-7} 程度で、値の開きは約 3 桁である。

$m = 300$ のとき、区間の中央付近では相対残差の値は 10^{-12} 程度で、両端付近では値は 10^{-9} 程度で、値の開きは約 3 桁である。

$m = 400$ のとき、区間の中央付近では相対残差の値は 10^{-13} 程度で、両端付近では値は 10^{-10} 程度で、値の開きは約 3 桁である。

7.1.4 (例題 M) 各 m に対する近似固有値の絶対誤差

図 29 は各 m の場合について、横軸に求めた近似対の固有値をとり、縦軸に近似固有値の真の値からの絶対誤差の常用対数の値をとってプロットしたグラフである。

$m = 200$ の場合はベクトルの数 m が不十分であるため、近似固有値の誤差は大きい。固有値の区間 $[500, 510]$ の中央付近では絶対誤差が $10^{-10} \sim 10^{-8}$ 程度で相対精度は 10 桁半から 12 桁半程度であるが、両端付近では絶対誤差は 10^{-2} 程度で相対精度は 4 桁半程度であることが分かる。

$m = 250$ の場合の近似固有値は、区間 $[500, 510]$ の中央付近では絶対誤差が $10^{-13} \sim 10^{-12}$ 程度で相対精度は 14 桁半～15 桁半程度であるが、両端付近では絶対誤差は 10^{-9}

表 7 (例題 M) 経過時間の内訳 (単位: 秒)

Table 7 (Exam M) Breakdown of elapsed time (in second).

処理ステップ	$m=200$	$m=250$	$m=300$	$m=400$
フィルタ対角化全体	2,524.1	3,027.7	3,572.0	5,315.4
・乱数入力ベクトル生成	0.3	0.3	0.4	0.5
・入力ベクトル B -正規直交化	18.4	23.8	28.9	40.4
・フィルタ作用適用	2,415.2	2,907.5	3,441.8	5,163.1
・不変部分空間の基底作成	61.4	67.2	72.1	82.6
・レイリー・リッツ法適用	28.8	28.8	28.8	28.8

表 8 各例題の問題サイズ、フィルタの次数 n と形状パラメタ

Table 8 Problem size, filters' degree n and shape parameters of each example.

例題	N	w_L	n	μ (μ')	g_p	g_s
L1	210,000	3,051	24	1.5	3.15×10^{-7}	3.75×10^{-14}
L2	210,000	3,051	32	2.0	1.13×10^{-5}	1.45×10^{-15}
L3	336,000	4,261	24	1.5	3.15×10^{-7}	3.75×10^{-14}
L	504,000	5,671	24	1.5	3.15×10^{-7}	3.75×10^{-14}
M1	336,000	4,261	10	2.0	2.16×10^{-5}	9.85×10^{-15}
M2	336,000	4,261	15	2.0	6.38×10^{-4}	9.71×10^{-15}
M	336,000	4,261	20	2.0	1.18×10^{-3}	9.77×10^{-16}

程度で相対精度が 11 桁半程度であることが分かる。

$m = 300$ や $m = 400$ の場合には、近似固有値は区間 $[500, 510]$ 全体で絶対誤差が 10^{-12} 程度以下であり相対精度は 14 桁半程度以上であることが分かる。

経過時間

フィルタを適用するベクトルの数を $m = 200$, $m = 250$, $m = 300$, $m = 400$ としたそれぞれの場合について、処理ステップごとの経過時間の例を表 7 に示す。

求めた近似対の品質の確認には相対残差の値 Θ を計算した。これは計算量が少ないので素早く求めることができる。

7.2 例題の比較

表 8 に、下端固有対を求める例題である L1, L2, L3, L および中間固有対を求める例題である M1, M2, M のそれぞれについて、行列次数 N 、下帯幅 w_L 、フィルタの次数 n 、フィルタの遷移域の形状比 μ (μ')、通過域での伝達率の最小値 g_p 、阻止域での伝達率の大きさの最大値 g_s をまとめた。

また表 9 には、例題 L1, L2, L3, L および例題 M1, M2, M のそれぞれについて、フィルタをベクトル m 個の組に対して適用するのにかかる経過時間をまとめた。

たとえば例題 M1, M2, M の 3 つは有限要素法による離散化された固有値問題の係数行列が共通であるため、用いるレゾルベントのシフトとして用いた虚数 ρ' の値がそれぞれ異なっているとしても、行列 $C = A - \rho' B$ に対して複素対称版の改訂コレスキ分解を行う計算の手間はみな同じである。また、求めた行列分解の結果を利用して連立 1 次方程式の解を計算する手間についてもベクトルの数 m が同じであ

表 9 各例題のフィルタ適用の経過時間 (単位: 秒)

Table 9 Elapsed time to apply a filter for each example (in second).

例題	$m=200$	$m=250$	$m=300$	$m=400$
L1	626	747	855	1,070
L2	782	945	1,088	1,374
L3	1,538	1,838	2,061	2,557
L	3,619	4,271	4,790	6,139
M1	1,684	1,920	2,204	3,056
M2	2,044	2,408	2,829	4,114
M	2,415	2,908	3,442	5,163

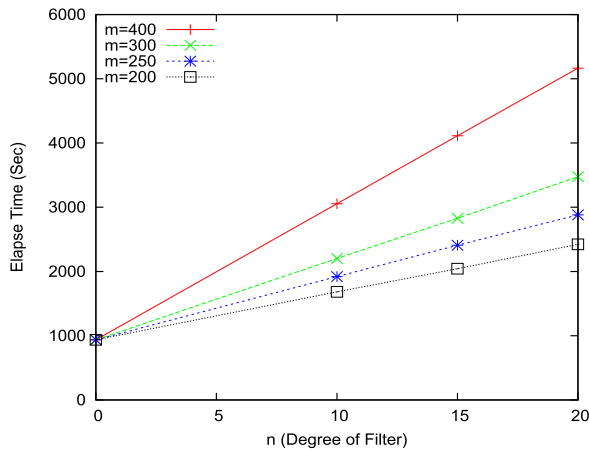


図 30 (例題 M1, M2, M) フィルタ適用の経過時間 (その 1, 単位: 秒)

Fig. 30 (Exam M1, M2, M) Elapsed time to apply a filter (case 1, in second).

