

直交化付きフィルタ適用による 固有値問題の近似対の反復改良について

村上 弘^{1,a)}

概要：実対称定値一般固有値問題 $A\mathbf{v} = \lambda B\mathbf{v}$ の固有対で固有値が指定された区間にあるものをフィルタを用いて近似して求める。フィルタの特性が悪く不要なベクトルの削減能力が不十分である、あるいは必要な固有ベクトルに対する伝達率の最大最小比が大きい場合には、ランダムなベクトルの組を B -正規直交化してフィルタを適用した結果から得られる近似対の精度は良くない。しかしそのような場合でも B -正規直交化してフィルタを適用する処理を数回反復すれば近似対を改良できることを数値実験により示す。

キーワード：フィルタ対角化, 固有値問題, レゾルベント, 固有対, 改良, 正規直交化

Refinements of Approximate Pairs of an Eigenproblem by Iterations of a Filter with an Orthonormalization

HIROSHI MURAKAMI^{1,a)}

Abstract:

For a real symmetric-definite generalized eigenproblem $A\mathbf{v} = \lambda B\mathbf{v}$, by using a filter we solve those approximate pairs whose eigenvalues are in a specified interval. When the property of the filter is not so good, such that the filter does not reduce those unnecessary eigenvectors so well or it has a large max-min ratio among transfer rates of eigenvectors which are required, then accuracies of those approximate pairs are not good which are calculated from the result of an application of the combination of a B -orthonormalization and a filtering to a set of randomly generated vectors. However, even in that case also we showed by numerical experiments that we can improve approximate eigenpairs by a few iterations of the combination of a B -orthonormalization and a filtering.

Keywords: filter diagonalization, eigenproblem, resolvent, eigenpair, improvement, orthonormalization

1. はじめに

(今回の報告は以前の報告 [34], [36] の改訂版である.)

フィルタ対角化法を用いて実対称定値一般固有値問題 $A\mathbf{v} = \lambda B\mathbf{v}$ の固有対 $(\lambda, \mathbf{v})$ でその固有値 λ が指定された区間 $[a, b]$ にあるものを近似して求める (文献 [4], [5], [6], [9]). フィルタ $\mathcal{F}$ はレゾルベント $\mathcal{R}(\rho) \equiv (A - \rho B)^{-1}B$ を用いて構成するが、たとえば以下の4種のを掲げられる。

- 1) (実数あるいは虚数をシフトとする) レゾルベントの線形結合の実部であるもの (式 (1)) (文献 [8], [9], [14], [16], [17], [18], [20]) .

$$\mathcal{F} \equiv c_{\infty} I + \operatorname{Re} \sum_{j=1}^n c_j \mathcal{R}(\rho_j). \quad (1)$$

- 2) 実数シフト ρ の単一のレゾルベントの実多項式 P であるもの (式 (2)) (文献 [22], [23], [26], [35]) .

$$\mathcal{F} \equiv P(\mathcal{R}(\rho)). \quad (2)$$

- 3) 虚数シフト ρ' の単一のレゾルベントの虚部の実多項式 P であるもの (式 (3)) (文献 [10], [19], [21], [22],

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

[23], [26], [30], [35]) .

$$\mathcal{F} \equiv P(\text{Im } \mathcal{R}(\rho')) . \quad (3)$$

- 4) 混合型(「レゾルベントの線形結合の実部」の実多項式)であるもの(文献[25], [27], [28], [29], [31], [32], [33]).

ベクトル $\mathbf{x}$ へのレゾルベント $\mathcal{R}(\rho)$ の作用 $\mathbf{y} \leftarrow \mathcal{R}(\rho)\mathbf{x}$ は、シフト行列 $C(\rho) \equiv A - \rho B$ を係数とする連立1次方程式 $C(\rho)\mathbf{y} = B\mathbf{x}$ を解いて実現する。この計算がフィルタの適用において主要となる部分である。

本報告では大規模な連立1次方程式は直接法で解くものと仮定する。その場合は行列分解の演算量と特に分解結果を保持する記憶量が計算を実施する上での制約になりがちであるので、必要なレゾルベントの数をなるべく減らしたいとする。

必要なレゾルベントの数が最も少ない1になるフィルタの構成として、「シフトが実数の単一のレゾルベントの実多項式」と「シフトが虚数の単一のレゾルベントの虚部の実多項式」がある。しかしこれらは、そのシフトや多項式を調整しても、急峻な遮断特性と通過域での伝達率の均一性の実現が困難であるので、伝達特性を十分に良くすることはできない(ちなみにシフトが虚数のものの方がシフトが実数のものに比べて実現できるフィルタの特性が良い)。

フィルタの特性はレゾルベントを2~4個用いることができればかなり改善できる(文献[27], [28], [29], [31], [32], [33])。しかし用いるレゾルベントの数に比例して行列分解に費やす演算量と(特に)分解結果を保持するための記憶量は増えてしまう。

用いたフィルタの特性が良くなければ得られる近似対の品質も良くない、そこで「特性の良くないフィルタ」の適用を反復することにより近似対を改良することを考える。同一のフィルタを反復して用いるのであれば、新たな行列分解は必要ではない。しかしフィルタの適用を単純に反復すると、それに伴って「必要な固有ベクトル」の間の伝達率の比が2乗、3乗と反復回数のべきで拡大してしまい、計算に用いている数値の有効精度の制約から、フィルタで濾過した後のベクトルに含まれる必要な固有ベクトルの情報は伝達率が相対的に小さいものについては有効な精度が低下したりあるいは消失してしまう。そこでこの反復による成長率の違いによる積み残しの誤差が原因の精度の低下抑止策として、フィルタを適用するたびに正規直交化を施すことにする。これは構造解析の分野において「(直交化付き)同時逆反復法」と呼ばれている方法(文献[2], [3], [11], [12], [13])の類似であるといえる(その原理は「直交化(同時)反復法」(Orthogonal Iteration)[7]である)。

そこで本報告では、ランダムに選んだ初期ベクトルの組から始めて、「 B -正規直交化を施してフィルタを適用する」

ことを数回反復することにより近似固有対を改良する実験を試みて、その効果を確認した。実験に用いたベクトルの組に対する B -正規直交化の方法は、 B を内積の計量とする特異値分解(B -SVD)である。

1.1 B -正規直交化法について

m 個のベクトルを列として並べた組 Y の B -正規直交化には、 B -特異値分解(B -SVD)を利用した。

- 1) m 次の対称正定値行列 $G \leftarrow Y^T B Y$ を作る。

- 2) 固有値分解 $G \Rightarrow U D U^T$ を Rutishauser の Jacobi 法 [1] で行い、 m 次の対角行列 D と直交行列 U を求める(D の対角は減少順に)。

- 3) $Y \leftarrow Y U$ は B -直交する列ベクトルの組。

元の Y が悪条件や、 G を作る際の丸め誤差の蓄積などから、上記の計算で得られた Y の列の B -直交性は十分ではないことがある。 B -直交性の向上を狙って、上記1)から3)までを高々2, 3回繰り返し、途中で G がほぼ対角になれば、 B -直交性は十分として繰り返しを抜ける。 G の対角要素の平方根(ノルム)が閾値未満の Y の列は棄てて、残った各列に対応するノルムの逆数を乗じれば、 Y は B -正規直交化されたベクトルの組になる。その際に新しい Y の列の数は、切断された列の数だけ減る。

注意すべき点として、フィルタの適用では対象とするベクトルの組を任意に分割して並行に処理できるが、直交化については対象となるベクトルの組が揃わないと処理を開始できないため、同期がそこで要求されてしまい並列分散処理用には不便になること、およびベクトルの組に対する直交化の操作は線形的作用ではないので、途中で直交化を挟んだ全体の処理は線形的作用素ではなくなることである。

2. フィルタ対角化法の概略

行列 A , B が実対称で B が正定値の一般固有値問題 (4) の固有対 $(\lambda, \mathbf{v})$ で固有値 λ が区間 $[a, b]$ にあるものを近似する。

$$A\mathbf{v} = \lambda B\mathbf{v} \quad (4)$$

フィルタは調整された線形作用素で、固有値が区間内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た固有ベクトルはなるべく阻止するようにする。するとフィルタは「必要な固有ベクトルで張られた不変部分空間」への射影作用素の近似になる。そこで線形独立性の高いベクトルのフィルタによる像ベクトルを十分多く集めて、その線形結合によりうまく不変部分空間の近似空間の基底を構成する。そのような基底に Rayleigh-Ritz 法を適用すると必要な近似対が得られる。

3. 今回の実験に用いた簡易型のフィルタ

フィルタは単一のレゾルベントの多項式型とした。その多項式にはチェビシェフ多項式を用いた。この種類のフィルタの伝達特性はあまり良いものにはならない。

区間 $[a, b]$ が固有値分布の下端であるならば、シフト ρ を実数にとることができて、フィルタは以下の式 (5) で与えられる。

$$\mathcal{F} \equiv g_S T_n(2\gamma \mathcal{R}(\rho) - I). \quad (5)$$

シフト ρ' を虚数に選べば区間 $[a, b]$ の位置は自由に設定できて、フィルタは以下の式 (6) で与えられる。

$$\mathcal{F} \equiv g_S T_n(2\gamma' \text{Im } \mathcal{R}(\rho') - I). \quad (6)$$

ここで $T_n(x)$ は x の n 次のチェビシェフ多項式を表し、 I は恒等作用素、そうして Im は虚部をとり出す作用素である。 γ や γ' は実数の定数であり、 g_S は阻止域に於ける伝達関数の大きさの上限である。

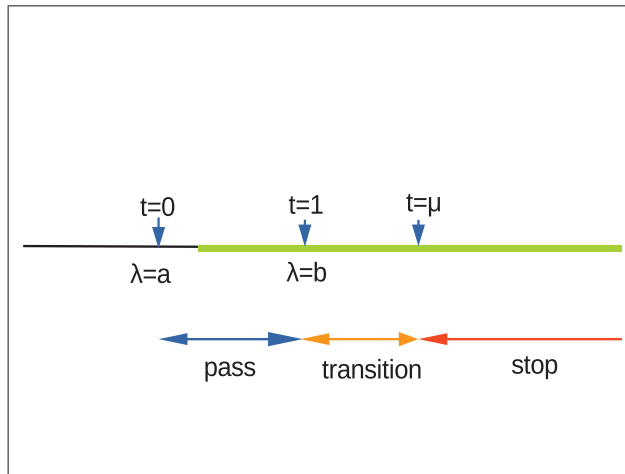


図 1 固有値 λ と正規化座標 t の関係 (下端固有値用)
通過域 $t \in [0, 1]$; 遷移域 $t \in (1, \mu)$; 阻止域 $t \in [\mu, \infty)$.
緑色の太線部分は固有値が存在する領域

3.1 実験に用いたフィルタの設計法

フィルタをパラメタの 3 つ組 (n, μ, g_S) で指定する (ほかのやり方として (μ, g_S, g_P) を指定してそれらを近似的に満たすようにすることも可能である)。ここで n はチェビシェフ多項式の次数であり、 μ は遷移域と阻止域の境界位置 (の正規化座標) を表す。 g_S は阻止域での伝達関数の大きさの上限値であり、伝達関数の通過域での値は g_P 以上 1 以下である。

各ベクトルに含まれる固有ベクトルのうちに必要なものに対する不要なものの割合は (数値丸め誤差の影響を無視する近似のもとでは) フィルタを適用するたびに g_S/g_P 倍以下に減少する。

以下に、フィルタの具体的な構成法を示す。

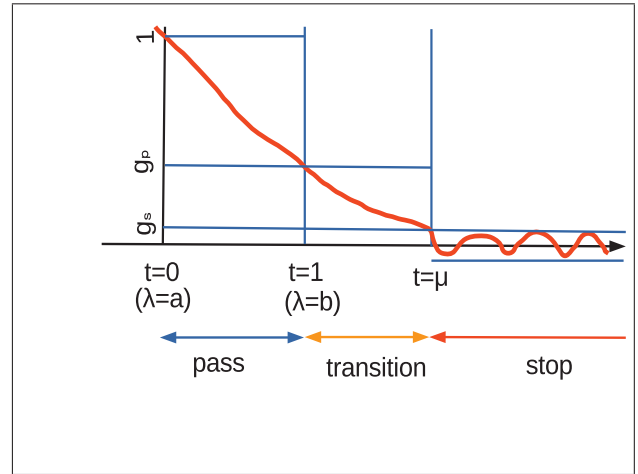


図 2 伝達関数 $g(t)$ の概形 (下端固有値用)

3.1.1 レゾルベントのシフトを実数にする場合

区間 $[a, b]$ は固有値分布の下端で、 a は最小固有値 $\lambda_{\min}$ 以下とする。指定されたパラメタの組 (n, μ, g_S) からシフト ρ とレゾルベントの係数 γ (および g_P) を以下の式 (7) で計算する。

$$\begin{cases} \sigma & \leftarrow \mu / \sinh \left(\frac{1}{2n} \cosh^{-1} \frac{1}{g_S} \right), \\ \rho & \leftarrow a - (b - a) \sigma, \\ \gamma & \leftarrow (b - a)(\sigma + \mu), \\ g_P & \leftarrow g_S \cosh \left\{ 2n \sinh^{-1} \sqrt{(\mu - 1)/(1 + \sigma)} \right\}. \end{cases} \quad (7)$$

すると、フィルタ $\mathcal{F}$ は以下の式 (8) で与えられる。

$$\mathcal{F} \equiv g_S T_n(2\gamma \mathcal{R}(\rho) - I). \quad (8)$$

$\lambda \in [a, b]$ を $t \in [0, 1]$ に移す線形変換 $t \equiv \frac{\lambda - a}{b - a}$ により、 λ に対する正規化座標 t を定義する。引数 t の伝達関数 $g(t)$ と、引数 λ の伝達関数 $f(\lambda)$ は以下の式 (9) になる：

$$\begin{cases} g(t) & = g_S T_n \left(2 \frac{\mu + \sigma}{t + \sigma} - 1 \right), \\ f(\lambda) & = g_S T_n \left(2\gamma \frac{1}{\lambda - \rho} - 1 \right). \end{cases} \quad (9)$$

3.1.2 レゾルベントのシフトを虚数にする場合

虚数のシフトを使えば、区間 $[a, b]$ の位置は任意にとれる。指定されたパラメタの組 (n, μ, g_S) から式 (10) でシフト ρ' とレゾルベントの係数 γ' (および g_P) を計算する。

$$\begin{cases} \sigma & \leftarrow \mu / \sinh \left(\frac{1}{2n} \cosh^{-1} \frac{1}{g_S} \right), \\ \rho' & \leftarrow \frac{a+b}{2} + \left(\frac{b-a}{2} \right) \sigma \sqrt{-1}, \\ \gamma' & \leftarrow \left(\frac{b-a}{2} \right) \frac{\mu^2 + \sigma^2}{\sigma}, \\ g_P & \leftarrow g_S \cosh \left\{ 2n \sinh^{-1} \sqrt{(\mu^2 - 1)/(1 + \sigma^2)} \right\}. \end{cases} \quad (10)$$

すると、フィルタ $\mathcal{F}$ は以下の式 (11) で与えられる。

$$\mathcal{F} \equiv g_S T_n (2\gamma' \operatorname{Im} \mathcal{R}(\rho') - I). \quad (11)$$

$\lambda \in [a, b]$ を $t \in [-1, 1]$ に移す線形変換 $\lambda \equiv \frac{1}{2}(a+b) + \frac{1}{2}(b-a)t$ で、 λ に対する正規化座標 t を定義する。

引数 t の伝達関数 $g(t)$ と引数 λ の伝達関数 $f(\lambda)$ は以下の式 (12) になる。

$$\begin{cases} g(t) &= g_S T_n \left(2 \frac{\mu^2 + \sigma^2}{t^2 + \sigma^2} - 1 \right), \\ f(\lambda) &= g_S T_n \left(2\gamma' \operatorname{Im} \frac{1}{\lambda - \rho'} - 1 \right). \end{cases} \quad (12)$$

4. 直交化付きフィルタ反復による改良法の概要

正規直交化付きでフィルタの適用を反復することで改良された近似固有対を求める計算手順を図 3 に示す。今回は、簡単のため同一のフィルタを IT 回 (2~4 回) 適用する。そうして毎回のフィルタの適用前ごとに必ず B -正規直交化を施すものとする。 B -正規直交化は閾値付きの B -特異値分解を用いて行い、閾値は用いた浮動小数点数システムのマシンイプシロンの 100 倍とし、特異値が閾値以下の特異ベクトルは除去した。

- (1) フィルタ $\mathcal{F}$ の準備として、ここでシフト行列 $C(\rho) \equiv A - \rho B$ を分解する；
- (2) $Y^{(0)} \leftarrow$ ランダムなベクトル m 個の組；
- (3) for $i := 1, 2, \dots, \text{IT}$ do
 $X^{(i)} \leftarrow Y^{(i-1)}$ の B -正規直交化；
 $Y^{(i)} \leftarrow \mathcal{F} X^{(i)}$ ；
enddo
注：途中の B -正規直交化においてベクトルの組 $Y^{(i-1)}$ の階数不足により $X^{(i)}$ の階数が切断を受けて低下したら、その後の $X^{(i)}$ や $Y^{(i)}$ のベクトルの数はその低下した階数へ変更。
- (4) 上記のステップ (3) で得られた最後の X と Y について、以下のようにする。
「必要な固有値全部を持つ不変部分空間」の近似空間の基底 Z を Y の列の線形結合で構成する (その際に X およびフィルタの伝達特性の値 g_S と g_P の情報も用いる) [15]。
- (5) 基底 Z 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する。

図 3 正規直交化付きフィルタ適用の反復による対角化の手順

4.1 フィルタの適用による固有ベクトルの含有比の変化

いま対象にしている一般固有値問題のすべての固有対を $(\lambda_i, \mathbf{v}_i)$, $i=1, 2, \dots, N$ とする。任意のベクトル $\mathbf{x}$ にフィルタ $\mathcal{F}$ を適用すると、対応する固有ベクトル展開では、任意の i 番目の固有ベクトル $\mathbf{v}_i$ の係数はそれにさらにフィルタの伝達率 $f(\lambda_i)$ を乗じたものになる。今回のフィルタの伝達関数の設定では固有値 λ が通過域にあれば $g_P \leq f(\lambda) \leq 1$ であり、固有値 λ が阻止域にあれば $|f(\lambda)| \leq g_S (\ll g_P)$ である。よってベクトル $\mathbf{x}$ にフィルタを適用すると、固有ベ

クトル展開においては固有値 λ_p が通過域にある固有ベクトル $\mathbf{v}_p$ の係数はどれもその $f(\lambda_p)$ 倍に変わり、固有値が阻止域にある任意の固有ベクトルに対する展開係数の大きさは g_S 以下の数が乗じられたものになる。

よって「(多くの) 阻止したい固有ベクトルの強度」の「通過させたい p 番目の固有ベクトルの強度」に対する割合は、フィルタの適用により (どれも) 適用前の $g_S/f(\lambda_p)$ 倍以下になる。この比の値は 1 よりも小さく、それは「特定の信号 p の強度に対するノイズの強度の比率」がフィルタの適用で受けた低減の倍率である。この倍率が小さいほど、フィルタの適用により除去したい固有値が阻止域にある固有ベクトルの「相対的」な含有率が減ることになる。通過させたい固有ベクトルの伝達率が大きいほどこの倍率は小さくなり、フィルタの適用により相対比で考えたときの品質改良の程度が高い。

通過域にある伝達率が g_P 以上の固有ベクトルのすべてを求める場合には、上記の比の値の上限は g_S/g_P になる。この上限は通過域にある固有値に対する伝達率の最小値が閾値 g_P に等しい場合にだけ達成されるものなので、最小値がそれよりも大きければそれだけ、実際の改良は上限からの予想よりも良いものになる。