ればみな同じになる。ただしフィルタの作用を完了させるまでには、例題 M1 は m 個のベクトルに対する連立 1 次方程式を計 $n = 10$ 回、例題 M2 は計 $n = 15$ 回、例題 M は計 $n = 20$ 回解く必要がある。これらの (帯行列に対する) 複素対称版の改訂コレスキ分解に要した経過時間は 936~937 秒程度であった。これは分解法をブロック化算法に変更すれば短縮されうと思われる (ちなみに同じ構造のプログラムコードで次数 N と下帯幅 w_L が同じで行列が実対称である場合に、実対称版の改訂コレスキ分解に要した経過時間はほぼ半分の 503~504 秒程度であった)。これら例題 M1 ($n = 10$), M2 ($n = 15$), M ($n = 20$) のそれぞれについて、 $m = 200, m = 250, m = 300, m = 400$ の各場合に対して、横軸にフィルタの次数 n をとり、縦軸に $N = 336,000$ 次のベクトル m 個に対してフィルタを適用するのにかった経過時間をグラフにプロットした (図 30)。 $n = 0$ のときの値は行列 C の分解だけの経過時間である 936.5 秒としている。経過時間の増分は次数 n にほぼ比例していることが分かる。しかし横軸にベクトルの数 m をとり、縦軸にフィルタの適用にかかった経過時間をグラフにプロットしてみた (図 31)。ただし $m = 0$ のときの値は仮に行列 C の分解だけの経過時間である 936.5 秒として

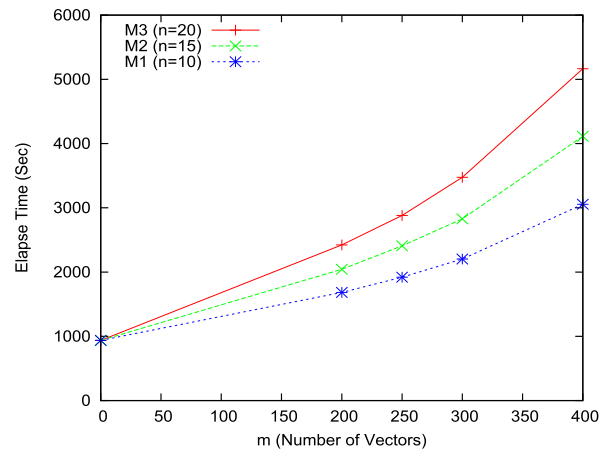


図 31 (例題 M1, M2, M) フィルタ適用の経過時間 (その 2, 単位: 秒)

Fig. 31 (Exam M1, M2, M) Elapsed time to apply a filter (case 2, in second).

いる。これを見ると残念ながら右辺ベクトルの数 m については増加の程度が比例するよりも急になっており、うまくスケールしていないことが分かる。これは現在のプログラムの実装が BLAS-2 レベルであって記憶参照の局所性があまり良くないからであろう。計算性能を向上させるのならば実装の改良が今後必要である。

8. レゾルベントの線形結合型のフィルタとの比較

シフトが虚数であるレゾルベント (と恒等作用素) の線形結合として表されるフィルタが実作用素であるならば、シフトの全体は k 対の複素共役対をなして、シフトが互いに複素共役であるレゾルベントの結合係数もまた互いに複素共役である。そこで、そのようなフィルタはたとえば虚部が正のシフトとそれに対する結合係数を $\rho_j, \gamma_j, j = 1, 2, \dots, k$ とすれば、以下の式 (37) の形になる (c_∞ は実数の定数である)。

$$\mathcal{F} = c_\infty I + \sum_{j=1}^k \text{Re} \{ 2\gamma_j \mathcal{R}(\rho_j) \}. \quad (37)$$

いま固有値を求める区間 $\lambda \in [a, b]$ を線形変換で正規化座標 $t \in [-1, 1]$ に対応させ、引数 t の伝達関数を $g(t)$ とする。伝達関数の通過域、遷移域、阻止域はそれぞれ $|t| \in [0, 1], |t| \in (1, \mu), |t| \in [\mu, \infty)$ である。そうして伝達関数 $g(t)$ の通過域、遷移域、阻止域における値域がそれぞれ $[g_p, 1], (g_s, g_p), [-g_s, g_s]$ となるように決める。

典型的な 4 種類のフィルタである「バターワース (Butterworth)」、「チェビシェフ (Chebyshev)」、「逆チェビシェフ (inverse-Chebyshev)」、「楕円 (elliptic)」は上記の式 (37) の形をしたレゾルベントの線形結合である。そうして用いるレゾルベントの数 k が同じであるときに、それら 4 種類のうちで最も優れた伝達特性を持つのは楕円フィルタであ

る [14]. あるいは 4 種類のうちで最も小さい k で優れた特性を実現するのが楕円フィルタであるともいえる. レゾルベントの線形結合型のフィルタの場合には, 同等以上の特性をより少ないレゾルベントの個数 k で実現できるのである. 少ない方を選べば解くべき連立 1 次方程式の数も少なくなり計算量が減らせるであろう. そこで, 今回のフィルタとの比較対象をレゾルベントの線形結合型のフィルタから選ぶとすれば, 4 種類のうちでは楕円フィルタが最も妥当であると考えられる.

そこで楕円フィルタを今回の (シフトが虚数の場合の) 単一のレゾルベントの虚部の n 次 Chebyshev 多項式として作られたフィルタと比較する. ただし楕円フィルタでは遷移域の幅は任意に狭くできるが, 今回のフィルタではそのようなことは不可能で実現可能な形状の良さには限界がある. そこで比較の際に条件をなるべく揃えるために, まず今回の方式のフィルタを形状パラメタの 3 つ組 (g_p, μ, g_s) を指定して構成し, それに対する楕円フィルタは形状パラメタの 3 つ組の値をなるべく近くしたうえでレゾルベントの数 k が最小のものを構成する (多項式の次数 n やレゾルベントの個数 k は整数値だけをとるので, 形状パラメタ 3 つの値をすべてを一致させることは一般には不可能である). そうして乱数を元にして生成された B -正規直交である m 個のベクトルの組 X にフィルタを適用する処理 $Y \leftarrow \mathcal{F}X$ にかかる経過時間を双方のフィルタについて比較してみる.

今回の虚数 ρ をシフトとする単一のレゾルベントの虚部から n 次 Chebyshev 多項式を用いて作られたフィルタの場合には, その作用を実現するためには, まず複素対称なシフト行列 $C \leftarrow A - \rho B$ を作って, その複素対称版の改訂コレスキ分解 $C \rightarrow LDL^T$ の結果を保持する. そうして, レゾルベントのベクトルの組 V への作用 $W \leftarrow \mathcal{R}(\rho)V$ を求めるのには, まず V に B を乗じて BV を作り, それを右辺としシフト行列 C を係数とする連立 1 次方程式 $CW = BV$ を求めておいた行列分解の結果を利用して解く. レゾルベントの n 次多項式をベクトルの組に作用させる処理の中には, ベクトルの組にレゾルベントを適用する処理が合計 n 回現れる. 特に多項式が Chebyshev 多項式なので 3 項漸化式を用いて計算ができる.

他方で楕円フィルタは k 個のレゾルベントの線形結合 (の実部) である. m 個のベクトルの組に対して各レゾルベントを 1 度だけ適用するので, 各レゾルベントの作用を並行に計算しようとするに対応するシフト行列の行列分解の結果を保持するための記憶量がフィルタの作用を実現する計算の隘路になりがちである. そこで記憶容量が隘路になる大規模問題の場合には, (処理の並行性は犠牲になるが, 全体の演算量を増すことなしに) フィルタのベクトルの組 X への作用は, $j = 1, 2, \dots, k$ について順次に各レゾルベントの項 $2\gamma_j \mathcal{R}(\rho_j)$ を X に対して適用して

得た結果の実部を足していくことで実現できる. それにはまず BX を作って保持する. そうして $Y \leftarrow 0$ とする. それから順次に $j = 1, 2, \dots, k$ について, まずシフト行列 $C^{(j)} \leftarrow A - \rho_j B$ を作り, それから右辺が BX である連立 1 次方程式 $C^{(j)}Z^{(j)} = BX$ を係数 $C^{(j)}$ が複素対称行列であるから複素対称版の改訂コレスキ法で解いて $Z^{(j)}$ を求める. そうして $Y \leftarrow Y + \text{Re}(2\gamma_j Z^{(j)})$ とする. 処理をこのように逐次的に行えばレゾルベントの項 k 個についての並行性は利用できなくなる. しかしその代わりに処理全体としてシフト行列 1 個分 (対称性から下半分) の記憶場所を確保できれば, 逐次に $j = 1, 2, \dots, k$ に対して連立 1 次方程式 $C^{(j)}Z^{(j)} = BX$ を解く際には行列 $C^{(j)}$ とその分解をいつもその同じ場所に置いて入れ替えながら作業を行うことができる (もしも対称行列の下半分を数個分入れるだけの記憶容量が確保できるなら, その個数ずつ同時並行に処理を行うことが可能である).

以下の実験ではフィルタが今回の虚数をシフトに持つ単一のレゾルベントの虚部の n 次多項式である場合, あるいはシフトが虚数のレゾルベント k 個の線形結合の実部である楕円フィルタの場合, そのどちらの場合にもレゾルベントの作用を与える連立 1 次方程式の係数行列は行列の下半分 1 個分を入れる記憶量を (A と B 以外に) 確保して計算をしている. 行列 C と各 $C^{(j)}$ はどちらも次数は N で下帯幅は w_L である.

例 1: この例 1 で扱った一般固有値問題 $A\mathbf{v} = \lambda B\mathbf{v}$ は, 以前と同じく立方体領域における零-Dirichlet 境界条件で (符号が逆の) ラプラシアン固有値問題を有限要素法を用いて離散化の要素分割を $(N_1, N_2, N_3) = (20, 30, 40)$ として得られたものであり, 係数行列 A と B の次数は $N = 24,000$ で下帯幅は $w_L = 621$ である. 各行列の帯内部は実際にはきわめて疎であるが, 計算の例題としては A と B の帯内部が密であるかのようになっている.

シフトが虚数の単一のレゾルベントの多項式であるフィルタは, Chebyshev 多項式の次数が $n = 40$ で, 形状パラメタの 3 つ組を $g_p = 3.699 \times 10^{-4}$, $\mu' = 1.5$, $g_s = 3.628 \times 10^{-13}$ とした. 比較する楕円フィルタはなるべくそれに近い形状パラメタを持つように選んだ. 双方の g_p と μ' (μ) は完全に一致させたが g_s の値については楕円フィルタの方が今回のフィルタのものよりも値が小さくなるようにした (つまり形状パラメタだけで較べると楕円フィルタの方が特性が優れたものとなるようにした). そのような条件を満たす最小の k の値は 8 であった. 得られた楕円フィルタは虚数をシフトとするレゾルベント $k = 8$ 個の線形結合の実部であり, その伝達関数の形状パラメタの 3 つ組は $g_p = 3.699 \times 10^{-4}$, $\mu = 1.5$, $g_s = 1.936 \times 10^{-14}$ である.

フィルタを適用するベクトルの数を $m = 100$, $m = 150$, $m = 200$, $m = 300$ と変えてそれぞれの経過時間を測定した (表 10). なお m 個のベクトルの組に対する各処理は m

表 10 例 1: フィルタを m 個のベクトルの組に適用した処理の経過時間 (単位: 秒)

Table 10 Ex-1: Elapsed time to apply a filter to a set of m vectors (in second).

	$m=100$	$m=150$	$m=200$	$m=300$
今回のフィルタ	15.85	21.87	27.91	41.66
楕円フィルタ	6.68	7.66	8.68	10.99

個のベクトルをまとめていっせいに適用している。計算に用いたシステムは 5.1 節で述べたものである。主記憶上にただ 1 つのシフト行列およびその分解を置ける場所を確保し、そうして行列分解の内部や分解結果を利用して連立 1 次方程式を解く処理の内部だけを OpenMP を用いた物理 8 コア上の 8 スレッドにより並列に計算している。ただしこの例 1 は必要な記憶量の少ない小規模な問題である。

表 10 の数値から、この例 1 については、今回の単一のレゾルベントの多項式であるフィルタよりも楕円フィルタの方が経過時間が短いことが分かる。その違いは m が大きくなるほど広がる。この例 1 では楕円フィルタの作用 $Y \leftarrow FX$ は、まず最初に m 個のベクトルの組 X に行列 B を乗じて BX を作る処理を 1 回だけ行い、 $j = 1, 2, \dots, k$ に対して順次にシフト行列 $C^{(j)}$ を係数とし右辺が BX である連立 1 次方程式を行列分解を用いて解きながら解ベクトルの組の複素線形結合の実部を Y に加算して実現している。

単一のレゾルベントの多項式をフィルタとする場合には、レゾルベントに対応する係数行列 C の分解は 1 回だけで済むが、その後に「 m 個のベクトルの組 V に行列 B を乗じて作ったベクトルの組 BV を右辺として行列 C を係数に持つ連立 1 次方程式をすでに得ている C の行列分解の結果を利用して解く」という作業が逐次的に多項式の次数に等しい $n = 40$ 回だけ繰り返され行われる。