5. 例題に用いた実対称定値一般固有値問題

例題とした実対称定値一般固有値問題 (13) は、1 辺の長さが π の立方体領域において零ディリクレ境界条件を課された 3 次元ラプラス作用素の固有値方程式 (14) を有限要素法 (FEM) を用いて離散化近似して得られたものである。

$$A \mathbf{v} = \lambda B \mathbf{v}. \quad (13)$$

$$-\Delta \Psi(x, y, z) = \lambda \Psi(x, y, z). \quad (14)$$

立方体の各辺方向を $N_1 + 1$, $N_2 + 1$, $N_3 + 1$ に等分割して、区間の直積による直方体を FEM の要素とした (図 4)。各要素内の基底関数には各辺方向の 3 重線形関数を用いた。一般固有値問題の行列 A と B の次数は $N = N_1 N_2 N_3$

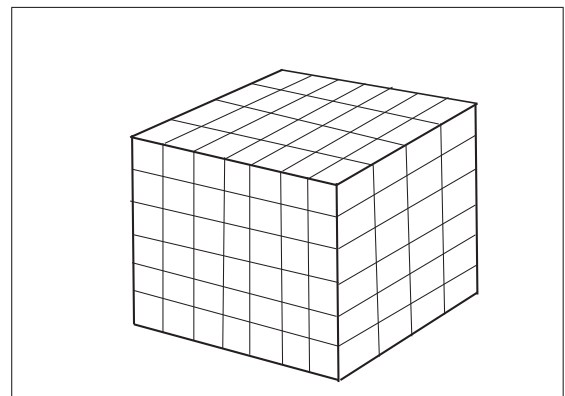


図 4 FEM 要素分割の概念図。 $(N_1, N_2, N_3) = (3, 5, 6)$ の例。

となる。

いま $N_1 \leq N_2 \leq N_3$ であるとして、帯幅が小さくなるように基底関数に番号を付けると、下帯幅（対角要素を含めない）は $w_L = 1 + N_1 + N_1 N_2$ にできる。離散化された一般固有値問題に対してフィルタ対角化法を適用して、固有値が区間 $[a, b]$ にある固有対の近似を求める。この例題の厳密な固有値は簡単な数式で計算できるので、任意区間内にある固有値の数は、厳密値の数え上げにより求められる。

5.1 例題の固有値の厳密値を与える式

辺長 π の立方体の各辺方向を $N_1 + 1, N_2 + 1, N_3 + 1$ に等分割した場合には、例題の 3 次元問題の固有値は添字の組 (k_1, k_2, k_3) で識別されて、各方向ごとの 1 次元問題の固有値の和で表される（式 (15)）。そうして各 1 次元問題の固有値は式 ((16)) で与えられる。ただしここで $\theta_k \equiv \frac{\pi k}{N+1}$ である。

$$E_{(k_1, k_2, k_3)}^{[N_1, N_2, N_3]} = \mathcal{E}_{k_1}^{[N_1]} + \mathcal{E}_{k_2}^{[N_2]} + \mathcal{E}_{k_3}^{[N_3]}, \quad k_i = 1, 2, \dots, N_i \quad (15)$$

$$\mathcal{E}_k^{[N]} = \frac{6k^2 (\sin \theta_k / \theta_k)^2}{(1 + \cos \theta_k)(2 + \cos \theta_k)}, \quad k = 1, 2, \dots, N. \quad (16)$$

6. 近似対の相対残差

近似対 $(\lambda, \mathbf{v})$ の相対残差 Θ を以下の式 (17) で定義する。

$$\Theta \equiv \frac{\|A\mathbf{v} - \lambda B\mathbf{v}\|}{\|\lambda B\mathbf{v}\|} = \frac{\|\mathbf{r}\|}{\|\lambda B\mathbf{v}\|}. \quad (17)$$

この Θ の値はベクトル $\mathbf{v}$ の規格化には依らないし、行列 A と B に共通の尺度を乗けても不変である。今回の実験ではベクトルのノルム $\|\cdot\|$ には 2-ノルムを用いた。近似対をこの相対残差 Θ を用いて評価し、値が小さいものほど精度が良いとする。相対残差の計算では複数の近似対の列ベクトルをまとめて行列 V にすると、 AV と BV を作るための行列 A と B 全体への記憶参照は 1 回ずつになり、しかも A と B が疎であるほど手間が少なく計算は容易になる。

ベクトルのノルムとして 2-ノルムを用いる場合の Θ の幾何学的な意味は以下のようなになる。図 5 中でベクトル $A\mathbf{v}$ と $\lambda B\mathbf{v}$ の挟む角を ϕ とすれば、関係式 $\sin \phi \leq \|\mathbf{r}\| / \|\lambda B\mathbf{v}\| = \Theta$ が成り立つ。

7. 直交化付きフィルタによる近似対の反復改良の実験例

直交化付きフィルタの反復により、改良された近似対を得る数値実験の具体的な例については、頁数が多いため記述を「付録」に移した。

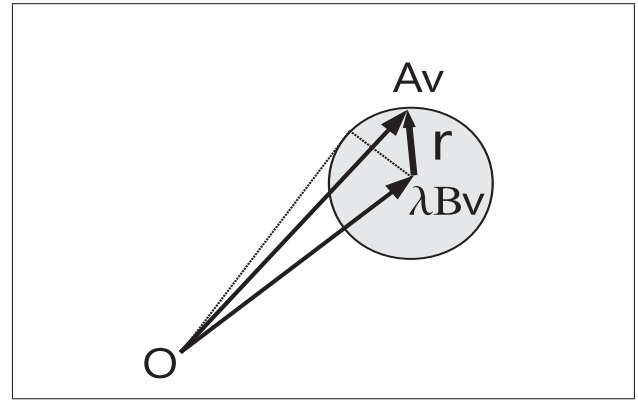


図 5 相対残差 Θ の幾何学的意味

8. おわりに

実対称定値一般固有値問題の固有対で固有値が指定された区間内のものをフィルタ対角化法を用いて近似して求めるものとする。フィルタはレゾルベントを利用して構成する。レゾルベントの作用はシフト行列を係数とする連立 1 次方程式を解くことで与えるが、その解法としては直接法を用いることにする。その場合は、問題が大規模であれば行列分解を行うための計算量と、分解を行った結果を格納するための記憶量が処理を実施する上での制約になりがちである。そこで制約を緩和するために、必要なレゾルベントの数が最も少ない「単一のレゾルベントの多項式型のフィルタ」を使いたいが、そのような種類のフィルタはあまり良い特性を持たせられないので、近似対の精度を十分高くできないという難点があった。そこで「特性のあまり良くないフィルタ」の適用を正規直交化と組み合わせることで反復することにより近似対を改良する手法を試みた。

本報告の付録中で行った実験により、あまり良くない特性のフィルタを使用しても、ベクトルの組に対して正規直交化を施してからフィルタを適用する操作を反復すれば近似対がうまく改良できることが確かめられた。

必要な近似対についての相対残差の最大値は、最初のうちは反復ごとに g_S/g_P 倍以下に減少し、数回の反復で十分小さくできることを実験で確かめた。

ただしベクトルの組に対する直交化の処理は組全体が揃うまで開始ができないので、計算を分散並列化するには不便になる。

レゾルベントの n 次多項式のフィルタを IT 回反復するには、最初に 1 度計算した行列分解の結果を保持できるのならば、その後はその分解結果を用いて m 個のベクトルの組を右辺とする連立 1 次方程式を合計で $n \times IT$ 回だけ「逐次的に」解く必要がある（そのほか合計で IT 回の B -正規直交化が各回のフィルタ適用の前に入る）。つまり処理を逐次化することと引き換えで、行列分解を行うために費やす演算量と得られた行列分解の結果を保持するための記憶量が低減できたことになる。

参考文献

- [1] Rutishauser, H.: The Jacobi Method for Real Symmetric Matrices, *Numer. Math.*, Vol.9, pp.1–10, (1966). Also in book: Bauer F.L.(eds) *Linear Algebra, Handbook for Automatic Computation*, vol.2, Springer (1971).
- [2] Rutishauser, H.: Computational aspects of F. L. Bauer's simultaneous iteration method, *Numer. Math.*, Vol.13, No.1, pp.4–13 (1969).
- [3] Rutishauser, H.: Simultaneous iteration method for symmetric matrices, *Handbook for Automatic Computation*, Springer-Verlag, pp.284–302 (1971). (Reprinted from *Numer. Math.*, Vol.16, pp.205–223 (1970).)
- [4] 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).
- [5] Polizzi, E.: A density matrix-based algorithm for solving eigenvalue problems, *Phys. Rev. B*, Vol.79, No.1, pp.115112–115117 (2009).
- [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] Golub, G.H. and Van Loan, C.F.: *Matrix computations*, 4th Ed., The John Hopkins Univ. Press (2013). (§8.2.4: 'Orthogonal Iteration').
- [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] 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).
- [10] Murakami, H.: 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).
- [11] 村田 健郎, 小国 力, 唐木 幸比古: 「スーパーコンピュータ: 科学技術計算への適用」, 丸善 (1985). (§8.1: ベキ乗法一族, §8.3: レーリー・リッツ法による同時逆反復法, §8.5: 一般固有値問題)
- [12] 村田 健郎, 三好 俊郎, ドンガラ, J.J., 長谷川 秀彦: 「行列計算ソフトウェア: WS, スーパーコン, 並列計算機」, 丸善 (1991). (§11.2: ベキ乗法一族, §11.4: レーリー・リッツ法による同時逆反復法, §11.8: 対称行列用の一般固有値問題)
- [13] 石上 裕之, 木村 欣司, 中村 佳正: 再直交化付ブロック逆反復法による固有ベクトルの並列計算, *HPCS2014 論文集*, pp.65–75 (2013).
- [14] 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).
- [15] 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).
- [16] 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).
- [17] 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).
- [18] 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).
- [19] 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).
- [20] 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).
- [21] 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).
- [22] 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).
- [23] 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*, Vol.2016-HPC-155, No.44, pp.1–27 (2016).
- [24] 村上 弘: レゾルベントの多項式によるフィルタを用いた実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.2016-HPC-157, No.4, pp.1–15 (2016).
- [25] 村上 弘: 実対称定値一般固有値問題を解くためのレゾルベントの多項式型フィルタの設計について, *情報処理学会研究報告*, Vol.2017-HPC-158, No.7, pp.1–10 (2017).
- [26] 村上 弘: チェビシェフ展開形で表わされた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.2019, pp.96–112 (2017).
- [27] 村上 弘: 実対称定値一般固有値問題を解くための少数のレゾルベントの多項式を用いたフィルタの設計法, *情報処理学会研究報告*, Vol.2017-HPC-159, No.4, pp.1–13 (2017).
- [28] 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実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-160, No.32, pp.1–32 (2017).
- [29] 村上 弘: 少数のレゾルベントで構成された多項式型フィルタによる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-161, No.7, pp.1–13 (2017).
- [30] 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).
- [31] 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).
- [32] 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).
- [33] 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).
- [34] 村上 弘: フィルタ対角化法による近似固有対の精度の改良について, *情報処理学会研究報告*, Vol.2018-HPC-167, No.29, pp.1–31 (2018).
- [35] 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).
- [36] 村上 弘: フィルタ対角化法による固有値問題の近似対の改良, *情報処理学会研究報告*, Vol.2019-HPC-168, No.18, pp.1–36 (2019).

付 録

A.1 B-正規直交化付きフィルタ適用の反復実験例（倍精度計算）

本節の数値実験で用いた数値とその演算にはすべて倍精度の浮動小数点（IEEE 754 規格の 2 進 64bit）を用いた。その有効精度は 10 進で約 15.95 桁である。

実験に使用した下端固有対用のフィルタ

固有値が下端の固有対（下端固有対）を求めるために用いたフィルタは、シフトが実数である単一のレゾルベントの n 次多項式である。しかも簡易型として多項式には Chebyshev 多項式を用いた。反復中ではフィルタは毎回同じものを用いた。

フィルタはパラメタの組 (n, μ, g_S) で指定したが、この節で紹介する実験例では常に $\mu = 1.5$, $g_S = 10^{-12}$ とし、次数 n だけを変えている（表 A.1, 図 A.1）。

g_P の値は、フィルタの通過域に於ける伝達率の最小値（最大値は 1）で、たとえば $n = 8$ の場合に $g_P = 8.8 \times 10^{-9}$ であるから、求めようとする固有ベクトルに対する伝達率は最大で約 8 桁違い得る。ある 1 つのベクトルに伝達率の異なる（正規化された）固有ベクトルが仮に同程度の係数で含まれていれば、それにフィルタを適用すると伝達率の小さい側は大きい側に比べて含まれている割合が約 8 桁減少して、数値丸めの影響で、そのベクトルの数値から取り出せる固有ベクトルの有効な精度は伝達率の小さい側のものは大きい側のものに比べて 8 桁程度失われる可能性があると考えられる。

A.1.1 実験に使用した中間固有対用のフィルタ

固有値が中間の固有対（中間固有対）を求めるために用いたフィルタは、シフトが虚数である単一のレゾルベントの虚部の n 次多項式である。しかも簡易型として多項式には Chebyshev 多項式を用いた。反復中ではフィルタは毎回同じものを用いた。

フィルタはパラメタの組 (n, μ, g_S) で指定したが、この節で紹介する実験例では常に $\mu = 1.5$, $g_S = 10^{-12}$ とし、次数 n だけを変えている（表 A.2, 図 A.2）。

シフトが虚数のフィルタは、シフトが実数であるものに比べて、ここでの $\mu = 1.5$, $g_S = 10^{-12}$ と固定している場合には表 A.1 と表 A.2 を見比べると次数 n が同じ場合に g_P の値が 2 桁弱大きいので伝達率の通過域全体での均一性が良くなっていること、そうして誤差縮小倍率の上限 g_S/g_P についても 2 桁弱小さいので反復あたりの近似対の改良がそれだけ速く進むと期待できる。

表 A.1 （下端固有対用フィルタ）通過域での最小の伝達率 g_P と誤差縮小倍率の上限 g_S/g_P

n	g_P	g_S/g_P
8	8.80×10^{-9}	1.14×10^{-4}
10	4.21×10^{-8}	2.38×10^{-5}
15	4.17×10^{-7}	2.40×10^{-6}
20	1.22×10^{-6}	8.23×10^{-7}

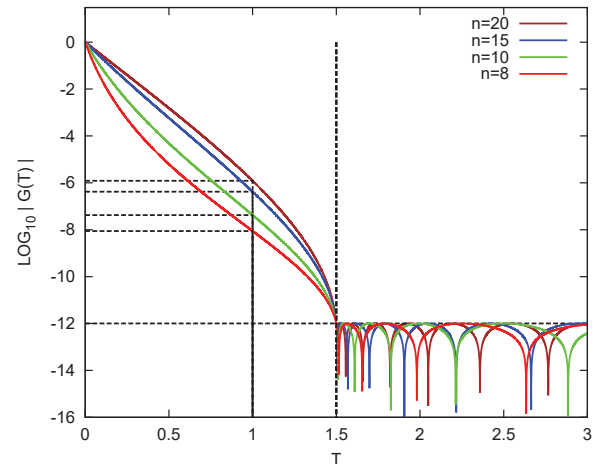


図 A.1 （下端固有対用フィルタ）伝達関数の大きさ $|g(t)|$ ($\mu = 1.5$, $g_S = 10^{-12}$)

表 A.2 （中間固有対用フィルタ）通過域での最小の伝達率 g_P と誤差縮小倍率の上限 g_S/g_P

n	g_P	g_S/g_P
8	5.91×10^{-7}	1.69×10^{-6}
10	4.20×10^{-6}	2.38×10^{-7}
15	5.56×10^{-5}	1.80×10^{-8}
20	1.63×10^{-4}	6.13×10^{-9}

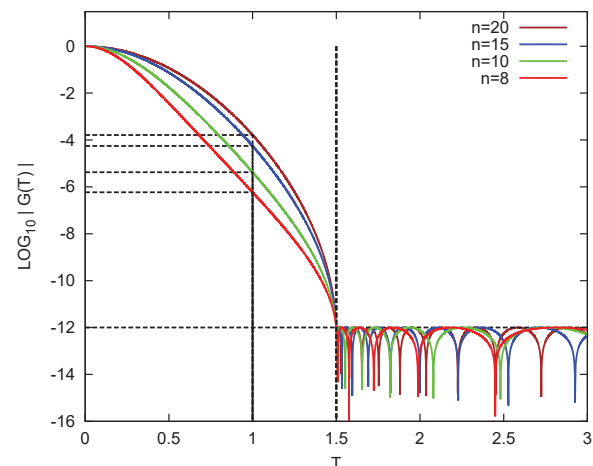


図 A.2 （中間固有対用フィルタ）伝達関数の大きさ $|g(t)|$ ($\mu = 1.5$, $g_S = 10^{-12}$, 右半分)

A.1.2 例題 R1：下端固有対用フィルタの反復（その1）

例題 R1 の FEM の要素分割は $(N_1, N_2, N_3) = (20, 30, 40)$ とした。それから導かれる一般固有値問題の行列 A と B の次数は $N = 24,000$ であり、下帯幅は $w_L = 621$ である。求める固有対は固有値が区間 $[a, b] = [0, 30]$ にあるものとした。そのような固有対の数は 54 個である。また、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a, b'] = [0, 45]$ にある固有値の数は 106 個であるので、フィルタで濾過するベクトルの数 m は 106 よりも多くすることが望ましい。

A.1.2.1 例題 R1：近似対の相対残差の反復による改良

右に掲げた図 A-3 から図 A-6 までの 4 枚の図は、フィルタの次数 n はどれも 8 に固定しておいてフィルタで濾過するベクトルの数 m をそれぞれ 80, 100, 120, 140 と変えた場合について、区間 $[a, b] = [0, 30]$ 内に固有値があるすべての近似対について横軸にその固有値をとり縦軸には相対残差 θ の常用対数値をとったグラフをフィルタの適用回数 IT を 1 から 4 の各場合に対してそれぞれ赤、緑、青、茶の折れ線でプロットしたものである。用いたフィルタの伝達率は固有値が通過域の右側に行くほど小さくなっているため、相対残差は区間の右側で大きくなる傾向を示している。そうして伝達率が大きい固有値を持つ近似対であるほど反復のたびに相対残差がより速やかに減少しており、固有値が区間の右側にある近似対の相対残差の減少は緩やかである傾向が見てとれる。ベクトルの数が $m = 80$ の場合には、反復回数 4 回目でもまだ区間の右側では残差は十分に小さくなっていないが、 $m = 100$ の場合には反復回数 4 回目でもまだ少し相対残差が減らせる余地がある程度になっていることが m が 106 を越えた場合の $m = 120$ や $m = 140$ の結果と比べることで見えてくる。