連立 1 次方程式の係数である複素対称の帯行列の下半分を格納する記憶量は、帯内を密であるかのように扱っているので約 Nw_L 語である。ベクトル m 個の組の 1 つ分の記憶量は Nm 語である。帯行列の改訂コレスキ分解の演算量は $O(Nw_L^2)$ でありそれを係数 c_D を用いて $c_D Nw_L^2$ と近似できたとする。行列 B を m 個のベクトルの組に乗じる演算量は $O(Nmw_L)$ でありそれを係数 c_M を用いて $c_M Nmw_L$ と近似できたとする。また右辺が m 個のベクトルの組である連立 1 次方程式をすでに得ている行列分解を用いて解く演算量も $O(Nmw_L)$ でありそれを係数 c_S を用いて $c_S Nmw_L$ と近似できたとする。するとフィルタの処理に掛かる演算の量は、フィルタが単一のレゾルベントの虚部の n 次多項式である場合には以下の式 (38) のようになり、フィルタが k 個のレゾルベントの線形結合の実部である場合には以下の式 (39) のようになる。

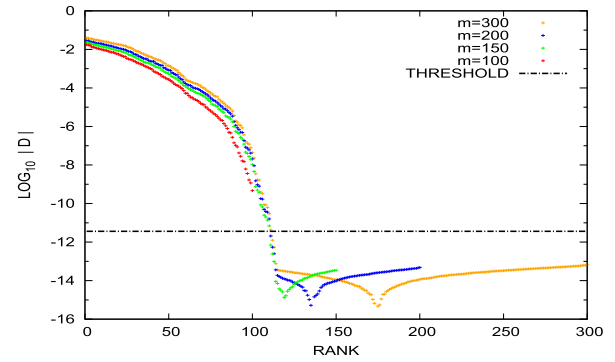


図 32 (例 1, 提案のフィルタ) 各 m に対する β の固有値の分布
Fig. 32 (Ex-1, by present filter) Eigenvalue distribution of β for each m .

$$\begin{aligned}
 & c_D Nw_L^2 + n(c_M + c_S)Nm w_L \\
 & = Nw_L^2 \times \{c_D + (c_S + c_M)(nm/w_L)\}. \\
 & \text{(フィルタは単一のレゾルベントの虚部の } n \text{ 次多項式)} \\
 & \hspace{15em} (38)
 \end{aligned}$$

$$\begin{aligned}
 & c_M Nmw_L + k(c_D Nw_L^2 + c_S Nmw_L) \\
 & = kNw_L^2 \times \{c_D + (c_S + c_M/k)(m/w_L)\} \\
 & \text{(フィルタは } k \text{ 個のレゾルベントの線形結合の実部)} \\
 & \hspace{15em} (39)
 \end{aligned}$$

よって定性的には係数行列の半帯幅 w_L がベクトルの個数 m (あるいは nm) に比べて十分に大きい状況であれば、フィルタ処理の演算の主要部は行列分解になるので、 k 個のレゾルベントを用いるのではなくて単一のレゾルベントを用いる方が演算量だけに関しては有利になりうる (m が少ないほど、また n が少ないほど有利になる傾向が強まる)。また係数 c_D , c_M , c_S を演算回数に対する比例係数ではなくて各種の演算処理に必要な経過時間に対する比例係数であるとすれば、経過時間についても定性的には同様の議論となるであろう。

今回提案のレゾルベントを 1 つだけ用いるフィルタによりベクトルの数が $m = 100$, $m = 150$, $m = 200$, $m = 300$ の各場合についてそれぞれ固有値が区間 $[a, b] = [200, 210]$ にある 87 個の固有値すべてを求めた場合について、 m 次の行列 β の固有値の分布とその切断値のグラフ (図 32)、切断を入れて解いた m 次の一般固有値問題 $\alpha \mathbf{u} = \phi \beta \mathbf{u}$ の固有値とフィルタの通過域における伝達率の最小値 g_p の線のグラフ (図 33)、求めた近似対の固有値に対する相対残差 Θ の値のグラフ (図 34) および近似固有値に対するその絶対誤差のグラフ (図 35) を示す。 $m = 100$ では β の固有値には閾値で切断されるものが無く、ベクトルの数はかなり不足して相対残差や固有値の誤差は大きい。

上述のレゾルベントを 8 つ用いた楕円フィルタによりベクトルの数が $m = 100$, $m = 150$, $m = 200$, $m = 300$ の各場合についてそれぞれ固有値が区間 $[a, b] = [200, 210]$ に

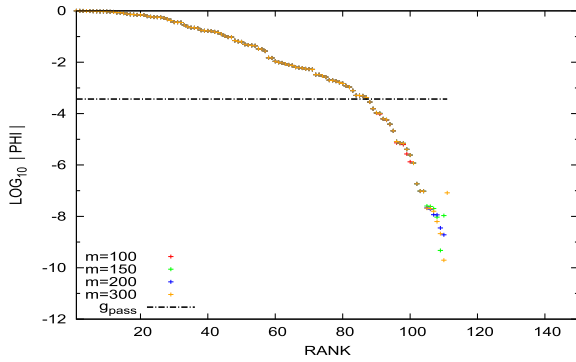


図 33 (例 1, 提案のフィルタ) 各 m に対する $|\phi|$ の値の分布
Fig. 33 (Ex-1, by present filter) Distribution of values of $|\phi|$ for each m .

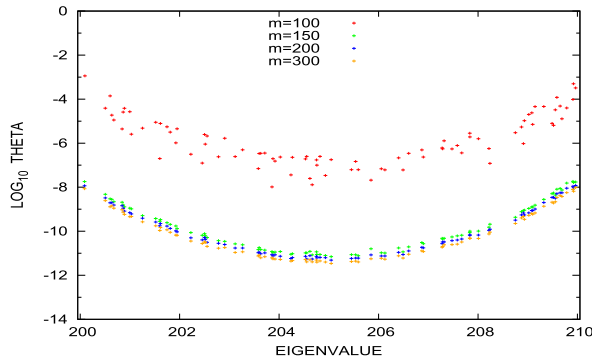


図 34 (例 1, 提案のフィルタ) 各 m に対する近似対の相対残差 Θ
Fig. 34 (Ex-1, by present filter) Values of Θ , norms of relative residuals of approximate pairs for each m .

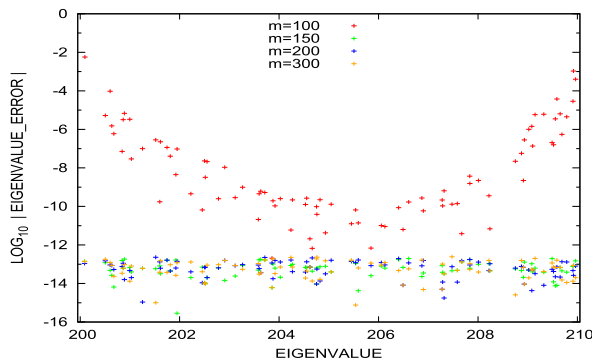


図 35 (例 1, 提案のフィルタ) 各 m に対する近似固有値の絶対誤差
Fig. 35 (Ex-1, by present filter) Absolute errors of approximate eigenvalues for each m .

ある 87 個の固有対すべてを求めた場合について同様のグラフを図 36, 図 37, 図 38 および図 39 に示す. $m = 100$ では与えたベクトルの数がまだ不足であり, 相対残差や固有値の誤差が大きい.

この例 1 に関しては, 両方のフィルタにより得られた近似対の相対残差や固有値の誤差をグラフをそれぞれ眺めてみると, ベクトルの数が不足している $m = 100$ を除いて考えれば, 結果の精度はほぼ互角であるように見える.

例 2: この例 2 で扱う一般固有値問題は前の例 1 の場合に比べて大規模な問題で, 有限要素法の要素分割の数

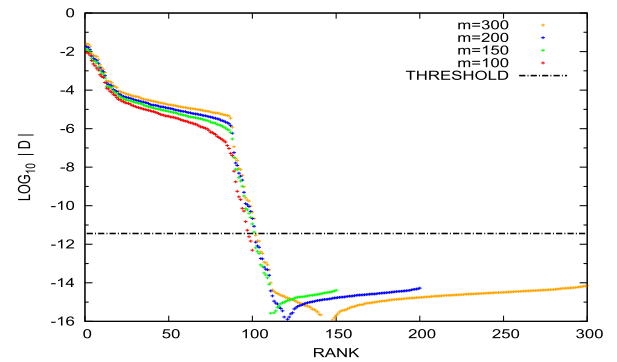


図 36 (例 1, 楕円フィルタ) 各 m に対する β の固有値の分布
Fig. 36 (Ex-1, by elliptic filter) Eigenvalue distribution of β for each m .

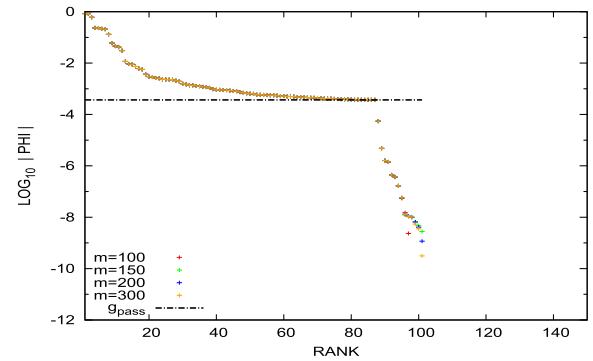


図 37 (例 1, 楕円フィルタ) 各 m に対する $|\phi|$ の値の分布
Fig. 37 (Ex-1, by elliptic filter) Distribution of values of $|\phi|$ for each m .

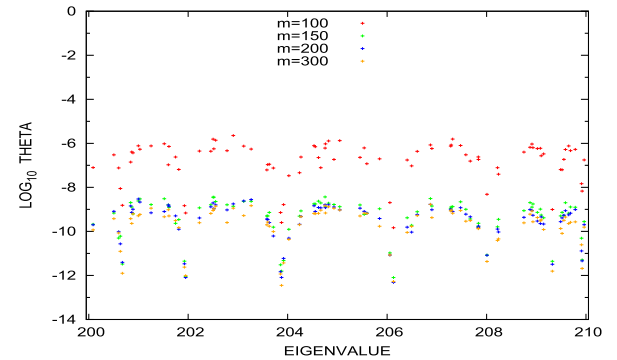


図 38 (例 1, 楕円フィルタ) 各 m に対する近似対の相対残差 Θ
Fig. 38 (Ex-1, by elliptic filter) Values of Θ , norms of relative residuals of approximate pairs for each m .

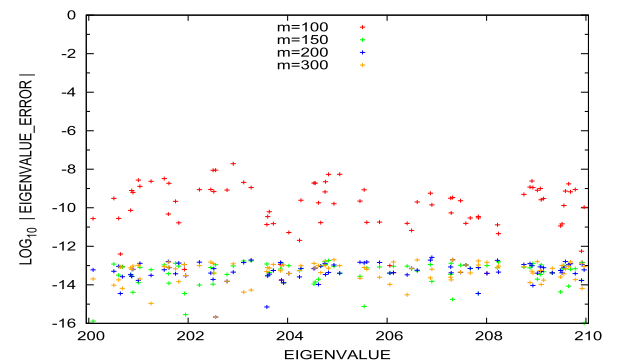


図 39 (例 1, 楕円フィルタ) 各 m に対する近似固有値の絶対誤差
Fig. 39 (Ex-1, by elliptic filter) Absolute errors of approximate eigenvalues for each m .

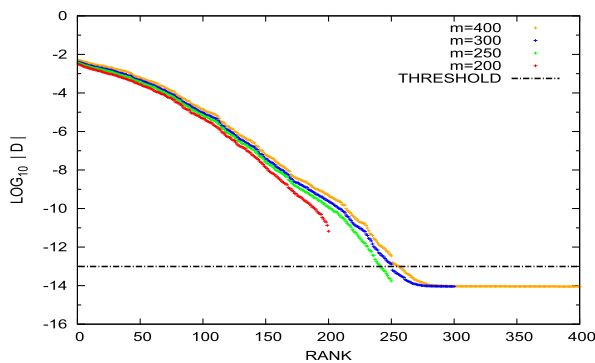


図 40 (例 2, 提案のフィルタ) 各 m に対する β の固有値の分布
Fig. 40 (Ex-2, by present filter) Eigenvalue distribution of β for each m .

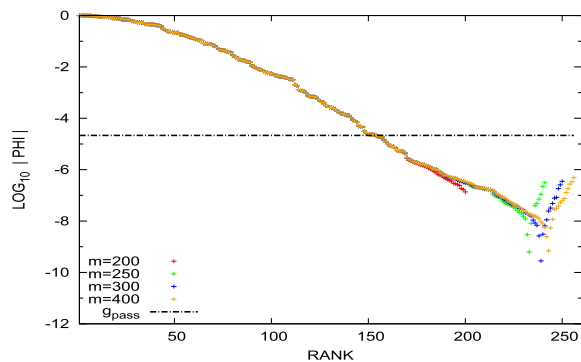


図 41 (例 2, 提案のフィルタ) 各 m に対する $|\phi|$ の値の分布
Fig. 41 (Ex-2, by present filter) Distribution of values of $|\phi|$ for each m .