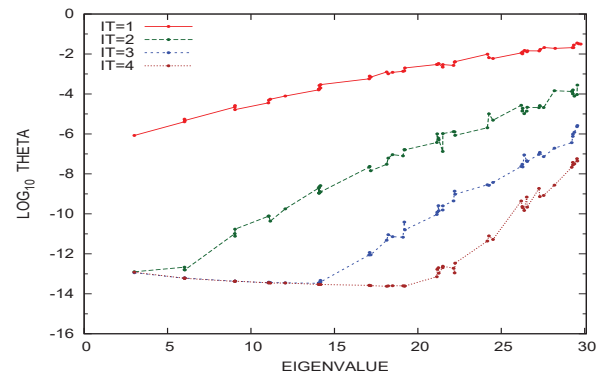


図 A-3 例題 R1：固有値と相対残差（次数 $n = 8$ ，ベクトルの数 $m = 80$ ）※ベクトルの数は 106 よりも多いことが望ましい

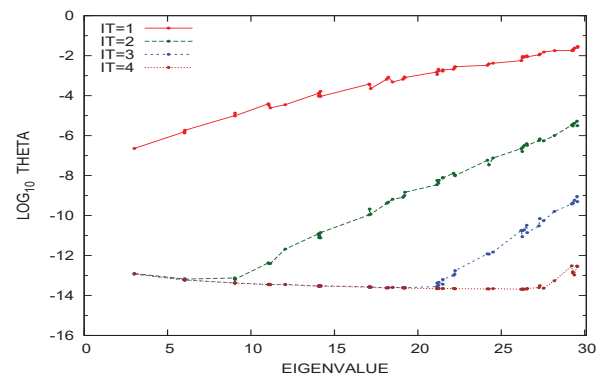


図 A-4 例題 R1：固有値と相対残差（次数 $n = 8$ ，ベクトルの数 $m = 100$ ）※ベクトルの数は 106 よりも多いことが望ましい

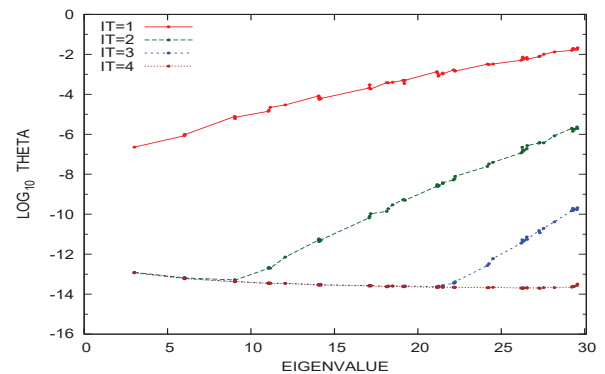


図 A-5 例題 R1：固有値と相対残差（次数 $n = 8$ ，ベクトルの数 $m = 120$ ）

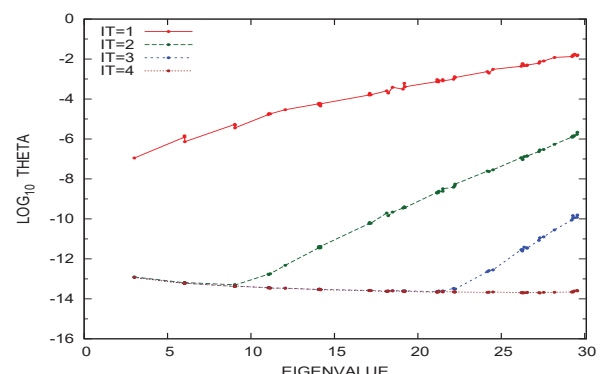


図 A-6 例題 R1：固有値と相対残差（次数 $n = 8$ ，ベクトルの数 $m = 140$ ）

A.1.2.2 例題 R1：反復による相対残差の最大値の減少

右に掲げた図 A・7 から図 A・10 までの 4 枚の図は、フィルタの次数 n をそれぞれ 8, 10, 15, 20 と変えた場合についてのものであり、ベクトルの数 m を 80, 100, 120, 140 とした各場合について色を茶, 青, 緑, 赤と変えた折れ線グラフとして横軸には反復回数 IT を 1 から 4 までとり、縦軸には固有値が区間 $[a, b] = [0, 30]$ に含まれる近似対の相対残差 θ の最大値を対数でとってプロットした。黒線は相対残差が反復ごとに g_S/g_P 倍になって減少すると仮定する場合のグラフの「傾き」を示すもので比較用である。

フィルタの次数 n が異なるなどの図についても、ベクトルの数 m が 106 より大きい $m = 120$ と $m = 140$ のグラフはほぼ重なっており、それらは反復 2 回目までは相対残差の最大値は黒線のグラフとほぼ同じ傾きで減少した様子が見てとれる。そうして相対残差の値が 10^{-13} 程度に達したところで減少は停滞している。

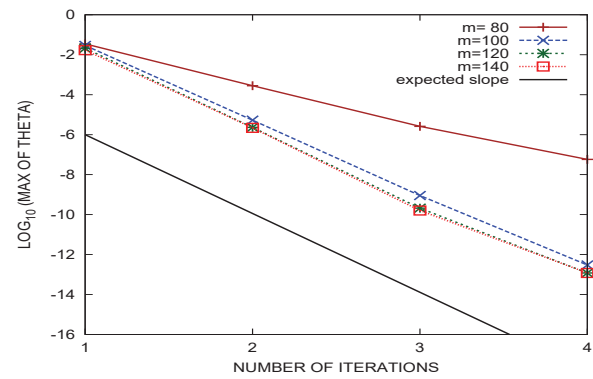


図 A・7 例題 R1：反復回数と相対残差の最大値（次数 $n = 8$, ベクトルの数 m は 106 よりも多いことが望ましい）

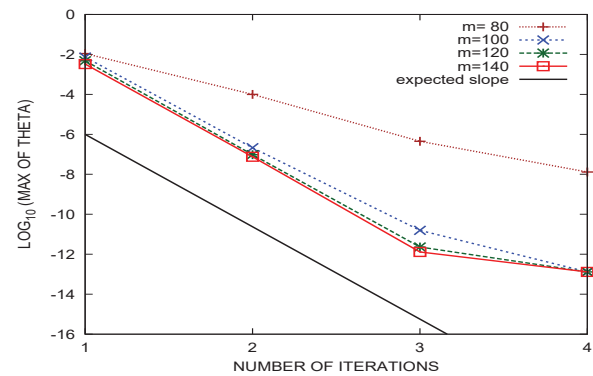


図 A・8 例題 R1：反復回数と相対残差の最大値（次数 $n = 10$, ベクトルの数 m は 106 よりも多いことが望ましい）

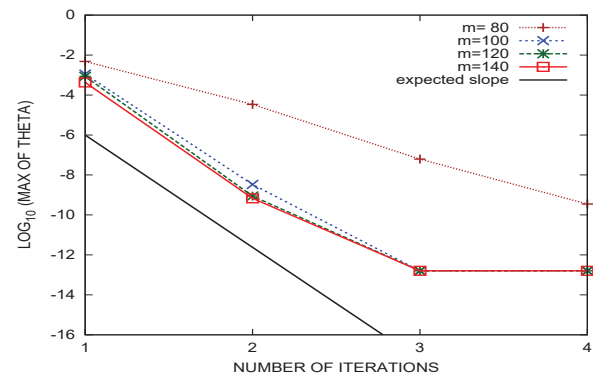


図 A・9 例題 R1：反復回数と相対残差の最大値（次数 $n = 15$, ベクトルの数 m は 106 よりも多いことが望ましい）

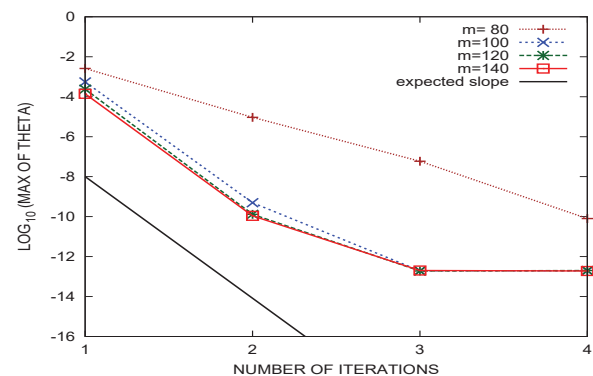


図 A・10 例題 R1：反復回数と相対残差の最大値（次数 $n = 20$, ベクトルの数 m は 106 よりも多いことが望ましい）

A.1.3 例題 R2：下端固有対用フィルタの反復（その2）

例題 R2 の FEM の要素分割は $(N_1, N_2, N_3) = (20, 30, 40)$ とした。それから導かれる一般固有値問題の行列 A と B の次数は $N = 24,000$ であり、下帯幅は $w_L = 621$ である。求める固有対は固有値が区間 $[a, b] = [0, 100]$ にあるものとした。そのような固有対の数は 378 個である。また $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a, b'] = [0, 150]$ にある固有値の数は 700 個であるので、フィルタで濾過するベクトルの数 m は 700 よりも多くすることが望ましい。

A.1.3.1 例題 R2：反復による相対残差の最大値の減少

右に掲げた図 A-11 から図 A-14 までの 4 枚の図は、フィルタの次数 n をそれぞれ 8, 10, 15, 20 と変えた場合についてのものであり、ベクトルの数 m を 500, 600, 700, 800 とした各場合について色を茶、青、緑、赤と変えた折れ線グラフとして横軸には反復回数 IT を 1 から 4 までとり、縦軸には固有値が区間 $[a, b] = [0, 100]$ に含まれる近似対の相対残差 θ の最大値を対数でとってプロットした。黒線は相対残差が反復ごとに g_S/g_P 倍になって減少すると仮定する場合のグラフの「傾き」を示すもので比較用である。

フィルタの次数 n が異なるどの図についても、ベクトルの数 m が 700 と 800 の場合のグラフは近く、それらはどの次数 n の場合にも反復 2 回目までは黒線のグラフとほぼ同じ傾きで相対残差の最大値が減少した様子が見てとれる。そして相対残差の値が 10^{-13} 程度に達したところで減少は停滞している。

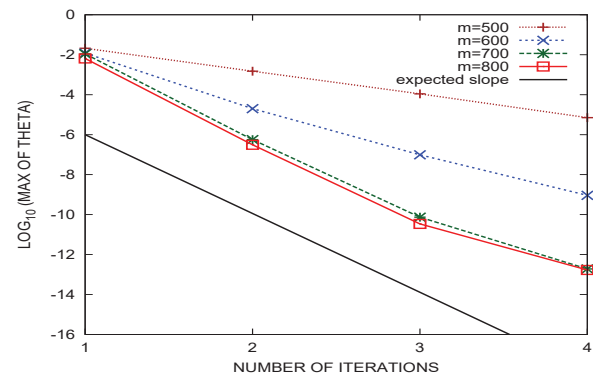


図 A-11 例題 R2：反復回数と相対残差の最大値（次数 $n = 8$ ，ベクトルの数 m は 700 よりも多いことが望ましい）

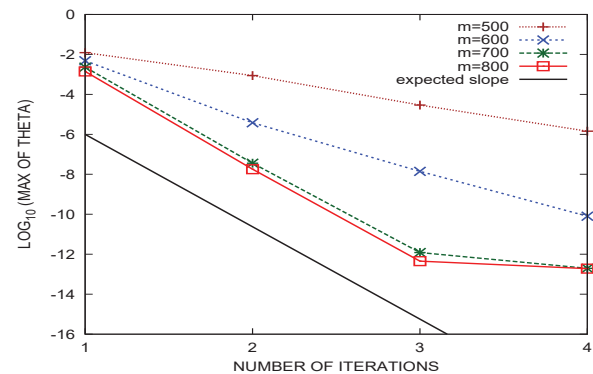


図 A-12 例題 R2：反復回数と相対残差の最大値（次数 $n = 10$ ，ベクトルの数 m は 700 よりも多いことが望ましい）

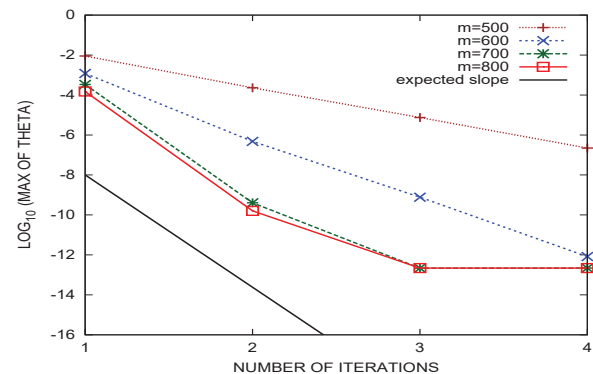


図 A-13 例題 R2：反復回数と相対残差の最大値（次数 $n = 15$ ，ベクトルの数 m は 700 よりも多いことが望ましい）

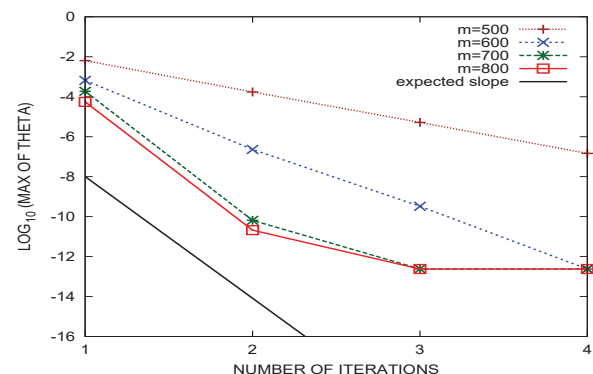


図 A-14 例題 R2：反復回数と相対残差の最大値（次数 $n = 20$ ，ベクトルの数 m は 700 よりも多いことが望ましい）

A.1.4 例題 C1：中端固有対用フィルタの反復（その 1）

例題 C1 の FEM の要素分割は $(N_1, N_2, N_3) = (20, 30, 40)$ とした。それにより導かれる一般固有値問題の行列 A と B の次数は $N = 24,000$ であり、下帯幅は $w_L = 621$ である。求める固有対は固有値は区間 $[a, b] = [300, 310]$ にあるものとした。そのような固有対の数は 90 個である。また、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a', b'] = [297.5, 312.5]$ にある固有値の数は 125 個であるので、フィルタで濾過するベクトルの数 m は 125 よりも多くすることが望ましい。

A.1.4.1 例題 C1：反復による相対残差の最大値の減少

右に掲げた図 A・15 から図 A・18 までの 4 枚の図は、フィルタの次数 n をそれぞれ 8, 10, 15, 20 と変えた場合についてのものであり、ベクトルの数 m を 100 から刻み 10 で 150 までとした各場合について色を変えた折れ線グラフとして横軸には反復回数 IT を 1 から 4 までとり、縦軸には固有値が区間 $[a, b] = [300, 310]$ に含まれる近似対の相対残差 θ の最大値を対数でとってプロットした。黒線は相対残差が反復ごとに g_S/g_P 倍になって減少すると仮定する場合のグラフの「傾き」を示すもので比較用である。

フィルタの次数 n が異なるどの図についても、ベクトルの数 m が 125 よりも大きい $m = 130, m = 140, m = 150$ の場合のグラフは非常に近く、それらはどの次数 n の場合にも反復 2 回目までは黒線のグラフとほぼ同じ傾きで相対残差の最大値が減少した様子が見てとれる。そうして相対残差の値が 10^{-14} 程度に達したところで減少は停滞している。

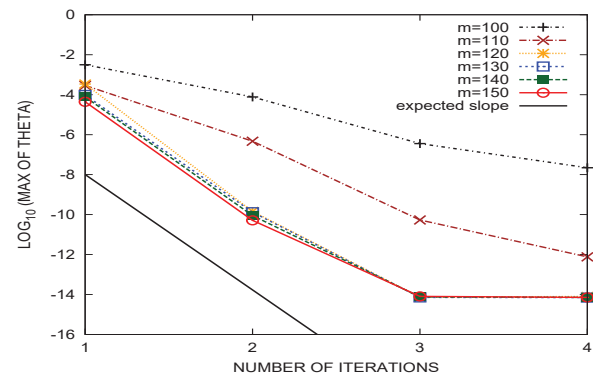


図 A・15 例題 C1：反復回数と相対残差の最大値（次数 $n = 8$ 、ベクトルの数 m は 125 よりも多いことが望ましい）

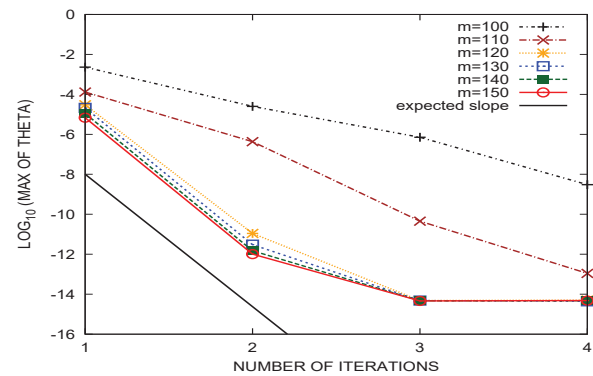


図 A・16 例題 C1：反復回数と相対残差の最大値（次数 $n = 10$ 、ベクトルの数 m は 125 よりも多いことが望ましい）

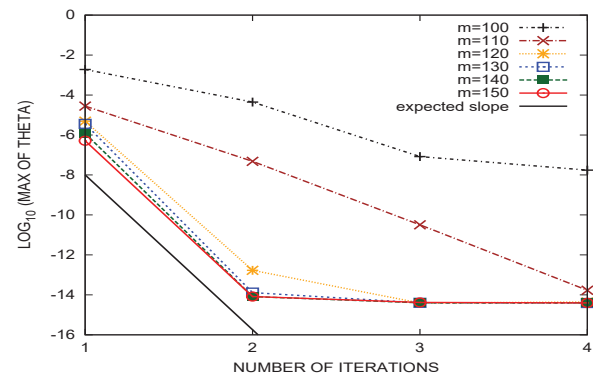


図 A・17 例題 C1：反復回数と相対残差の最大値（次数 $n = 15$ 、ベクトルの数 m は 125 よりも多いことが望ましい）

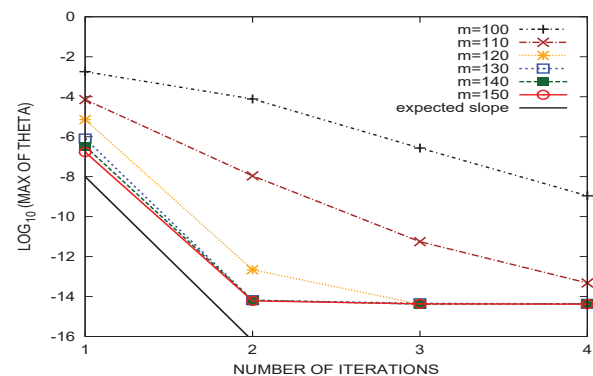


図 A・18 例題 C1：反復回数と相対残差の最大値（次数 $n = 20$ 、ベクトルの数 m は 125 よりも多いことが望ましい）

A.1.5 例題 C2：中端固有対用フィルタの反復（その2）

例題 C2 の FEM の要素分割は $(N_1, N_2, N_3) = (20, 30, 40)$ とした。それから導かれる一般固有値問題の行列 A と B の次数は $N = 24,000$ であり、下帯幅は $w_L = 621$ である。求める固有対の固有値は区間 $[a, b] = [1000, 1010]$ にあるものとした。そのような固有対の数は 92 個である。また、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a', b'] = [997.5, 1012.5]$ にある固有値の数は 145 個であるので、フィルタで濾過するベクトルの数 m は 145 よりも多くすることが望ましい。

A.1.5.1 例題 C2：反復による相対残差の最大値の減少

右に掲げた図 A-19 から図 A-22 までの 4 枚の図は、フィルタの次数 n をそれぞれ 8, 10, 15, 20 と変えた場合についてのものであり、ベクトルの数 m を 100 から刻み 10 で 150 までとした各場合について色を変えた折れ線グラフとして横軸には反復回数 IT を 1 から 4 までとり、縦軸には固有値が区間 $[a, b] = [1000, 1010]$ に含まれる近似対の相対残差 θ の最大値を対数でとってプロットした。黒線は相対残差が反復ごとに g_S/g_P 倍になって減少すると仮定する場合のグラフの「傾き」を示すもので比較用である。