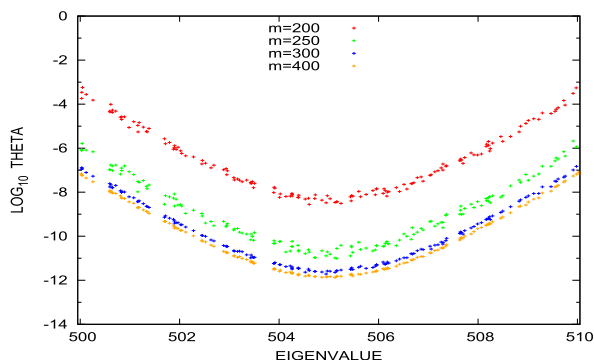


図 42 (例 2, 提案のフィルタ) 各 m に対する近似対の相対残差 Θ
Fig. 42 (Ex-2, by present filter) Values of Θ , norms of relative residuals of approximate pairs for each m .

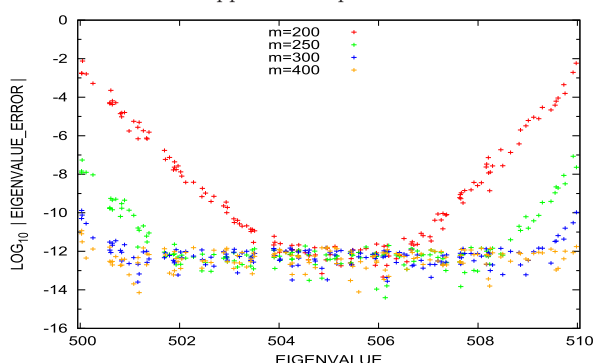


図 43 (例 2, 提案のフィルタ) 各 m に対する近似固有値の絶対誤差
Fig. 43 (Ex-2, by present filter) Absolute errors of approximate eigenvalues for each m .

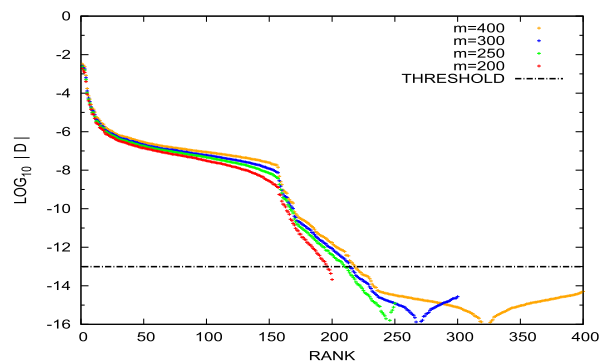


図 44 (例 2, 楕円フィルタ) 各 m に対する β の固有値の分布
Fig. 44 (Ex-2, by elliptic filter) Eigenvalue distribution of β for each m .

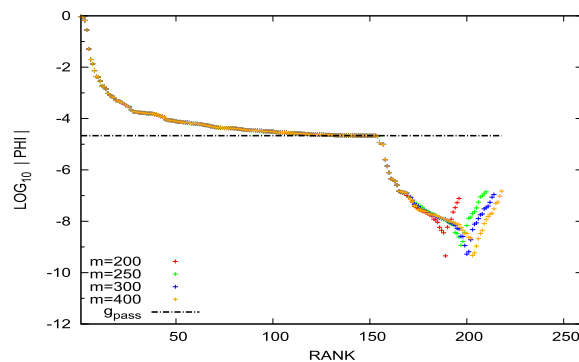


図 45 (例 2, 楕円フィルタ) 各 m に対する $|\phi|$ の値の分布
Fig. 45 (Ex-2, by elliptic filter) Distribution of values of $|\phi|$ for each m .

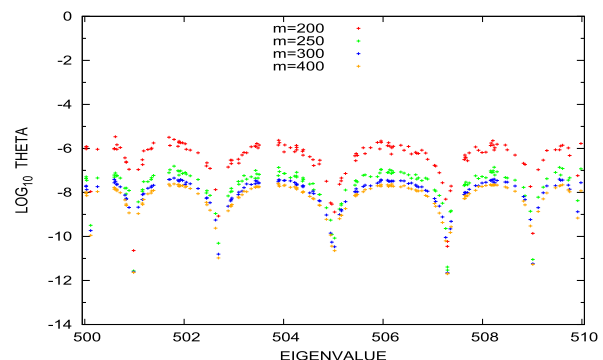


図 46 (例 2, 楕円フィルタ) 各 m に対する近似対の相対残差 Θ
Fig. 46 (Ex-2, by elliptic filter) Values of Θ , norms of relative residuals of approximate pairs for each m .

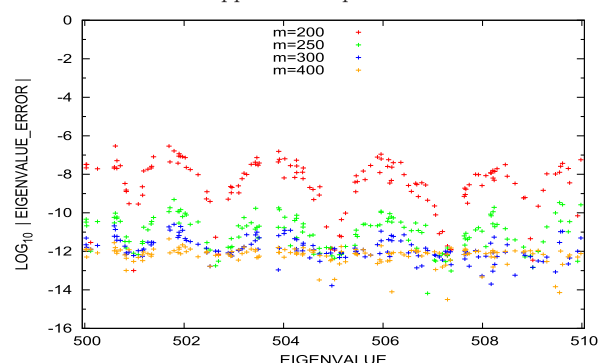


図 47 (例 2, 楕円フィルタ) 各 m に対する近似固有値の絶対誤差
Fig. 47 (Ex-2, by elliptic filter) Absolute errors of approximate eigenvalues for each m .

は $(N_1, N_2, N_3) = (60, 70, 80)$ であり、係数行列の次数は $N = 336,000$ 、下帯幅は $w_L = 4,261$ である。

虚数をシフトとする単一のレゾルベント（の虚部の）の多項式であるフィルタは、その Chebyshev 多項式の次数を $n = 10$ 、 $\mu' = 2.0$ 、 $\sigma = 0.64$ として設定した。それに対応する形状パラメタの3つ組は $g_p = 2.16 \times 10^{-5}$ 、 $\mu' = 2.0$ 、 $g_s = 9.85 \times 10^{-15}$ である。

比較対象とする楕円フィルタはなるべく伝達関数の形状パラメタの組が近くなるように選んだ（単一のレゾルベントの多項式の場合と比べて g_s の値だけが異なっていてしかも小さい（特性が良い）ように設定した）。その結果の楕円フィルタは、虚数をシフトとするレゾルベント $k = 7$ 個の複素線形結合の実部で、その伝達関数の形状パラメタの3つ組は $g_p = 2.16 \times 10^{-5}$ 、 $\mu = 2.0$ 、 $g_s = 2.09 \times 10^{-16}$ となった。

例1の場合と同様に2種類のフィルタを用いて、固有値が区間 $[a, b] = [500, 510]$ にある153個の固有対を求めた。それぞれのフィルタに適用するベクトルの数を $m = 200$ 、 $m = 250$ 、 $m = 300$ 、 $m = 400$ と変えた各場合について実験を行った。

m 次の行列 β の固有値の分布とその切断線のグラフ（図40、図44）、切断を入れて解いた m 次の一般固有値問題 $\alpha u = \phi \beta u$ の固有値とフィルタの通過域における伝達率の最小値 g_p の線のグラフ（図41、図45）、求めた近似対の固有値に対する相対残差 Θ の値のグラフ（図42、図46）および近似固有値に対するその絶対誤差のグラフ（図43、図47）を示す。

提案のフィルタは $m = 200$ では β の固有値に閾値で切断されるものがなく、ベクトルの数がかなり不足しているために相対残差や固有値の誤差は大きい。また楕円フィルタでもベクトルの数が $m = 200$ ではそれほど十分であるとはいえず、相対残差や固有値の誤差が大きい。

この例2について、両方のフィルタを用いて得られた近似対の相対残差や固有値の絶対誤差のグラフを（ベクトルの数がまだ不足しているあるいはそれほど十分ではない $m = 200$ の場合を除いて）比較してみると、楕円フィルタの方が相対残差の最大値は1桁小さく、また固有値の絶対誤差の最大値は2桁ほど小さい。その理由としては、この例2で用いたフィルタの g_s の値が提案のフィルタでは 9.85×10^{-15} であるのに対して楕円フィルタでは 2.09×10^{-16} で値が約 $1/50$ になっていることが考えられる。

注：フィルタ対角化法では、近似によって得られた各種フィルタはその特性に対する解くべき固有値の分布状況の適合性が異なることや、有限精度の数値と演算を用いて計算をしていることから生じる数値丸め誤差の影響もあるので、得られる近似固有対の精度は採用したフィルタの特性や実装法によって異なりうる。そのため必要な演算量あるいは経過時間だけで単純に比較することはできない。

単一のレゾルベントで多項式にチェビシェフ多項式を採用した簡易構成型のフィルタでは、フィルタの特性を良くしようとしても限界がある。たとえば遷移域の幅をうんと狭くすることはできないが、他方で楕円フィルタであれば用いるレゾルベントの数を増せば遷移域の幅をきわめて狭くすることが容易にできる（その代わり増えたレゾルベントに対応して連立1次方程式を多く解くために計算量が増える可能性がある）。

今回の単一のレゾルベントを用いた簡易型のフィルタは、実現可能なフィルタの特性やそれから期待できる近似対の精度に関してはレゾルベントを多く用いる楕円型フィルタに比べて一般には劣る。レゾルベントの作用を与える連立1次方程式を行列分解を用いて直接法で解く場合であれば、得られる解の精度はあまり良くななくても、そのかわりに必要な記憶容量や演算量など資源に制約がある状況ではたとえば計算が早く終わるなどの利点があるのではないかと考えて調べている。

9. おわりに

実対称定値一般固有値問題を解くためのフィルタとして単一のレゾルベント（の虚部の）の多項式を用いる方法について考察した。レゾルベントのシフトと多項式の持つ係数の自由度を数値最適化を用いてフィルタの伝達関数の形状を最適となるように決めることは避けて、その多項式である1つのチェビシェフ多項式で表される形に制限することで、伝達関数の形状を準最適にする簡易な設計法を採用した。得られる簡易型のフィルタは、阻止域における伝達率の大きさを微小にすることはチェビシェフ多項式の次数を上げることにより容易に実現できるが、通過域における伝達率の最大最小比を小さくすることは難しい。この最大最小比が大きいとそれだけ得られる近似対の精度が均一でなくなる可能性がある。

ベクトルの組に対して単一のレゾルベント（の虚部の）の多項式であるフィルタを作用させる計算の中では、そのレゾルベントを多項式の次数と同じ回数だけ繰り返し作用させる。レゾルベントの作用は連立1次方程式を解いて実現するが、その連立1次方程式は係数行列の分解を用いて直接法で解くとすれば、最初に行った分解の結果を保持できるのであればその分解を再利用することにより同じ係数の連立1次方程式は新たな分解を省いて前進消去と後退代入だけを用いて比較的少ない演算量で解くことができる。そのためレゾルベントの作用を与える連立1次方程式を解くのに直接法を用いる場合には、フィルタに「単一のレゾルベント」（の虚部）の多項式を用いる方法は「複数のレゾルベント」を用いる他の方法に比べて、分解しなければならない行列の数が複数ではなく1つになるので、行列分解の処理に掛かる計算量を少なくできる。

「単一のレゾルベント」からなる多項式をフィルタとし

て用いる場合に、そのレゾルベントのシフトに実数を用いる場合と虚数を用いる場合の得失については以下のようになる。

- 求めたい固有対の固有値の区間 $[a, b]$ が固有値分布の端にある場合には、レゾルベントのシフトに実数を用いることができる。レゾルベントのシフトに虚数を用いるのであれば、区間 $[a, b]$ の位置は任意に設定できる。
- レゾルベントの作用の計算は、シフトが実数の場合には実数の演算だけでできるが、シフトが虚数の場合には複素数の演算を用いる必要がある。それゆえ実数をシフトに選べると、演算と記憶に複素数を避けて実数だけを用いればよくなるので、行列分解に必要な演算量と記憶量を少なくできる利点がある。
- 実対称定値一般固有値問題の場合には、レゾルベントの作用を与える連立1次方程式については、シフトが実数で固有値分布の範囲の外側にある場合には係数行列は実対称定値であり、シフトが虚数の場合には係数行列は複素対称である。どちらの場合も行列の対称性を利用することで利用しない場合に比べて必要な演算量や記憶量を減らせる。

連立1次方程式はその係数行列が実対称定値であれば、ピボット選択をせずに数値的に安定に解ける。今の問題で係数行列が複素対称である場合にピボット選択をせずに解いた場合の数値的安定性は不明である。もしも数値安定性を保証するために係数行列の対称性の利用を諦めれば、たとえば片側（行）ピボット交換を行うLU分解法を用いることになる。その場合には行列が帯であるならば、分解の過程で下帯幅は元のままに保たれるが、上帯幅は元の2倍にまで広がる可能性がある。対称性を用いてピボット交換をしない場合に比べてLU分解を保持するための記憶量は3倍になり、分解を行うのに必要な演算量も増えてしまう。