フィルタの次数 n が異なるどの図についても、ベクトルの数 m が $m = 140$, $m = 150$ の場合のグラフは非常に近く、それらはどの次数 n の場合にも反復 2 回目までは黒線のグラフとほぼ同じ傾きで相対残差の最大値が減少した様子が見てとれる。そうして相対残差の値が 10^{-14} 程度に達したところで減少は停滞している。

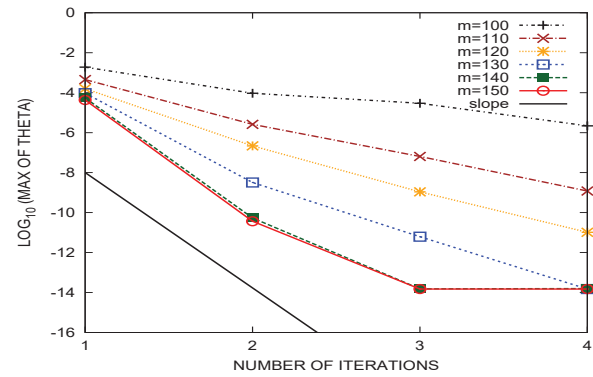


図 A-19 例題 C2：反復回数と相対残差の最大値（次数 $n = 8$, ベクトルの数 m は 145 よりも多いことが望ましい）

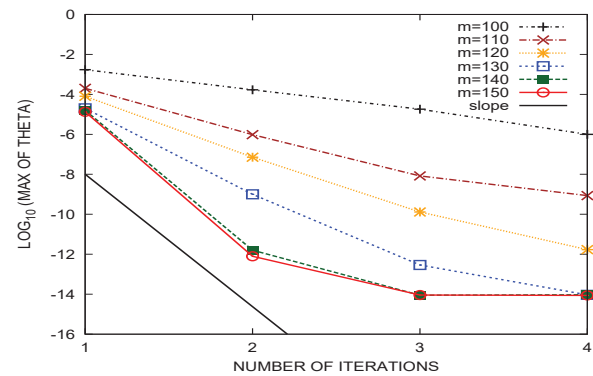


図 A-20 例題 C2：反復回数と相対残差の最大値（次数 $n = 10$, ベクトルの数 m は 145 よりも多いことが望ましい）

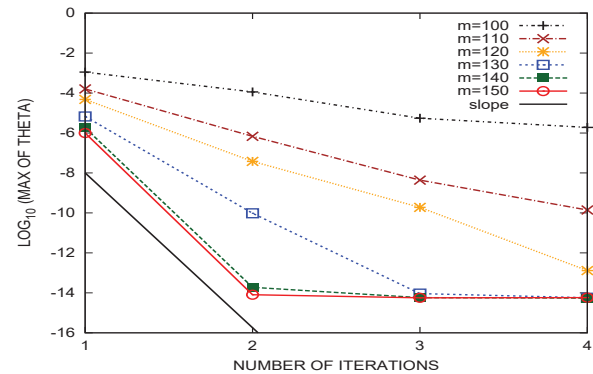


図 A-21 例題 C2：反復回数と相対残差の最大値（次数 $n = 15$, ベクトルの数 m は 145 よりも多いことが望ましい）

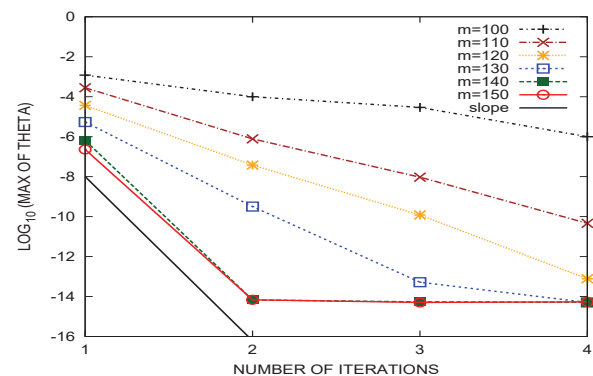


図 A-22 例題 C2：反復回数と相対残差の最大値（次数 $n = 20$, ベクトルの数 m は 145 よりも多いことが望ましい）

A.1.6 例題 C3：中端固有対用フィルタの反復（その 3）

例題 C2 では FEM の要素分割を $(N_1, N_2, N_3) = (50, 60, 70)$ とした。それから導かれる一般固有値問題の行列 A と B の次数は $N = 210,000$ であり、下帯幅は $w_L = 3051$ である（これは行列次数がある程度大きい問題である）。求める固有対の固有値は区間 $[a, b] = [100, 200]$ にあるものとした。そのような固有対の数は 801 個である。また、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a', b'] = [75, 225]$ にある固有値の数は 1192 個であるので、フィルタで濾過するベクトルの数 m は 1192 よりも多くすることが望ましい。

A.1.6.1 例題 C3：反復による相対残差の最大値の減少

右に掲げた図 A-23 から図 A-26 までの 4 枚の図は、フィルタの次数 n をそれぞれ 8, 10, 15, 20 と変えた場合についてのものであり、ベクトルの数 m を 1000, 1100, 1200, 1300 とした各場合について色を変えた折れ線グラフとして横軸には反復回数 IT を 1 から 4 までとり、縦軸には固有値が区間 $[a, b] = [100, 200]$ に含まれる近似対の相対残差 θ の最大値を対数でとってプロットした。黒線は相対残差が反復ごとに g_S/g_P 倍になって減少すると仮定する場合のグラフの「傾き」を示すもので比較用である。

フィルタの次数 n が異なるなどの図についても、ベクトルの数 m が 1192 を越えた $m = 1200$ と $m = 1300$ の場合のグラフはかなり近くて、それらは（特に $m = 1300$ の場合は）どの次数 n の場合にも反復 2 回目までは黒線のグラフとほぼ同じ傾きかあるいは少しより急な傾きで相対残差の最大値が減少した様子が見てとれる。そうして相対残差の値が 10^{-13} 程度に達したところで減少は停滞している。

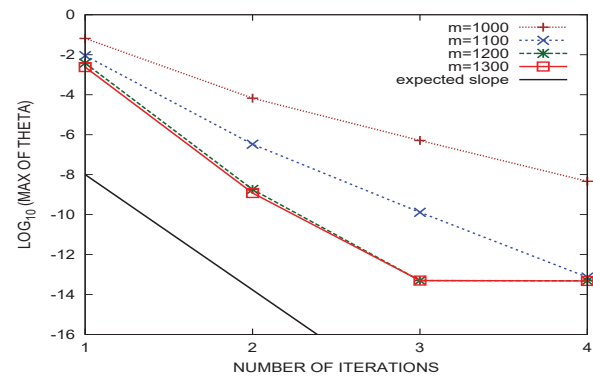


図 A-23 例題 C3：反復回数と相対残差の最大値（次数 $n = 8$, ベクトルの数 m は 1192 よりも多いことが望ましい）

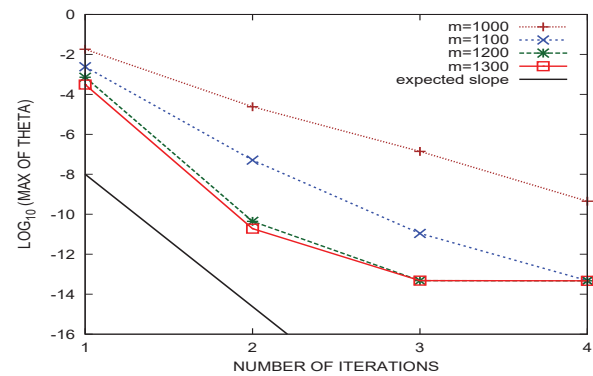


図 A-24 例題 C3：反復回数と相対残差の最大値（次数 $n = 10$, ベクトルの数 m は 1192 よりも多いことが望ましい）

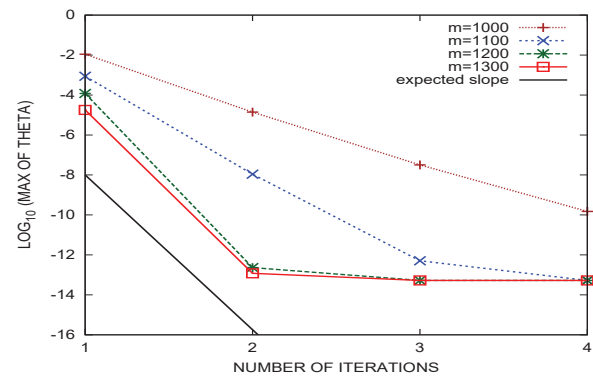


図 A-25 例題 C3：反復回数と相対残差の最大値（次数 $n = 15$, ベクトルの数 m は 1192 よりも多いことが望ましい）

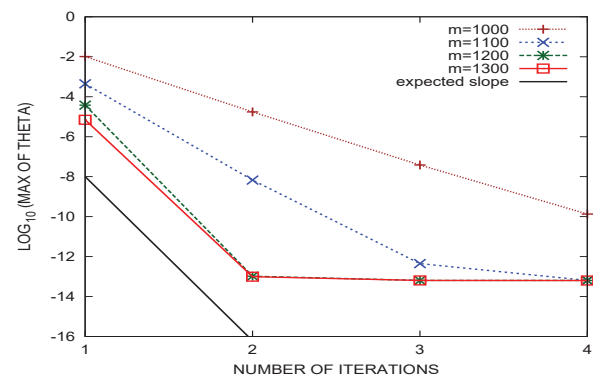


図 A-26 例題 C3：反復回数と相対残差の最大値（次数 $n = 20$, ベクトルの数 m は 1192 よりも多いことが望ましい）

A.2 B-正規直交化付きフィルタ適用の反復実験例（四倍精度計算）

本節では、実数型の数値とその演算には intel Fortran コンパイラがサポートしている四倍精度（IEEE 754 規格の 2 進 128bit）を用いた。四倍精度の浮動小数点数の有効精度は 10 進では約 34.02 桁である。

なお本節で扱う例題（R1-Q, R2-Q, C1-Q, C2-Q）はどれも、前節では倍精度で行っていた一般固有値問題の例題（R1, R2, C1, C2）の計算を、同じ問題設定で四倍精度で行ったものである。

A.2.1 例題 R1-Q：例題 R1 の四倍精度計算

例題 R1-Q の FEM 要素分割は $(N_1, N_2, N_3) = (20, 30, 40)$ とした。それから導かれる一般固有値問題の行列 A と B の次数は $N = 24,000$ であり、下帯幅は $w_L = 621$ である。求める固有対は固有値が区間 $[a, b] = [0, 30]$ にあるものとした。そのような固有対の数は 54 個である。また、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a, b'] = [0, 45]$ にある固有値の数は 106 個であるので、フィルタで濾過するベクトルの数 m は 106 よりも多くすることが望ましい。

例題 R1-Q：反復による相対残差の最大値の減少の様子

右に掲げた図 A-27 から図 A-30 までの 4 枚の図は、フィルタの次数 n をそれぞれ 8, 10, 15, 20 と変えた場合について、反復回数 IT を 1 から 8 まで横軸にとり、固有値が区間 $[a, b]$ に含まれる近似対の相対残差 θ の最大値を縦軸に対数でとって、フィルタで濾過した最初の乱数ベクトルの数 m を 80, 100, 120, 140 としたそれぞれの場合について色を（茶、青、緑、赤と）変えて折れ線でプロットしたグラフである。黒い線は、相対残差が反復ごとに g_S/g_P 倍になって減少する、と仮定する場合のグラフの「傾き」を示すためのものである。

フィルタの次数 n が異なる 4 枚の図のどれについても、ベクトルの数 m が 106 より大きい $m = 120$ と $m = 140$ の 2 つのグラフは線がほぼ重なっており、反復 5 回目までは黒線のグラフとほぼ同じ傾きで減少していることが見てとれる。そうして、減少は相対残差の値が 10^{-31} 程度に達したところで停滞している。

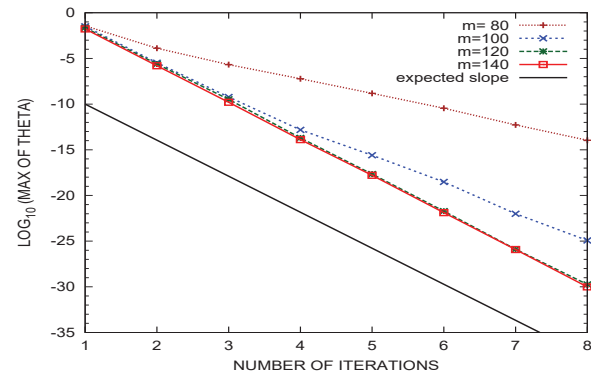


図 A-27 例題 R1-Q：反復回数と相対残差の最大値（次数 $n = 8$ ，ベクトルの数 m は 106 よりも多いことが望ましい）

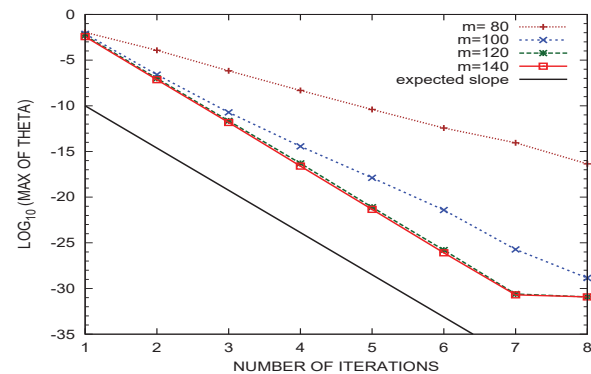


図 A-28 例題 R1-Q：反復回数と相対残差の最大値（次数 $n = 10$ ，ベクトルの数 m は 106 よりも多いことが望ましい）

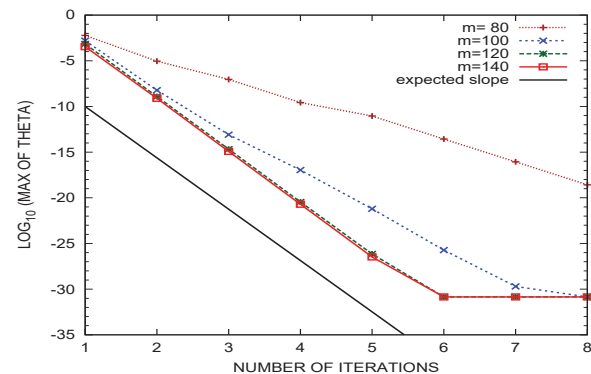


図 A-29 例題 R1-Q：反復回数と相対残差の最大値（次数 $n = 15$ ，ベクトルの数 m は 106 よりも多いことが望ましい）

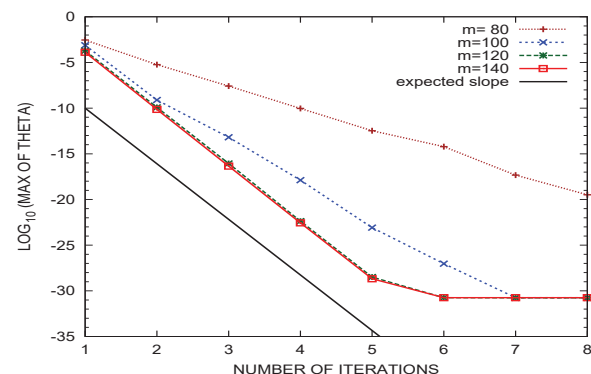


図 A-30 例題 R1-Q：反復回数と相対残差の最大値（次数 $n = 20$ ，ベクトルの数 m は 106 よりも多いことが望ましい）

A.2.2 例題 R2-Q：例題 R2 の四倍精度計算

例題 R2-Q の FEM 要素分割は $(N_1, N_2, N_3) = (20, 30, 40)$ とした。それから導かれる一般固有値問題の行列 A と B の次数は $N = 24,000$ であり、下帯幅は $w_L = 621$ である。求める固有対は固有値が区間 $[a, b] = [0, 100]$ にあるものとした。そのような固有対の数は 378 個である。また、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a, b'] = [0, 150]$ にある固有値の数は 700 個であるので、フィルタで濾過するベクトルの数 m は 700 よりも多くすることが望ましい。

例題 R2-Q：反復による相対残差の最大値の減少の様子

右に掲げた図 A・31 から図 A・34 までの 4 枚の図は、フィルタの次数 n をそれぞれ 8, 10, 15, 20 と変えた場合について、反復回数 IT を 1 から 8 まで横軸にとり、固有値が区間 $[a, b]$ に含まれる近似対の相対残差 θ の最大値を縦軸に対数でとって、フィルタで濾過した最初の乱数ベクトルの数 m を 500, 600, 700, 750 としたそれぞれの場合について色を（茶、青、緑、赤と）変えて折れ線でプロットしたグラフである。黒い線は、相対残差が反復ごとに g_S/g_P 倍になって減少する、と仮定する場合のグラフの「傾き」を示すためのものである。

フィルタの次数 n が異なる 4 枚の図のどれについても、ベクトルの数 m が 700 と 750 に対する 2 つのグラフの線はほぼ重なっており、反復 5 回目までは黒線のグラフとほぼ同じ傾きで減少していることが見てとれる。そうして、減少は相対残差の値が 10^{-31} 程度に達したところで停滞している。

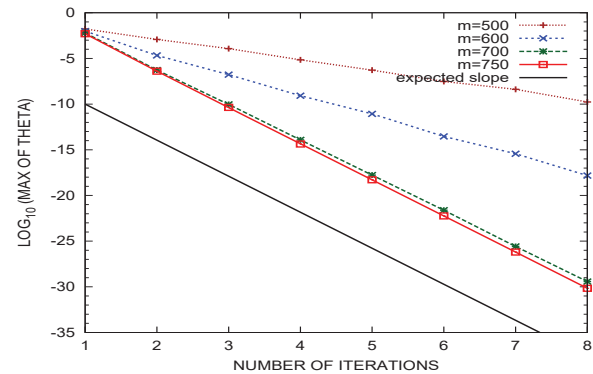


図 A・31 例題 R2-Q：反復回数と相対残差の最大値（次数 $n = 8$ ，ベクトルの数 m は 700 よりも多いことが望ましい）

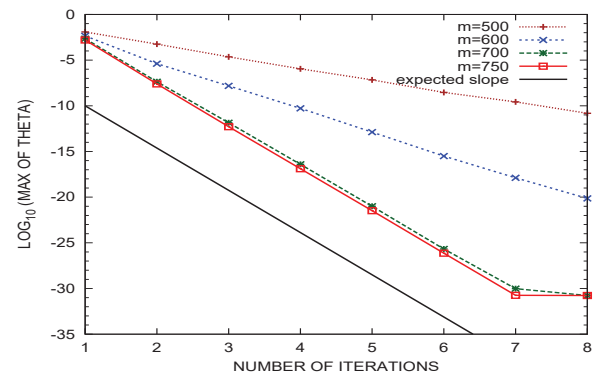


図 A・32 例題 R2-Q：反復回数と相対残差の最大値（次数 $n = 10$ ，ベクトルの数 m は 700 よりも多いことが望ましい）

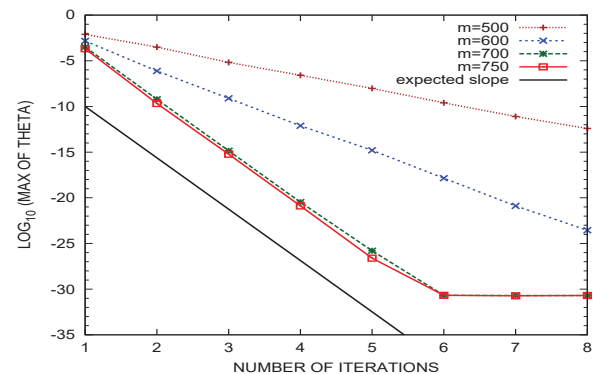


図 A・33 例題 R2-Q：反復回数と相対残差の最大値（次数 $n = 15$ ，ベクトルの数 m は 700 よりも多いことが望ましい）

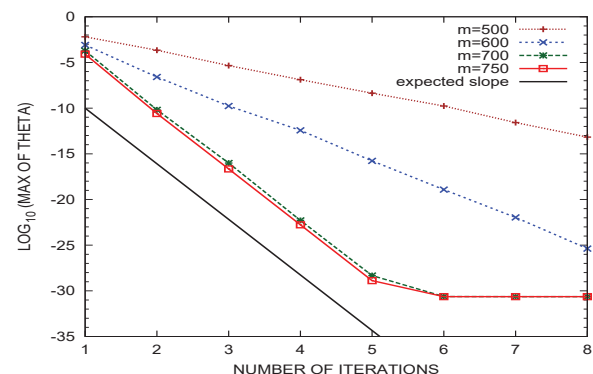


図 A・34 例題 R2-Q：反復回数と相対残差の最大値（次数 $n = 20$ ，ベクトルの数 m は 700 よりも多いことが望ましい）

A.2.3 例題 C1-Q：例題 C1 の四倍精度計算

例題 C1-Q の FEM 要素分割は $(N_1, N_2, N_3) = (20, 30, 40)$ とした。それから導かれる一般固有値問題の行列 A と B の次数は $N = 24,000$ であり、下帯幅は $w_L = 621$ である。求める固有対は固有値が区間 $[a, b] = [300, 310]$ にあるものとした。そのような固有対の数は 90 個である。また、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a', b'] = [297.5, 312.5]$ にある固有値の数は 125 個であるので、フィルタで濾過するベクトルの数 m は 125 よりも多くすることが望ましい。

例題 C1-Q：反復による相対残差の最大値の減少の様子

右に掲げた図 A・35 から図 A・38 までの 4 枚の図は、フィルタの次数 n をそれぞれ 8, 10, 15, 20 と変えた場合について、反復回数 IT を 1 から 8 まで横軸にとり、固有値が区間 $[a, b]$ に含まれる近似対の相対残差 θ の最大値を縦軸に対数でとって、フィルタで濾過した最初の乱数ベクトルの数 m が 100 から 10 刻みで 150 までについて色を変えて折れ線でそれぞれプロットしたグラフである。黒い線は、相対残差が反復ごとに g_S/g_P 倍になって減少する、と仮定する場合のグラフの「傾き」を示すためのものである。

フィルタの次数 n が異なる 4 枚の図のどれについても、ベクトルの数 m が 125 よりも多い $m = 130$, $m = 140$, $m = 150$ に対する 3 つのグラフは線がほぼ重なっており、反復 4 回目までは黒線のグラフとほぼ同じ傾きで減少していることが見てとれる。そうして、減少は相対残差の値が 10^{-32} 程度に達したところで停滞している。

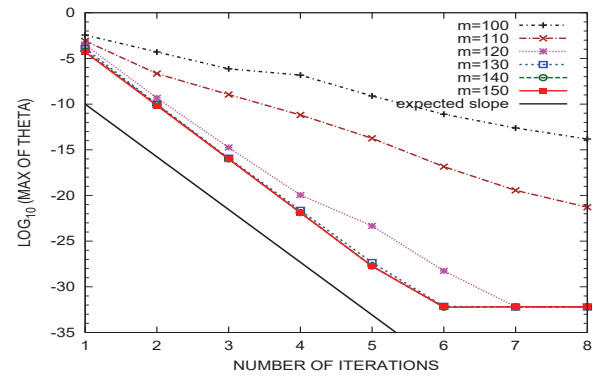


図 A・35 例題 C1-Q：反復回数と相対残差の最大値（次数 $n = 8$, ベクトルの数 m は 125 よりも多いことが望ましい）

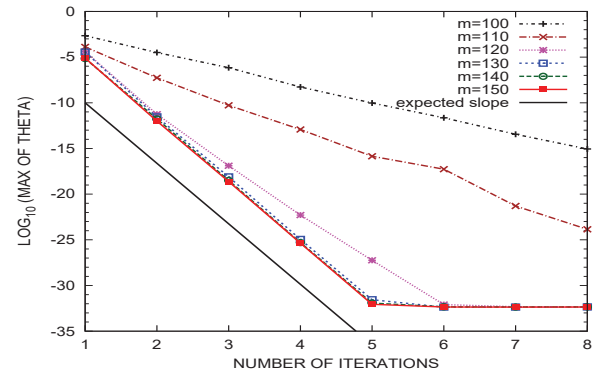


図 A・36 例題 C1-Q：反復回数と相対残差の最大値（次数 $n = 10$, ベクトルの数 m は 125 よりも多いことが望ましい）

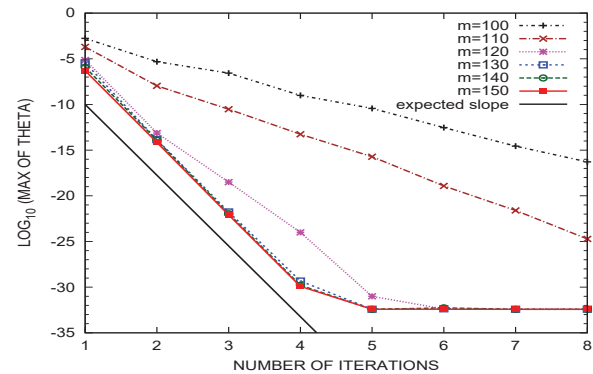


図 A・37 例題 C1-Q：反復回数と相対残差の最大値（次数 $n = 15$, ベクトルの数 m は 125 よりも多いことが望ましい）

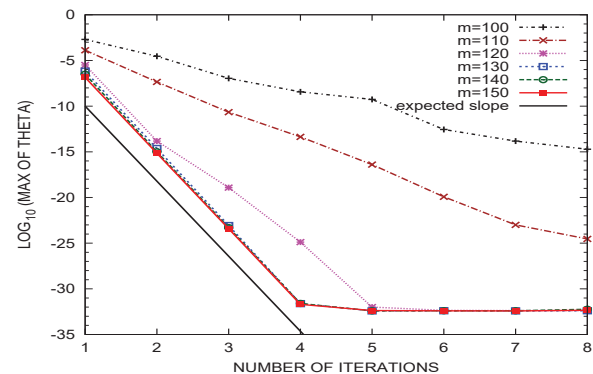


図 A・38 例題 C1-Q：反復回数と相対残差の最大値（次数 $n = 20$, ベクトルの数 m は 125 よりも多いことが望ましい）

A.2.4 例題 C2-Q：例題 C2 の四倍精度計算

例題 C2-Q の FEM 要素分割は $(N_1, N_2, N_3) = (20, 30, 40)$ とした。それから導かれる一般固有値問題の行列 A と B の次数は $N = 24,000$ であり、下帯幅は $w_L = 621$ である。求める固有対は固有値が区間 $[a, b] = [1000, 1010]$ にあるものとした。そのような固有対の数は 92 個である。また、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a', b'] = [997.5, 1012.5]$ にある固有値の数は 145 個であるので、フィルタで濾過するベクトルの数 m は 145 よりも多くすることが望ましい。

例題 C2-Q：反復による相対残差の最大値の減少の様子

右に掲げた図 A・39 から図 A・42 までの 4 枚の図は、フィルタの次数 n をそれぞれ 8, 10, 15, 20 と変えた場合について、反復回数 IT を 1 から 8 まで横軸にとり、固有値が区間 $[a, b]$ に含まれる近似対の相対残差 θ の最大値を縦軸に対数でとって、フィルタで濾過した最初の乱数ベクトルの数 m が 100 から 10 刻みで 150 までについて色を変えて折れ線でそれぞれプロットしたグラフである。黒い線は、相対残差が反復ごとに g_S/g_P 倍になって減少する、と仮定する場合のグラフの「傾き」を示すためのものである。

フィルタの次数 n が異なる 4 枚の図のどれについても、ベクトルの数 m が 145 よりも多い $m = 150$ の場合の（赤色の）グラフの線は、反復 4 回目までは黒線のグラフとほぼ同じ傾きで減少していることが見てとれる。そうして、減少は相対残差の値が 10^{-32} 程度に達したところで停滞している。

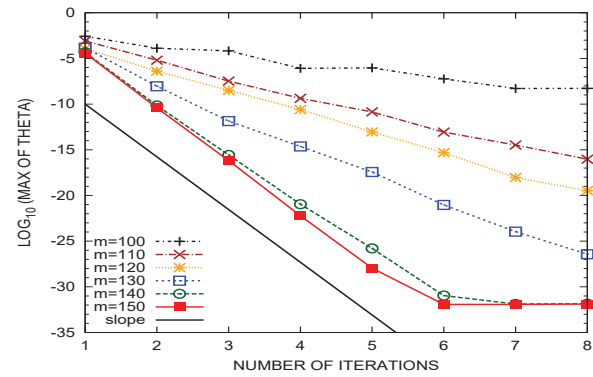


図 A・39 例題 C2-Q：反復回数と相対残差の最大値（次数 $n = 8$ ，ベクトルの数 m は 145 よりも多いことが望ましい）

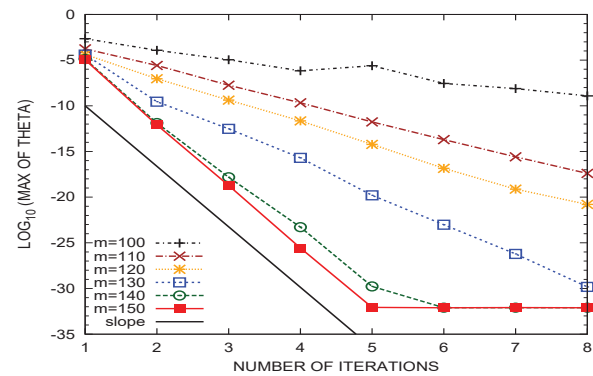


図 A・40 例題 C2-Q：反復回数と相対残差の最大値（次数 $n = 10$ ，ベクトルの数 m は 145 よりも多いことが望ましい）

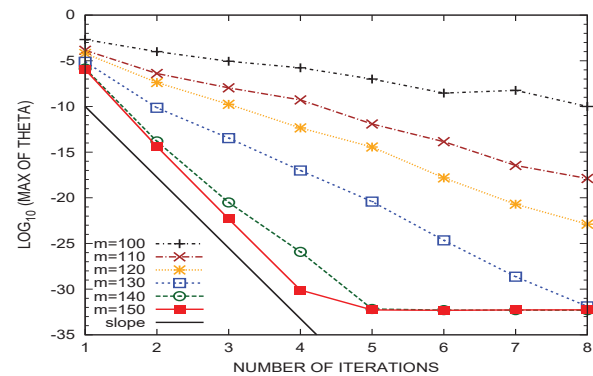


図 A・41 例題 C2-Q：反復回数と相対残差の最大値（次数 $n = 15$ ，ベクトルの数 m は 145 よりも多いことが望ましい）

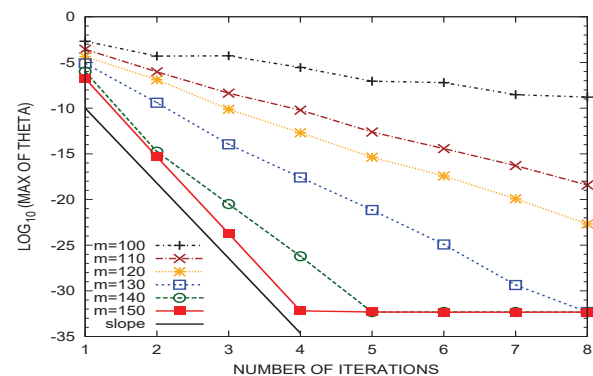


図 A・42 例題 C2-Q：反復回数と相対残差の最大値（次数 $n = 20$ ，ベクトルの数 m は 145 よりも多いことが望ましい）

A.3 B-正規直交化付きフィルタ適用の反復実験例（単精度計算）

本節の実験では、実数型の数値と演算にはすべて IEEE 754 規格の単精度浮動小数点（2 進 32bit）を用いた。その有効精度は 10 進では約 7.22 桁である。紹介する結果は、倍精度用のプログラムを簡単に単精度用に書き換えて実行することで得られたものである。単精度の計算は倍精度計算に比べて数値の有効桁数が少ないので、どうしても計算に余裕を持たせることが難しい。しかし最近は低精度の数値を用いて計算を低消費電力でおこなうことが注目されているので、今回はこのような実験もおこなってみた。

注：この節で扱った一般固有値問題の例題につけた名称は、倍精度あるいは四倍精度の場合の一般固有値問題の例題とは対応していない。たとえば、例題 C1-S は単精度計算により例題 C1 や C1-Q と同じ一般固有値問題を扱ったものではない。

単精度計算による実験に用いたフィルタについて

固有値が固有値分布の下端にある固有対を求めるためのフィルタとしては、シフトが実数である単一のレゾルベントの多項式であってその多項式が n 次のチェビシェフ多項式で表される簡易型のものを用いた。また固有値が固有値分布の中間にある固有対を求めるためのフィルタとしては、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であってその多項式が n 次のチェビシェフ多項式で表される簡易型のものを用いた。

フィルタの特性はパラメタの 3 つ組 (n, μ, g_S) で指定した。 μ は常に 1.5 に設定した。実験に用いたフィルタは、求めようとする固有対の固有値の区間が下端である場合と中間である場合のそれぞれについて、次数を $n = 4$ にして g_S の値を 10^{-3} , 10^{-4} , 10^{-5} の 3 通りに変えた場合と、 g_S の値を 10^{-5} にして次数 n を 6, 8, 10 の 3 通りに変えた場合の合計 6 通りとした。パラメタをそれら 6 通りの組み合わせに選んだ場合の g_P の値と g_S/g_P の値を下端固有対用および中間固有対用のフィルタについて、表 A.3 と表 A.4 に示す。また、それらのフィルタの伝達関数 $g(t)$ のグラフを下端固有対用で次数が $n = 4$ の場合のものを図 A.43 に、下端固有対用で $g_S = 10^{-5}$ の場合のものを図 A.44 に、中間固有対用で次数が $n = 4$ の場合のものを図 A.45 に、中間固有対用で $g_S = 10^{-5}$ の場合のものを図 A.46 にそれぞれ示す。

表 A.3 （単精度） 下端固有対用フィルタ, $\mu = 1.5$

n	g_S	g_P	g_S/g_P
4	10^{-3}	1.93×10^{-2}	5.19×10^{-2}
4	10^{-4}	3.56×10^{-3}	2.81×10^{-2}
4	10^{-5}	5.33×10^{-4}	1.88×10^{-2}
6	10^{-5}	1.53×10^{-3}	6.54×10^{-3}
8	10^{-5}	2.55×10^{-3}	3.92×10^{-3}
10	10^{-5}	3.34×10^{-3}	2.99×10^{-3}

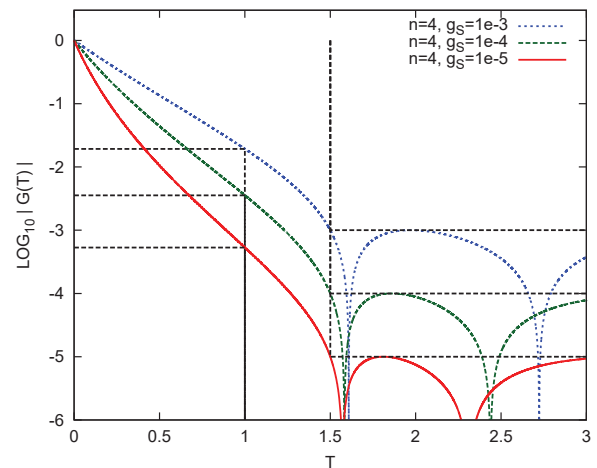


図 A.43 （単精度） 下端固有対用フィルタ, 伝達関数の大きさ $|g(t)|$ ($\mu = 1.5$, $n = 4$ のもの)

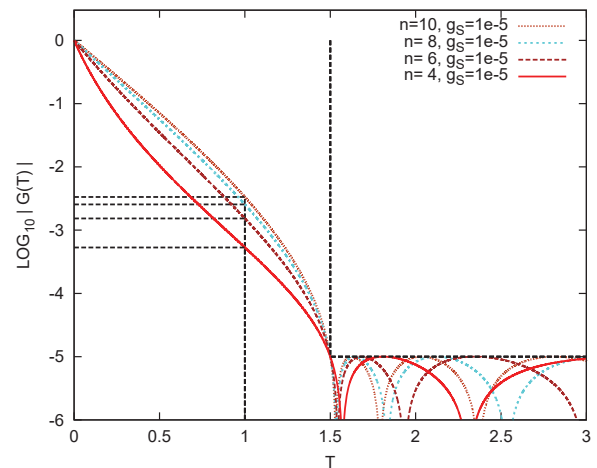


図 A.44 （単精度） 下端固有対用フィルタ, 伝達関数の大きさ $|g(t)|$ ($\mu = 1.5$, $g_S = 10^{-5}$ のもの)

表 A.4 (単精度) 中間固有対用フィルタ, $\mu = 1.5$

n	g_S	g_P	g_S/g_P
4	10^{-3}	7.16×10^{-2}	1.40×10^{-2}
4	10^{-4}	1.88×10^{-2}	5.33×10^{-3}
4	10^{-5}	3.69×10^{-3}	2.71×10^{-3}
6	10^{-5}	1.25×10^{-2}	8.01×10^{-4}
8	10^{-5}	2.11×10^{-2}	4.73×10^{-4}
10	10^{-5}	2.74×10^{-2}	3.65×10^{-4}

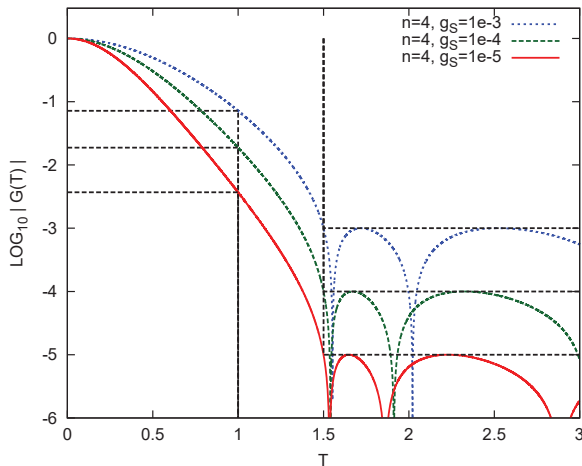


図 A.45 (単精度) 中間固有対用フィルタ, 伝達関数の大きさ $|g(t)|$, 右半分 ($\mu = 1.5$, $n = 4$ のもの)

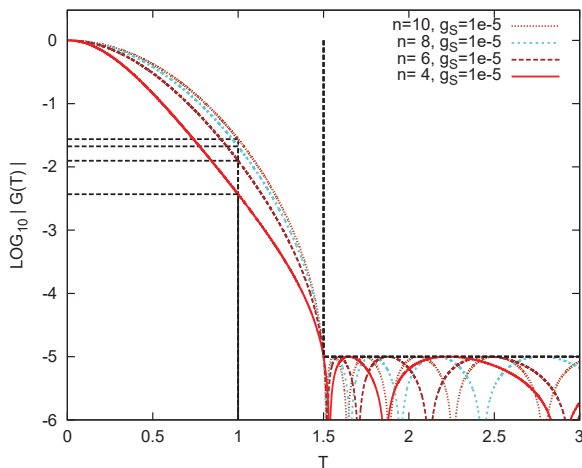


図 A.46 (単精度) 中間固有対用フィルタ, 伝達関数の大きさ $|g(t)|$, 右半分 ($\mu = 1.5$, $g_S = 10^{-5}$ のもの)

単精度計算による実験について

以下の例題 R1-S, 例題 R2-S, 例題 C1-S, 例題 C2-S, 例題 C3-S, 例題 C4-S の実験では直交化付きフィルタの反復回数 IT は高々 6 とした. まず 1 回目の適用では一様乱数からベクトルの組を作り, それを B -正規直交化して得られたベクトルの組に対してフィルタを適用している. しかし単精度計算では倍精度計算との相違点として, 「求めたい固有ベクトル」が一様乱数から作ったベクトルの組に含まれている割合が丸め誤差のレベル (マシンイプシロン) よりも十分に大きくはないことがフィルタを用いた近似対の計算で精度上の困難をもたらす原因として挙げられる. そのほか, 単精度計算では実現できるフィルタの特性には制限がある (たとえば倍精度計算の場合のように, 阻止域での伝達率の上限 g_S を 10^{-12} にすることは無理である). そのため, ベクトルの組に含まれる「不要な固有ベクトル」の割合に対してフィルタ適用ごとに乗じられる低減化率の上限 g_S/g_P を十分に小さくできなければ, フィルタ適用 1 回の結果から近似対を求めたとしてもその相対残差は小さくなかったり, あるいは近似対の数すら正しくないことが起こり得る. そのような場合でも, ベクトルの組に B -直交化付きフィルタを反復して適用することにより, ベクトルの組に含まれる「不要な固有ベクトル」の割合が減らせるのであれば, 得られる近似対は改良されることが考えられる. (なお今後計算を単精度で行う場合に取り組むべきこととして, 計算過程における数値丸めの影響をできるだけ減らすように計算法を見直すことが考えられる. たとえばベクトルの組に対する B -正規直交化や, フィルタ適用後のベクトルの組から不変部分空間の近似基底をとりだす計算の過程などである.)