- 得られる伝達関数の形状に関しては、計算に用いる数値と演算の精度が固定であるとする場合に、虚数をシフトに用いる場合に比べて実数を用いる場合が不利である点としては、両方のフィルタで遷移域の幅を与えるパラメタ μ を同じにする場合に、通過域における伝達率の最大最小比 $1/g_p$ が大きいので、得られる近似対の精度が均一ではなくなる傾向も大きいことがあげられる。

虚数のシフトを用いて中間固有値を求める方法に比べると実数のシフトを用いて下端固有対を求める方法は利用上の制約が強いが、固有値が固有値分布の端にある少数の固有対だけが必要で、しかも固有値の精度が5~6桁程度であれば十分であるような問題への適用が考えられる。たとえば弾性体の振動問題を有限要素法で離散化近似して得られる実対称定値一般固有値問題である。まず固有値は振動数の2乗なので正值になることが分かっている。またもと

もと離散化に用いた有限要素分割の細かさや要素内の展開基底関数の程度により、離散化の近似の程度が十分良いのは普通は振動数の低い少数の振動モードだけであるので、そのような少数の近似対だけが必要なのであれば、解に要求される要求もあまり高くない場合が多いであろうと考えられる。また構造物の問題では現実には理想的な幾何形状からのずれや材質の不均質さ、測定の誤差などもあるので、固有振動解をきわめて高い精度（たとえば有効桁が6~7桁以上）で求める必要性はあまりないであろうと考えられる。

レゾルベントの多項式であるフィルタは線形作用素なので、 m 個のベクトルの組 X にフィルタを適用して Y を作る計算 $Y \leftarrow \mathcal{F}X$ は、ベクトルに対する行列の乗算と同様に、 X や Y の各列の計算は相互に依存性なく完全に独立に行えるので、 m 個のベクトルの組を任意に分割して計算することができる。フィルタの作用の計算の中ではレゾルベントの適用が n 回現れるが、フィルタの作用の全体の計算が終わるまでは他の分割した組との間で処理の同期をとる必要はない。それに対してクリロフ部分空間法の系統の解法では、作用素（たとえばレゾルベント）を適用するたびにベクトルの組に対して直交化の処理が入るので、そこでは必ず処理の同期をとる必要がある。

ここで述べたフィルタ対角化法は、フィルタの作用を適用する（乱数を元にして生成された）ベクトルの数 m に関しては、通過域または遷移域に固有値が含まれる対の数を事前に見積もる必要がある。ただしフィルタを適用して得られたベクトルの組を調べることで、ベクトルの数が不足していればそのことを検出できるので、もしも実際に不足していた場合には、フィルタを適用して得られたベクトルを少しずつ追加して十分な数にまですることは（その「十分な数」があまり大きくなければ）可能である。

実対称定値一般固有値問題の少数の例題に対して、単一のレゾルベントを用いた多項式をフィルタとして採用し、そのレゾルベントのシフトが実数の場合と虚数の場合の両方について、倍精度演算を用いて数値実験を行い、通常の現実的な用途に対しては十分であろうと思われる程度の精度でもって近似固有対が得られることを確かめた。

参考文献

- [1] Toledo, S. and Rabani, E.: Very Large Electronic Structure Calculations Using an Out-of-Core Filter-Diagonalization Method, *J. Comput. Phys.*, Vol.180, No.1, pp.256-269 (2002).
- [2] Sakurai, T. and Sugiura, H.: A Projection Method for Generalized Eigenvalue Problems Using Numerical Integration, *J. Comput. Appl. Math.*, Vol.159, pp.119-128 (2003).
- [3] Sakurai, T. and Tadano, H.: CIRR: A Rayleigh-Ritz Type Method with Contour Integral for Generalized Eigenvalue Problems, *Hokkaido Math. J.*, Vol.36, No.4, pp.745-757 (2007).
- [4] Polizzi, E.: A Density Matrix-based Algorithm for Solv-

- ing Eigenvalue Problems, *Phys. Rev. B*, Vol.79, No.1, pp.115112-115117 (2009).
- [5] Ikegami, T., Sakurai, T. and Nagashima, U.: A Filter Diagonalization for Generalized Eigenvalue Problems Based on the Sakurai-Sugiura Projection Method, *J. Compu. Appl. Math.*, Vol.233, No.8, pp.1927-1936 (2010).
- [6] Galgon, M., Krämer, L. and Lang, B.: The FEAST Algorithm for Large Eigenvalue Problems, *PAMM: Proc. Appl. Math. Mech.*, Vol.11, pp.747-748 (2011).
- [7] Güttel, S., Polizzi, E., Tang, P.T.P. and Viaud, G.: Zolotarev Quadrature Rules and Load Balancing for the FEAST Eigensolver, *SIAM J. Sci. Comput*, Vol.37, No.4, pp.A2100-A2122 (2015).
- [8] Austin, A.P. and Trefethen, L.N.: Computing Eigenvalues of Real Symmetric Matrices with Rational Filters in Real Arithmetic, *SIAM J. Sci. Comput*, Vol.37, No.3, pp.A1365-A1387 (2015).
- [9] Murakami, H.: Filter Diagonalization Method by Using a Polynomial of a Resolvent as the Filter for a Real Symmetric-Definite Generalized Eigenproblem, *Eigenvalue Problems Algorithms, Software and Applications, in Petascale Computing: EPASA2015, Tsukuba, Japan, September 2015*, Springer LNCSE-117, ISBN 978-3-319-62424-2 (2018).
- [10] 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).
- [11] 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).
- [12] 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).
- [13] 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).
- [14] 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).
- [15] 村上 弘: フィルタ対角化法について, 日本応用数理学会 2014 年度年会予稿集, pp.329-330 (2014).
- [16] 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).
- [17] 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).
- [18] 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).
- [19] 村上 弘: 1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).
- [20] 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, 日本応用数理学会 2015 年度年会予稿集 (統合版), pp.442-443 (2015).
- [21] 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).

- [22] 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).
- [23] 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 (2016).
- [24] 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).



村上 弘 (正会員)

1960 年生. 1992 年北海道大学大学院理学博士号 (化学第二学専攻) 取得. 現在, 首都大学東京数理学専攻の准教授. 数理的な問題の数値的あるいは記号的方法による効率あるいは精度の良い解法やその並列化手法の研究に従事. 日本応用数理学会, 日本数式処理学会, 日本コンピュータ学会, SIAM, AMS, ACM, IEEE 各会員.