A.3.1 例題 R1-S：下端固有対を解いた例 1 (単精度計算)

この例題 R1-S で用いた FEM の要素分割は $(N_1, N_2, N_3) = (50, 60, 70)$ である。それから導かれる一般固有値問題の係数行列 A と B の次数は $N = 210000$ であり、下帯幅は $w_L = 3051$ である。求める固有対は固有値が下端の区間 $[a, b] = [0, 30]$ にあるものとした。そのような固有対の数は 54 個である。そうして、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a, b'] = [0, 45]$ にある固有値の数は 108 個であり、フィルタを適用するベクトルの数 m はそれよりも多くすることが望ましいので、この例では $m = 120$ とした。(ベクトルの数 m を増すほど、近似対の品質は向上する。特にフィルタの適用回数 IT が 1 の場合にはそうなる。)

フィルタの次数が小さい $n = 4$ の場合 (表 A・5 から表 A・7 まで、図 A・47 から図 A・49 まで) では、得られた近似対の数はフィルタの適用回数が IT=1 の場合には 54 個にはならず正しくない (数が正しくない場合には下線を付けている)。IT=2 の場合は、 g_S が 10^{-3} のときは数は正しくないが、 $g_S = 10^{-4}$ と $g_S = 10^{-5}$ のときには正しい 54 個になっている。そうして IT=3 と IT=4 の場合には近似対の数はみな正しい。ところが、 $g_S = 10^{-5}$ のときの結果 (表 A・6、図 A・48) は $g_S = 10^{-4}$ のときの結果 (表 A・7、図 A・49) よりもむしろ精度が悪くなったことが、 $\Theta_{\max}$ の値からあるいはグラフで IT=4 の場合を比べるとわかる。これは現状の計算処理内での切断操作が影響しているのではないと思われる。

こんどは g_S を 10^{-5} にして、フィルタの次数 n を 4 から 10 まで 2 ずつ増やした場合 (表 A・7 から表 A・10 まで、図 A・49 から図 A・52 まで) では、近似対の数はフィルタの適用回数 IT=1 のときには正しくないが、IT が 2 以上ではどれも正しいく 54 個になっている。フィルタの次数 n が 6 以上である場合には、固有値が区間の右端付近にある近似対も相対残差が反復とともに順調に減少したことがわかる。

表 A・5 例題 R1-S：次数 $n = 4$, $g_S = 10^{-3}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>0</u>	---
2	<u>48</u>	9.9E-01
3	54	5.8E-02
4	54	2.2E-03
5	54	4.1E-04

表 A・6 例題 R1-S：次数 $n = 4$, $g_S = 10^{-4}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>0</u>	---
2	54	6.9E-01
3	54	1.3E-02
4	54	4.2E-04
5	54	4.2E-04

表 A・7 例題 R1-S：次数 $n = 4$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>4</u>	5.2E-01
2	54	2.8E-01
3	54	9.2E-03
4	54	2.4E-03
5	54	1.6E-03

表 A・8 例題 R1-S：次数 $n = 6$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>11</u>	2.3E-01
2	54	1.7E-02
3	54	8.3E-04
4	54	4.3E-04

表 A・9 例題 R1-S：次数 $n = 8$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>15</u>	3.8E-01
2	54	1.6E-02
3	54	4.2E-04
4	54	4.2E-04

表 A・10 例題 R1-S：次数 $n = 10$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>18</u>	4.2E-01
2	54	1.1E-02
3	54	4.1E-04
4	54	4.1E-04

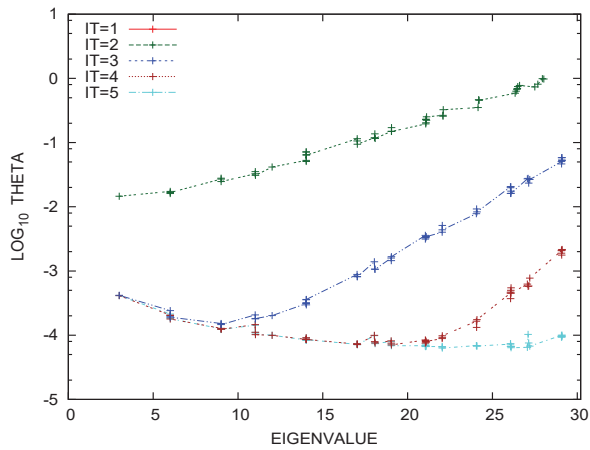


図 A-47 例題 R1-S：固有値と相対残差 ($n = 4$, $g_s = 10^{-3}$)

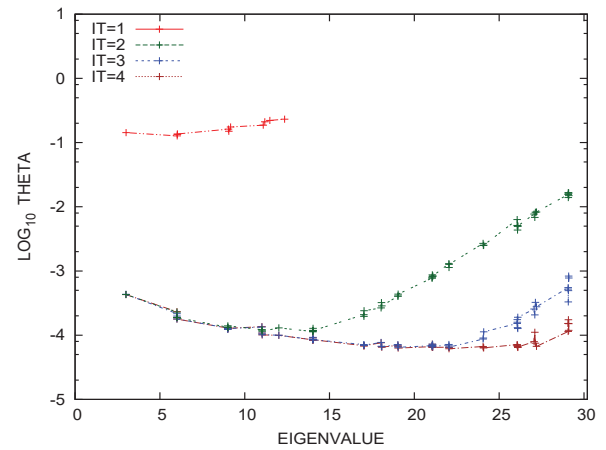


図 A-50 例題 R1-S：固有値と相対残差 ($n = 6$, $g_s = 10^{-5}$)

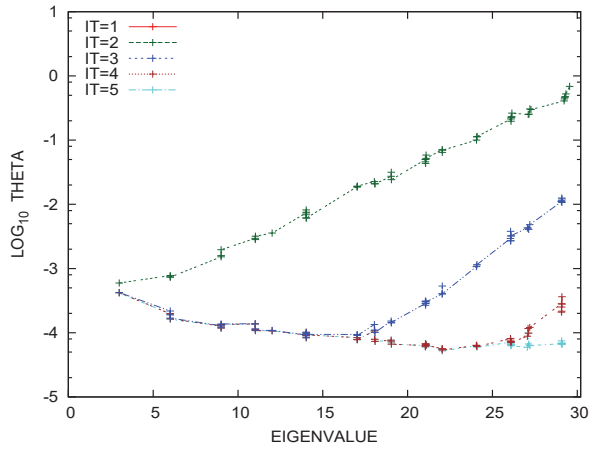


図 A-48 例題 R1-S：固有値と相対残差 ($n = 4$, $g_s = 10^{-4}$)

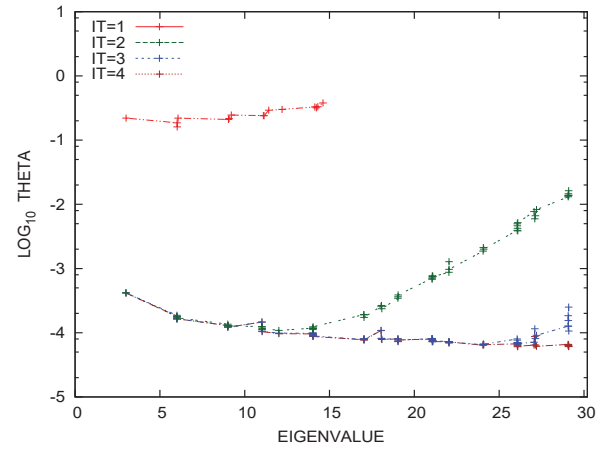


図 A-51 例題 R1-S：固有値と相対残差 ($n = 8$, $g_s = 10^{-5}$)

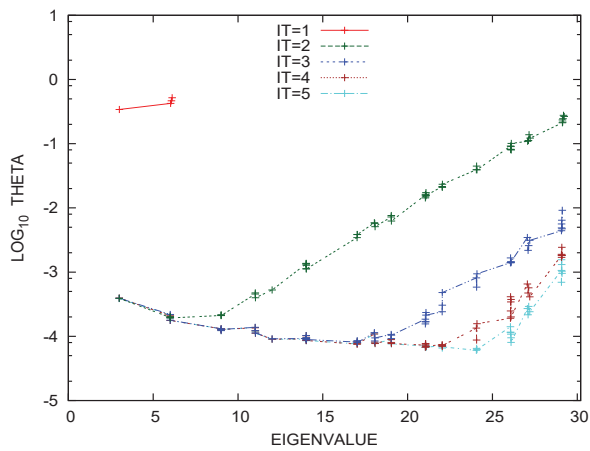


図 A-49 例題 R1-S：固有値と相対残差 ($n = 4$, $g_s = 10^{-5}$)

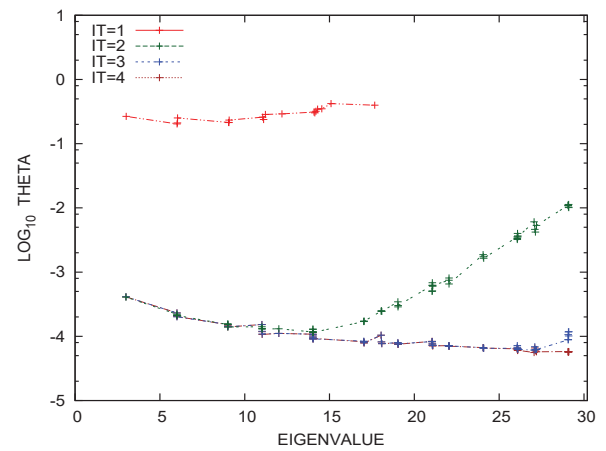


図 A-52 例題 R1-S：固有値と相対残差 ($n = 10$, $g_s = 10^{-5}$)

A.3.2 例題 R2-S：下端固有対を解いた例 2 (単精度計算)

この例題 R2-S で用いた FEM の要素分割は $(N_1, N_2, N_3) = (50, 60, 70)$ である。それから導かれる一般固有値問題の係数行列 A と B の次数は $N = 210000$ であり、下帯幅は $w_L = 3051$ である。求める固有対は固有値が下端の区間 $[a, b] = [0, 100]$ にあるものとした。そのような固有対の数は 402 個である。そうして、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a, b'] = [0, 150]$ にある固有値の数は 764 個であり、フィルタを適用するベクトルの数 m はそれよりも多くすることが望ましいので、この例では $m = 800$ とした。

フィルタの次数が小さい $n = 4$ の場合 (表 A.11 から表 A.13 まで、図 A.53 から図 A.55 まで) には、得られた近似対の数はフィルタの適用回数が IT=1 の場合には 402 個とは違っており正しくない。IT=2 の場合は、 $g_S = 10^{-3}$ のときには正しくないが、 $g_S = 10^{-4}$ と $g_S = 10^{-5}$ のときには正しい数 402 個になっており、IT が 3 以上の場合には得られた近似対の数はみな正しい。ところが、 $g_S = 10^{-5}$ のとき (図 A.54) の結果は $g_S = 10^{-4}$ のとき (図 A.55) の結果に比べてむしろ精度が悪くなっていることが、グラフにおいて区間の右端付近の固有値を持つ近似対の相対残差を IT=4 や IT=5 の場合について比較すればわかる。これは現状の計算処理内での切断操作が影響しているのではないと思われる。

こんどは g_S を 10^{-5} にして、フィルタの次数 n を 4 から 10 まで 2 ずつ増やした場合 (表 A.13 から表 A.16 まで、図 A.55 から図 A.58 まで) では、得られた近似対の数はフィルタの適用回数が IT=1 のときには正しくないが、IT が 2 以上ではどれも正しく 402 個になっている。フィルタの次数 n が 6 以上である場合には、反復とともに固有値が区間の右端付近にある収束の遅い近似対も相対残差が順調に減少したことがわかる。

表 A.11 例題 R2-S：次数 $n = 4$, $g_S = 10^{-3}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>5</u>	2.3E+00
2	<u>394</u>	2.3E-01
3	402	1.6E-02
4	402	9.6E-04
5	402	3.8E-04
6	402	3.8E-04

表 A.12 例題 R2-S：次数 $n = 4$, $g_S = 10^{-4}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>82</u>	5.0E-01
2	402	6.8E-02
3	402	2.3E-03
4	402	4.1E-04
5	402	3.9E-04
6	402	4.1E-04

表 A.13 例題 R2-S：次数 $n = 4$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>139</u>	1.6E-01
2	402	2.7E-02
3	402	2.7E-03
4	402	8.7E-04
5	402	4.5E-04
6	402	4.5E-04

表 A.14 例題 R2-S：次数 $n = 6$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>222</u>	2.3E-01
2	402	1.1E-02
3	402	3.8E-04
4	402	3.8E-04

表 A.15 例題 R2-S：次数 $n = 8$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>264</u>	1.8E-01
2	402	3.6E-03
3	402	4.0E-04
4	402	3.9E-04

表 A.16 例題 R2-S：次数 $n = 10$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>287</u>	1.7E-01
2	402	2.2E-03
3	402	3.9E-04
4	402	4.0E-04

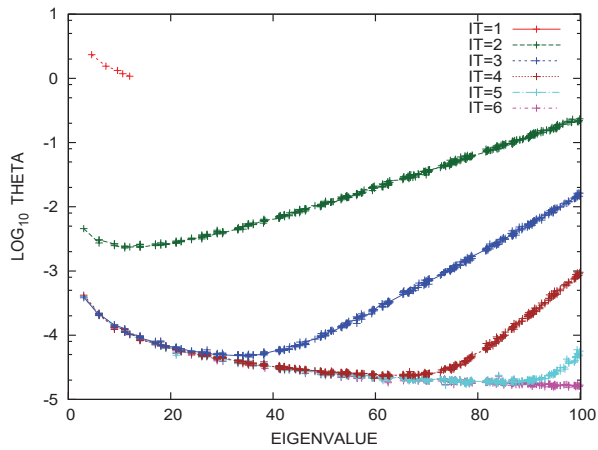


図 A.53 例題 R2-S：固有値と相対残差 ($n = 4$, $g_S = 10^{-3}$)

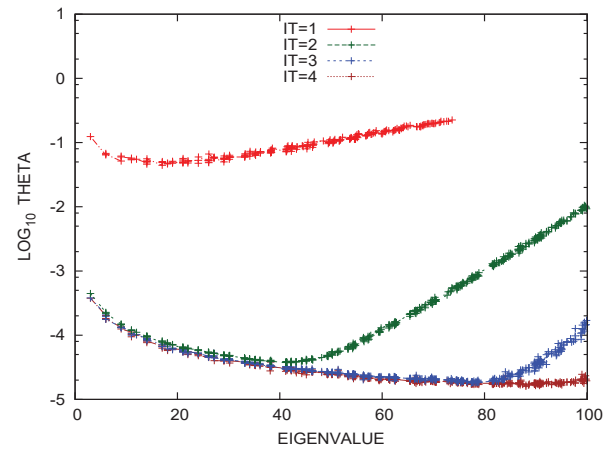


図 A.56 例題 R2-S：固有値と相対残差 ($n = 6$, $g_S = 10^{-5}$)

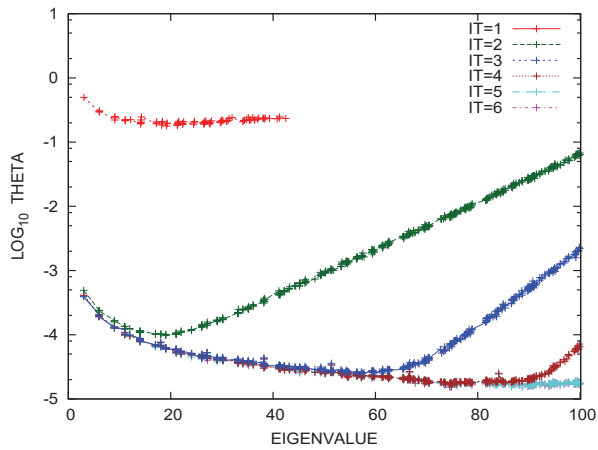


図 A.54 例題 R2-S：固有値と相対残差 ($n = 4$, $g_S = 10^{-4}$)

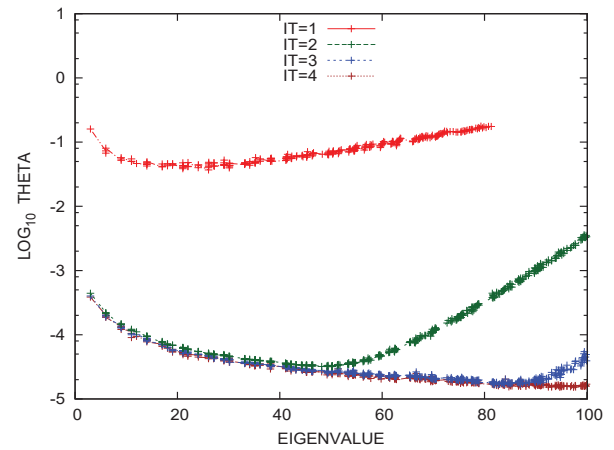


図 A.57 例題 R2-S：固有値と相対残差 ($n = 8$, $g_S = 10^{-5}$)

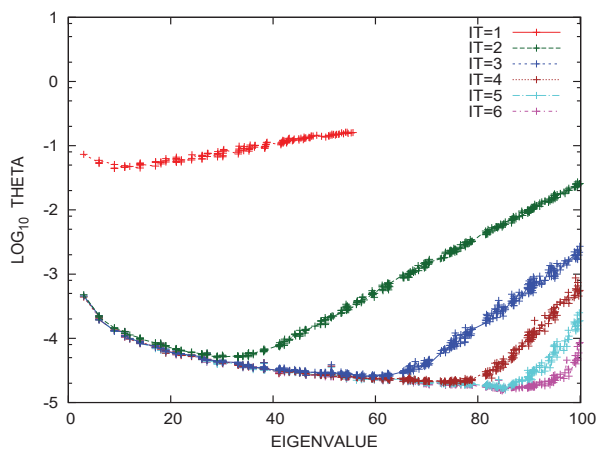


図 A.55 例題 R2-S：固有値と相対残差 ($n = 4$, $g_S = 10^{-5}$)

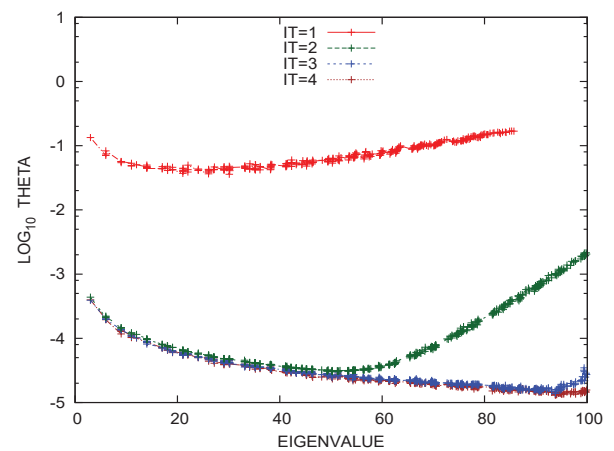


図 A.58 例題 R2-S：固有値と相対残差 ($n = 10$, $g_S = 10^{-5}$)

A.3.3 例題 C1-S：中間固有対を解いた例 1 (単精度計算)

この例題 C1-S で用いた FEM の要素分割は $(N_1, N_2, N_3) = (50, 60, 70)$ である。それから導かれる一般固有値問題の係数行列 A と B の次数は $N = 210000$ であり、下帯幅は $w_L = 3051$ である。求める固有対は固有値が区間 $[a, b] = [0, 30]$ にあるものとした（ただしここでは下端の区間を中間の区間のようにして扱った）。そのような固有対の数は 54 個である。そうして、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a', b'] = [-7.5, 37.5]$ にある固有値の数は 78 個であり、フィルタを適用するベクトルの数 m はそれよりも多くすることが望ましいので、この例では $m = 100$ とした。

フィルタの次数が小さい $n = 4$ の場合（表 A.17 から表 A.19 まで、図 A.59 から図 A.61 まで）には、得られた近似対の数はフィルタの適用回数 $IT=1$ の場合にはどれも正しい数 54 個とは異なっていて正しくないが、 IT が 2 以上の場合にはみな正しい。反復回数 IT を 2, 3, 4 と増すにつれて近似対の相対残差が減少していく様子がこれら 3 枚のグラフからわかる。またこれら 3 つの表からは、相対残差の最大値 $\Theta_{\max}$ が反復により減少した様子もわかる。

こんどは g_S を 10^{-5} にして、フィルタの次数 n を 4 から 10 まで 2 ずつ増やした場合（表 A.19 から表 A.22 まで、図 A.61 から図 A.64 まで）をみると、近似対の数はフィルタの適用回数が $IT=1$ のときには正しい数 54 個からは異なっているが、 IT が 2 以上のときにはどれも正しい数 54 個になっている。グラフからフィルタの次数 n が 4, 6, 8, 10 のどの場合についても、近似対の相対残差は反復回数 IT とともに順調に減少したことがわかる（これら 4 枚の図では、 $IT=3$ のグラフ（青線）と $IT=4$ のグラフ（茶線）の線がほとんど重なっており、改良が反復 3 回でほぼ終了していることがわかる）。

表 A.17 例題 C1-S：次数 $n = 4$, $g_S = 10^{-3}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>0</u>	---
2	54	8.0E-01
3	54	5.3E-03
4	54	6.4E-04

表 A.18 例題 C1-S：次数 $n = 4$, $g_S = 10^{-4}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>16</u>	5.5E-01
2	54	7.2E-02
3	54	6.0E-04
4	54	5.6E-04

表 A.19 例題 C1-S：次数 $n = 4$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>33</u>	1.2E+00
2	54	9.9E-03
3	54	5.1E-04
4	54	5.1E-04

表 A.20 例題 C1-S：次数 $n = 6$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>40</u>	8.3E-01
2	54	1.3E-03
3	54	6.3E-04
4	54	6.3E-04

表 A.21 例題 C1-S：次数 $n = 8$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>41</u>	7.0E-01
2	54	1.1E-03
3	54	7.4E-04
4	54	7.5E-04

表 A.22 例題 C1-S：次数 $n = 10$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>45</u>	8.1E-01
2	54	1.6E-03
3	54	9.8E-04
4	54	9.8E-04

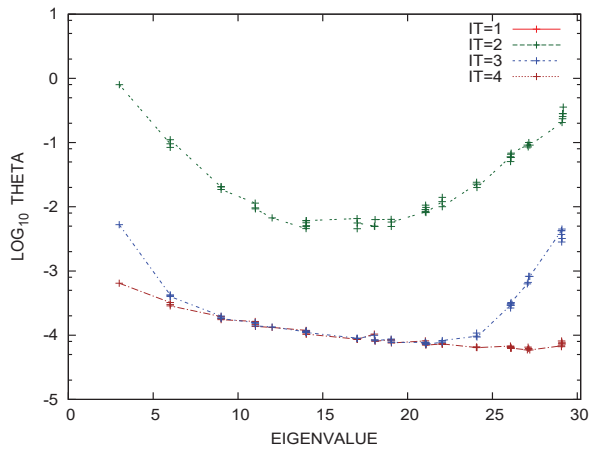


図 A.59 例題 C1-S : 固有値と相対残差 ($n = 4$, $g_s = 10^{-3}$)

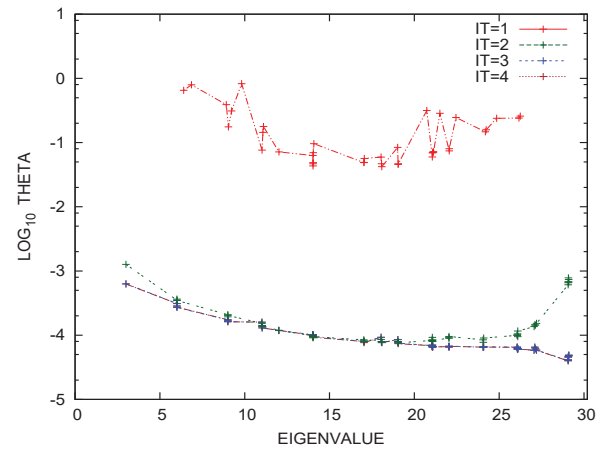


図 A.62 例題 C1-S : 固有値と相対残差 ($n = 6$, $g_s = 10^{-5}$)

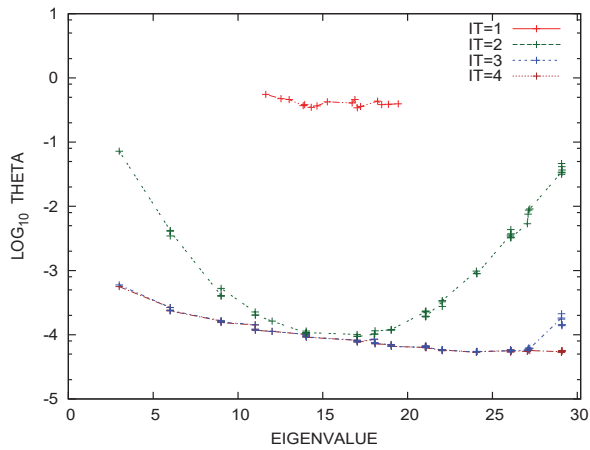


図 A.60 例題 C1-S : 固有値と相対残差 ($n = 4$, $g_s = 10^{-4}$)

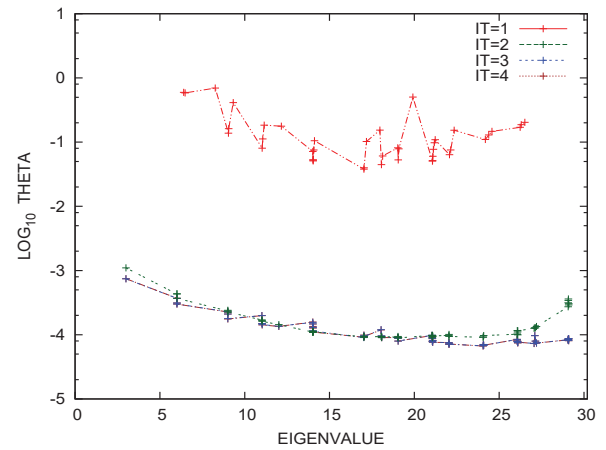


図 A.63 例題 C1-S : 固有値と相対残差 ($n = 8$, $g_s = 10^{-5}$)

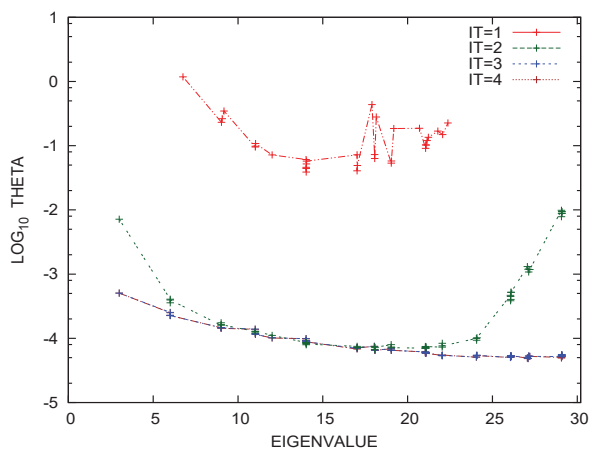


図 A.61 例題 C1-S : 固有値と相対残差 ($n = 4$, $g_s = 10^{-5}$)

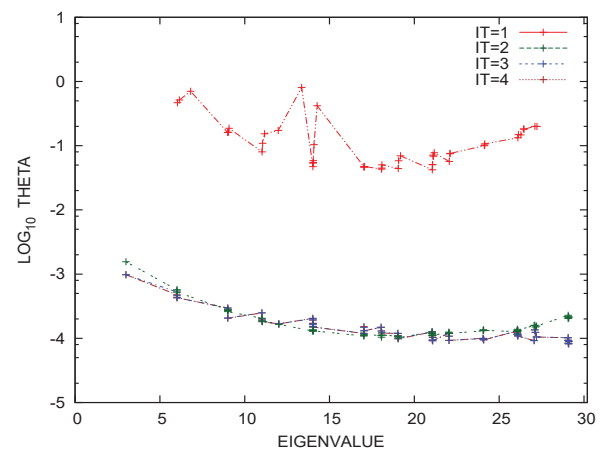


図 A.64 例題 C1-S : 固有値と相対残差 ($n = 10$, $g_s = 10^{-5}$)

A.3.4 例題 C2-S：中間固有対を解いた例 2 (単精度計算)

この例題 C2-S で用いた FEM の要素分割は $(N_1, N_2, N_3) = (50, 60, 70)$ である。それから導かれる一般固有値問題の係数行列 A と B の次数は $N = 210000$ であり、下帯幅は $w_L = 3051$ である。求める固有対は固有値が中間の区間 $[a, b] = [300, 310]$ にあるものとした。そのような固有対の数は 114 個である。そうして、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a', b'] = [297.5, 312.5]$ にある固有値の数は 172 個であり、フィルタを適用するベクトルの数 m はそれよりも多くすることが望ましいので、この例では $m = 200$ とした。

フィルタの次数が小さい $n = 4$ の場合 (表 A-23 から表 A-25 まで、図 A-65 から図 A-67 まで) には、得られた近似対の数はフィルタの適用回数 $IT=1$ の場合には正しい数 114 個とは異なっていて正しくない。そうして IT が 2 の場合は $g_S = 10^{-3}$ のときには正しくないが、 $g_S = 10^{-4}$ と $g_S = 10^{-5}$ のときは正しく、さらに $IT=3$ と $IT=4$ の場合にはすべてのときに正しい数になっている。これら 3 枚のグラフからは、得られた近似対の相対残差が反復回数 IT を 2, 3, 4 と増すにつれて減少した様子がわかる。またこれら 3 つの表から反復により相対残差の最大値 $\Theta_{\max}$ が減少した様子もわかる。

こんどは g_S を 10^{-5} にして、フィルタの次数 n を 4 から 10 まで 2 ずつ増やした場合 (表 A-25 から表 A-28 まで、図 A-67 から図 A-70 まで) をみると、得られた近似対の数はフィルタの適用回数が $IT=1$ のときには正しい数 114 個とは違うが、 IT が 2 以上のときにはどれも正しい数である。そうしてフィルタの次数 n が 4, 6, 8, 10 のどの場合についても、近似対の相対残差は反復回数 IT とともに順調に減少したことがグラフからわかる (これら 4 枚の図では、 $IT=3$ のグラフ (青線) と $IT=4$ のグラフ (茶線) の線がほとんど重なっているの、改良は反復 3 回でほぼ終了していることがわかる)。

表 A-23 例題 C2-S：次数 $n = 4$, $g_S = 10^{-3}$

IT	得られた近似対の数	$\Theta_{\max}$
1	0	---
2	121	8.3E-02
3	114	7.4E-04
4	114	1.3E-05

表 A-24 例題 C2-S：次数 $n = 4$, $g_S = 10^{-4}$

IT	得られた近似対の数	$\Theta_{\max}$
1	45	3.0E-02
2	114	7.0E-03
3	114	4.0E-05
4	114	7.8E-06

表 A-25 例題 C2-S：次数 $n = 4$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	80	2.3E-02
2	114	1.7E-03
3	114	9.5E-06
4	114	8.4E-06

表 A-26 例題 C2-S：次数 $n = 6$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	92	2.2E-02
2	114	1.8E-04
3	114	7.8E-06
4	114	7.8E-06

表 A-27 例題 C2-S：次数 $n = 8$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	97	2.1E-02
2	114	6.1E-05
3	114	9.3E-06
4	114	9.3E-06

表 A-28 例題 C2-S：次数 $n = 10$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	100	2.2E-02
2	114	3.8E-05
3	114	1.1E-05
4	114	1.1E-05

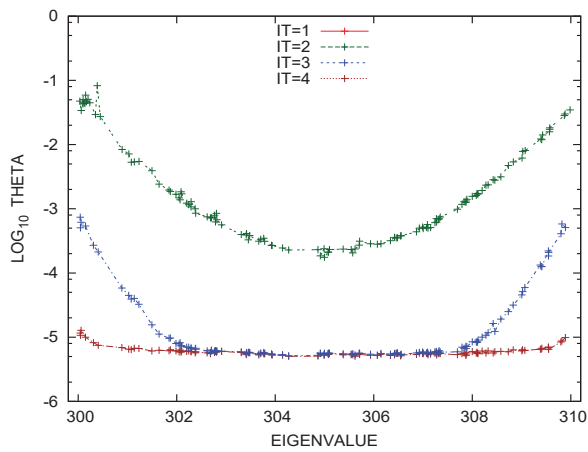


図 A.65 例題 C2-S：固有値と相対残差 ($n = 4$, $g_S = 10^{-3}$)

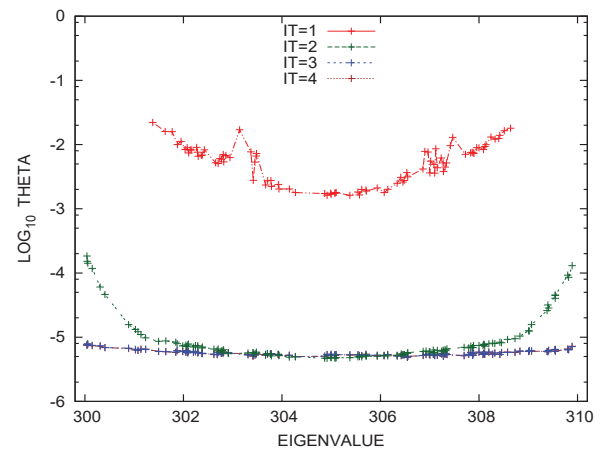


図 A.68 例題 C2-S：固有値と相対残差 ($n = 6$, $g_S = 10^{-5}$)

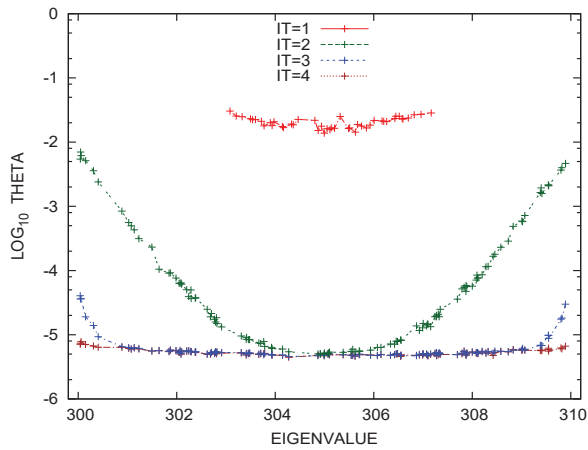


図 A.66 例題 C2-S：固有値と相対残差 ($n = 4$, $g_S = 10^{-4}$)

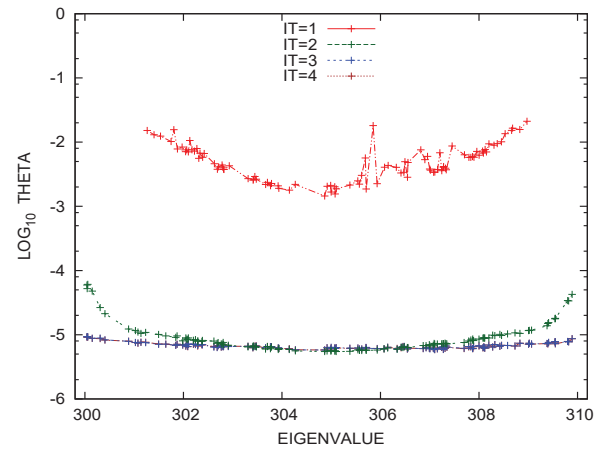


図 A.69 例題 C2-S：固有値と相対残差 ($n = 8$, $g_S = 10^{-5}$)

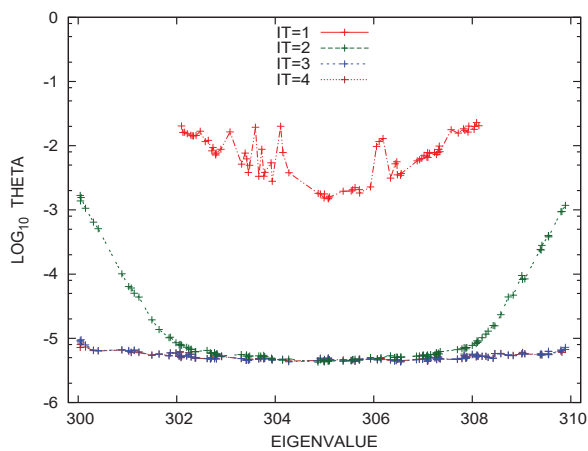


図 A.67 例題 C2-S：固有値と相対残差 ($n = 4$, $g_S = 10^{-5}$)

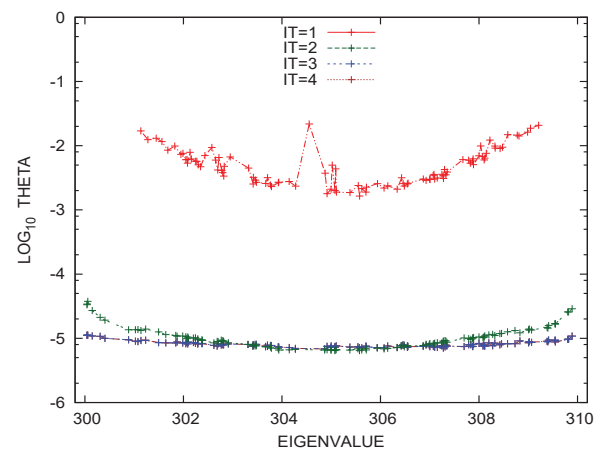


図 A.70 例題 C2-S：固有値と相対残差 ($n = 10$, $g_S = 10^{-5}$)

A.3.5 例題 C3-S：中間固有対を解いた例 3（単精度計算）

この例題 C3-S で用いた FEM の要素分割は $(N_1, N_2, N_3) = (50, 60, 70)$ である。それから導かれる一般固有値問題の係数行列 A と B の次数は $N = 210000$ であり、下帯幅は $w_L = 3051$ である。求める固有対は固有値が中間の区間 $[a, b] = [1000, 1010]$ にあるものとした。そのような固有対の数は 180 個である。そうして、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a', b'] = [997.5, 1012.5]$ にある固有値の数は 263 個であり、フィルタを適用するベクトルの数はそれよりも多くすることが望ましいので、この例では $m = 300$ とした。

フィルタの次数が小さい $n = 4$ の場合（表 A-29 から表 A-31 まで、図 A-71 から図 A-73 まで）には、得られた近似対の数はフィルタの適用回数 IT=1 の場合には正しい数 180 個とは違っている。そうして IT が 2 の場合は $g_S = 10^{-3}$ と $g_S = 10^{-4}$ のときには 1 つ多くて正しくないが、 $g_S = 10^{-5}$ のときには正しい数 180 個になっている。さらに IT=3 と IT=4 の場合にはすべて正しい数になっている。これら 3 枚のグラフからは、近似対の相対残差が反復回数 IT を 2, 3, 4 と増すにつれて減少した様子がわかる。またこれら 3 つの表から相対残差の最大値 $\Theta_{\max}$ の減少の様子もわかる。

こんどは g_S を 10^{-5} にして、フィルタの次数 n を 4 から 10 まで 2 ずつ増やした場合（表 A-31 から表 A-34 まで、図 A-73 から図 A-76 まで）をみると、得られた近似対の数はフィルタの適用回数が IT=1 のときには正しい数 180 個とは異なるが、IT が 2 以上のときにはどれも正しい数と一致している。そうして 4 枚の図のグラフからは、フィルタの次数 n が 4, 6, 8, 10 のそれぞれの場合について、近似対の相対残差が反復回数 IT とともに順調に減少したことがわかる（これら 4 枚の図はどれも、IT=3 のグラフ（青線）と IT=4 のグラフ（茶線）の線がほとんど重なっているので、改良は反復 3 回でほぼ終了していることがわかる）。

表 A-29 例題 C3-S：次数 $n = 4$, $g_S = 10^{-3}$

IT	得られた近似対の数	$\Theta_{\max}$
1	0	---
2	181	1.2E-02
3	180	1.7E-04
4	180	5.8E-06

表 A-30 例題 C3-S：次数 $n = 4$, $g_S = 10^{-4}$

IT	得られた近似対の数	$\Theta_{\max}$
1	85	8.2E-03
2	181	2.0E-03
3	180	1.1E-05
4	180	7.1E-06

表 A-31 例題 C3-S：次数 $n = 4$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	122	8.4E-03
2	180	4.5E-04
3	180	1.1E-05
4	180	1.2E-05

表 A-32 例題 C3-S：次数 $n = 6$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	146	6.4E-03
2	180	4.6E-05
3	180	5.8E-06
4	180	5.4E-06

表 A-33 例題 C3-S：次数 $n = 8$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	154	6.4E-03
2	180	1.8E-05
3	180	5.3E-06
4	180	4.9E-06

表 A-34 例題 C3-S：次数 $n = 10$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	158	6.1E-03
2	180	1.2E-05
3	180	5.3E-06
4	180	5.3E-06

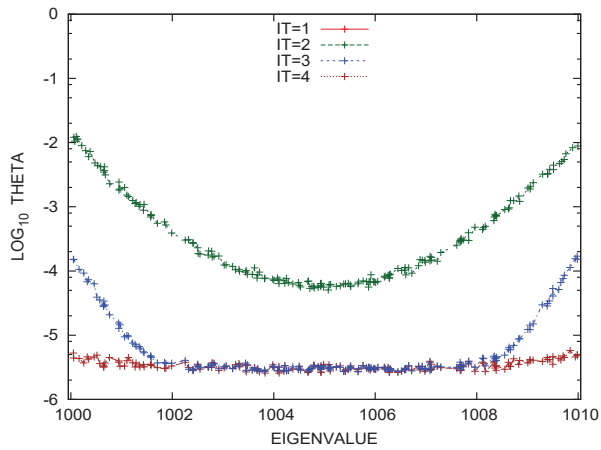


図 A.71 例題 C3-S : 固有値と相対残差 ($n = 4$, $g_s = 10^{-3}$)

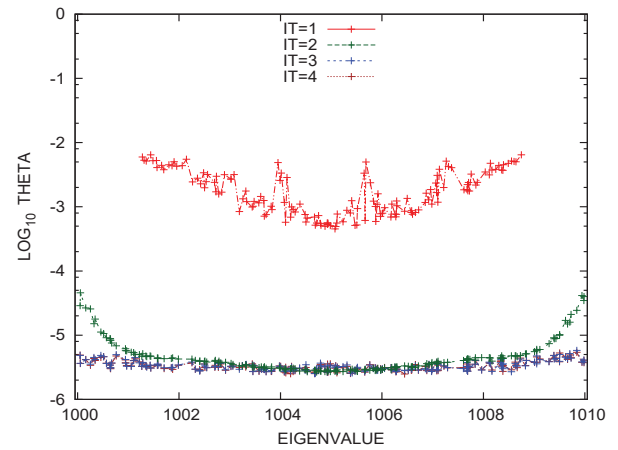


図 A.74 例題 C3-S : 固有値と相対残差 ($n = 6$, $g_s = 10^{-5}$)

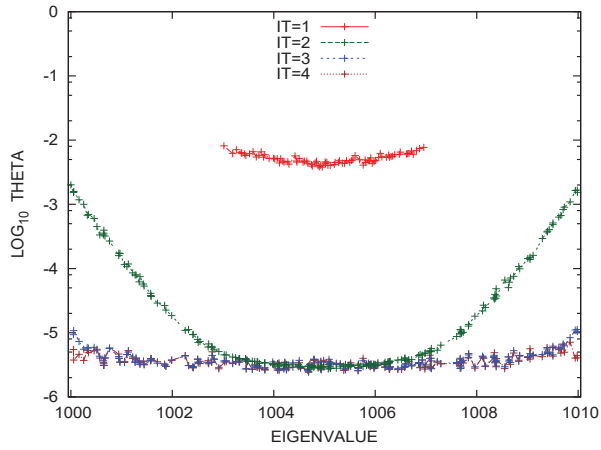


図 A.72 例題 C3-S : 固有値と相対残差 ($n = 4$, $g_s = 10^{-4}$)

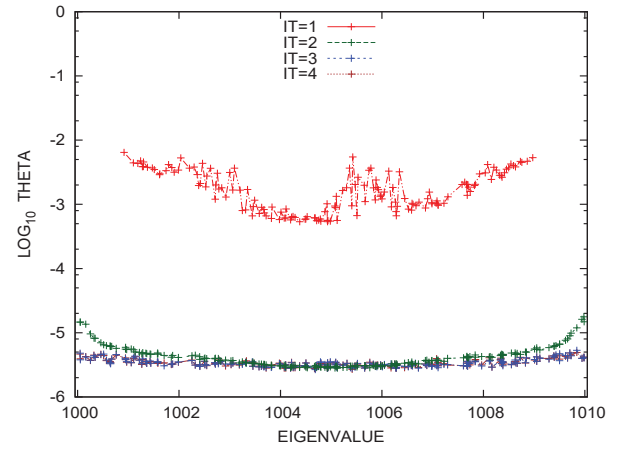


図 A.75 例題 C3-S : 固有値と相対残差 ($n = 8$, $g_s = 10^{-5}$)

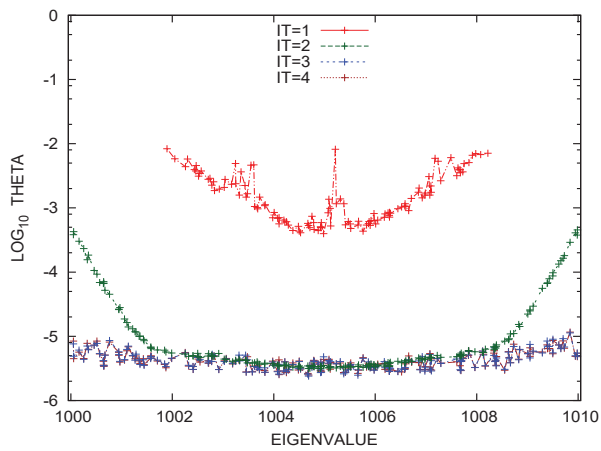


図 A.73 例題 C3-S : 固有値と相対残差 ($n = 4$, $g_s = 10^{-5}$)

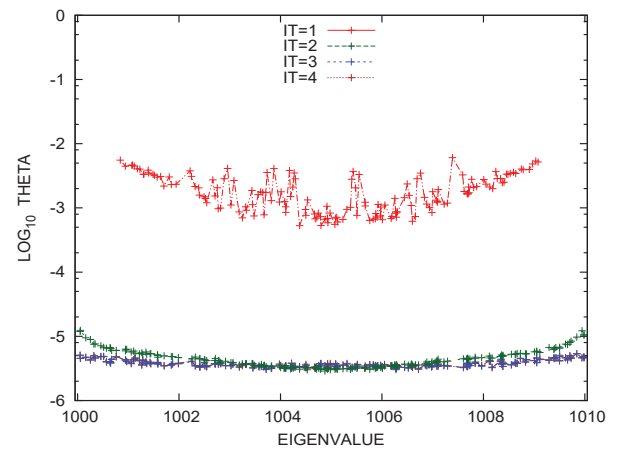


図 A.76 例題 C3-S : 固有値と相対残差 ($n = 10$, $g_s = 10^{-5}$)

A.3.6 例題 C4-S：中間固有対を解いた例 4（単精度計算）

この例題 C4-S で用いた FEM の要素分割は $(N_1, N_2, N_3) = (50, 60, 70)$ である。それから導かれる一般固有値問題の係数行列 A と B の次数は $N = 210000$ であり、下帯幅は $w_L = 3051$ である。求める固有対は固有値が中間の区間 $[a, b] = [100, 200]$ 内にあるものとした。そのような固有対の数は 801 個である。そうして、 $\mu = 1.5$ に対応するフィルタの通過域と遷移域を合併した区間 $[a', b'] = [75, 225]$ 内にある固有値の数は 1192 個であり、フィルタを適用するベクトルの数 m はそれよりも多くすることが望ましいので、この例では $m = 1300$ とした。

フィルタの次数が小さい $n = 4$ の場合（表 A-35 から表 A-37 まで、図 A-77 から図 A-79 まで）には、得られた近似対の数はフィルタの適用回数 $IT=1$ の場合には正しい数 801 個とは違っているが、そうして IT が 2 以上の場合には正しい数に一致している。これら 3 枚のグラフからは、近似対の相対残差が反復回数 IT を 2, 3, 4 と増すにつれて減少した様子がわかる。またこれら 3 つの表から相対残差の最大値 $\Theta_{\max}$ の減少の様子もわかる。

こんどは g_S を 10^{-5} にして、フィルタの次数 n を 4 から 10 まで 2 ずつ増やした場合（表 A-37 から表 A-40 まで、図 A-79 から図 A-82 まで）をみると、得られた近似対の数はフィルタの適用回数が $IT=1$ のときには正しい数 801 個とは異なるが、 IT が 2 以上のときにはどれも正しい数と一致している。そうしてフィルタの次数 n が 4, 6, 8, 10 の 4 枚の図のグラフから、それぞれの場合について、近似対の相対残差が反復回数 IT とともに順調に減少したことがわかる（これら 4 枚の図はどれも、 $IT=3$ のグラフ（青線）と $IT=4$ のグラフ（茶線）の線がほとんど重なっているので、改良は反復 3 回でほぼ終了していることがわかる）。

表 A-35 例題 C4-S：次数 $n = 4$, $g_S = 10^{-3}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>329</u>	1.9E-01
2	801	5.3E-02
3	801	9.4E-04
4	801	3.1E-05

表 A-36 例題 C4-S：次数 $n = 4$, $g_S = 10^{-4}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>598</u>	1.8E-01
2	801	8.8E-03
3	801	5.6E-05
4	801	2.6E-05

表 A-37 例題 C4-S：次数 $n = 4$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>701</u>	3.0E-01
2	801	2.4E-03
3	801	2.9E-05
4	801	2.6E-05

表 A-38 例題 C4-S：次数 $n = 6$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>800</u>	3.2E-01
2	801	2.2E-04
3	801	2.7E-05
4	801	2.7E-05

表 A-39 例題 C4-S：次数 $n = 8$, $g_S = 10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>825</u>	3.3E-01
2	801	8.3E-05
3	801	3.6E-05
4	801	3.5E-05

表 A-40 例題 C4-S：次数 $n=10$, $g_S=10^{-5}$

IT	得られた近似対の数	$\Theta_{\max}$
1	<u>828</u>	3.2E-01
2	801	6.0E-05
3	801	3.8E-05
4	801	3.8E-05

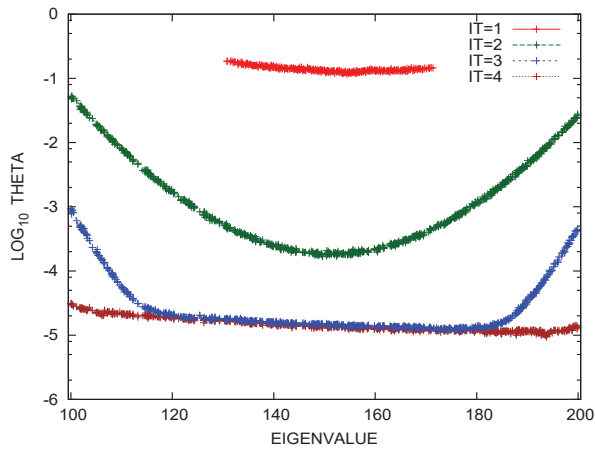


図 A.77 例題 C4-S : 固有値と相対残差 ($n = 4$, $g_s = 10^{-3}$)

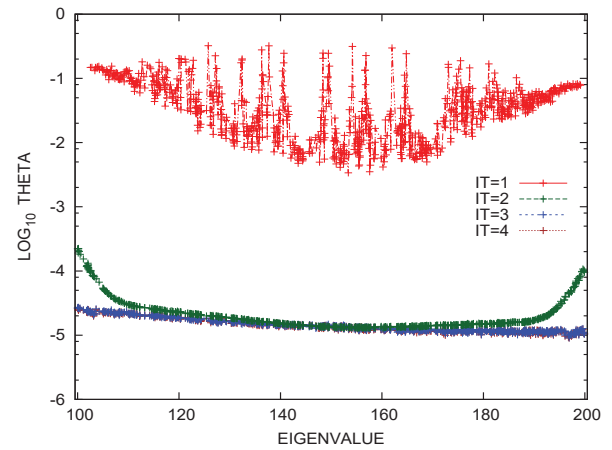


図 A.80 例題 C4-S : 固有値と相対残差 ($n = 6$, $g_s = 10^{-5}$)

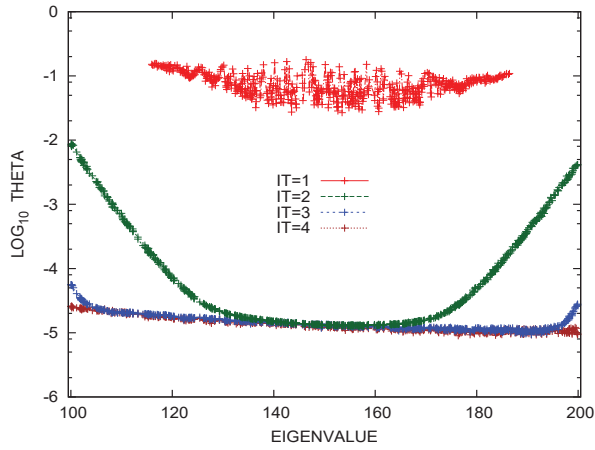


図 A.78 例題 C4-S : 固有値と相対残差 ($n = 4$, $g_s = 10^{-4}$)

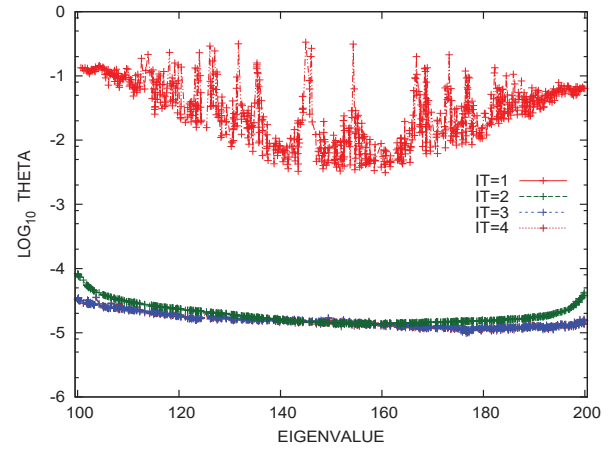


図 A.81 例題 C4-S : 固有値と相対残差 ($n = 8$, $g_s = 10^{-5}$)

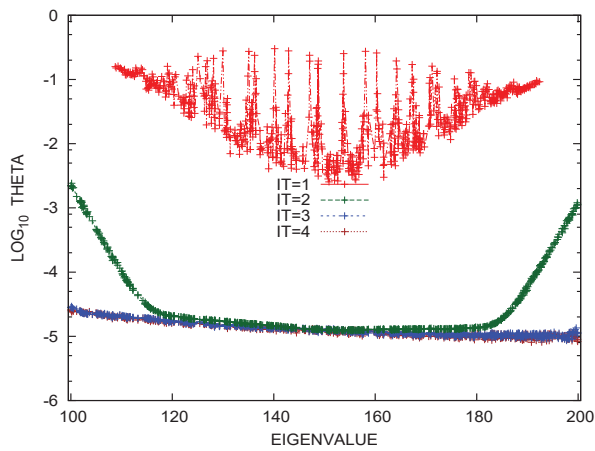


図 A.79 例題 C4-S : 固有値と相対残差 ($n = 4$, $g_s = 10^{-5}$)

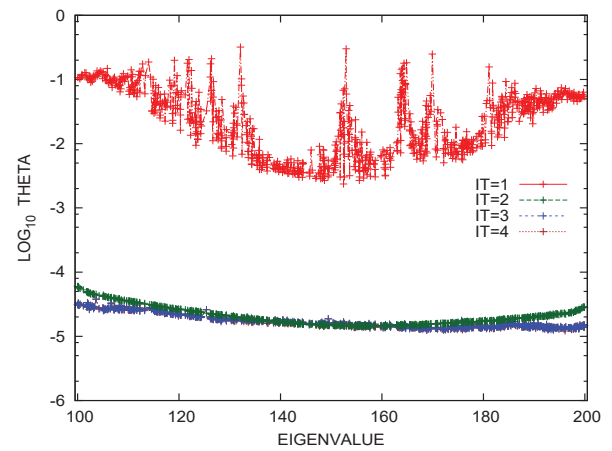


図 A.82 例題 C4-S : 固有値と相対残差 ($n = 10$, $g_s = 10^{-5}$)