フィルタを用いた固有値問題の近似解法について

On Approximate Solution Method of Eigenvalue Problems by Using Filters

東京都立大学・数理科学専攻　　村上弘[*1]
Hiroshi Murakami
Department of Mathematical Sciences, Tokyo Metropolitan University

Abstract

We approximate eigenpairs of an eigenproblem whose eigenvalues are in a specified interval by using a filter. In this study, we assume the filter is a Chebyshev polynomial of a linear combination of resolvents. The filter is applied to a set of vectors to reduce the proportion of unwanted eigenvectors contained in the resulted set, so to make it spans a good approximation of the required invariant subspace. In general situations, a filter is applied to a random set of vectors, but before the original filter is applied if another filter with lower performance but requires less effort is applied to reduce the proportion of unwanted eigenvectors, the approximation of the invariant subspace can be improved, which we confirmed in our experiments.

1　はじめに

指定された区間内に固有値がある固有値問題の少数の固有対全部を一斉に近似して求める．今回用いるフィルタはレゾルベントの線形結合の Chebyshev 多項式とする．ベクトルの組に対してフィルタを適用すると，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る．そのようにして得られたベクトルの組からうまく近似不変部分空間の基底を構成して，その基底に Rayleigh-Ritz 法を適用して必要な固有対の近似を取り出す．通常はランダムなベクトルの組にフィルタを直接適用するが，前処理として性能は低いが手間の少ない別のフィルタを適用する操作によりあらかじめ不要な固有ベクトルを含む割合を減らしておいてから本来のフィルタを適用することで固有対の近似精度を向上できる．そのことを実験により示す．

係数行列 A と B が実対称で，B は正定値である実対称定値一般固有値問題 $A\mathbf{v} = \lambda B\mathbf{v}$ の必要な固有対を求めることにする（この問題では，固有値はすべて実数で，固有ベクトルもすべて実ベクトルにとれる）．固有値が指定された区間 $[a, b]$ にある「必要な固有対」の数がたとえば数百程度以下の場合に，そのような固有対全部の近似をフィルタを利用して一斉に求めることにする．用いるフィルタ $\mathcal{F}$ は線形作用素で，必要な固有ベクトルは良く伝達するが不要なものはほとんど伝達しない性質を持つようにうまく構成する．十分多くのランダムに生成されたベクトルの組を B-正規直交化してベクトルの組 X を作り，それに対してフィルタ $\mathcal{F}$ を適用して得られるベクトルの組は「不要な固有ベクトル」を含む割合が小さいものになる．

「不要な固有ベクトル」を含む割合が少ないベクトルの組から，線形結合をうまく作って不変部分空間 $S_{[a,b]}$（固有値が区間 $[a, b]$ にある固有ベクトル全体で張られた）を良く近似する空間の線形独立性の良い基底を構成する．その基底に Rayleigh-Ritz 法を適用すると「必要な固有対」の近似が一斉に得られる．フィルタの伝達率が「必要な固有ベクトル」に対してばらつきが大きいと得られる近似固有対の精度もそれだ

[*1] 〒 191-0397 東京都八王子市南大沢 1-1　E-mail: mrkmhrsh@tmu.ac.jp

けばらつく傾向が生じる．しかしベクトルの組に対して「B-正規直交化に続いてフィルタを適用する」という操作を繰り返せば，必要な固有対の近似精度の不均一さを改善できることは既に示されている．

今回採用するフィルタの構成は「レゾルベントの線形結合」を Chebyshev 多項式に合成したものである．Chebyshev 多項式の次数が n のとき，3 項漸化式に従って「レゾルベントの線形結合」をベクトルの組に適用する回数は n になる．従来，フィルタを適用する前のベクトルの組として，ランダムに生成したベクトルの組を B-正規直交化したものを用いてきたが．今回は，Chebyshev 多項式の次数が n である本来のフィルタの適用対象とするベクトルの組として，ランダムなベクトルの組を B-正規直交化をしたものに前処理として Chebyshev 多項式の次数を下げた別のフィルタを適用してさらに B-正規直交化を施したものを用いることにする．

2　フィルタの構成

シフトが ρ のレゾルベントは $\mathcal{R}(\rho) \equiv (A - \rho B)^{-1} B$ とする．今回使用するフィルタ $\mathcal{F}$ は，K 個のレゾルベント $\mathcal{R}(\rho_j)$ の線形結合である $\mathcal{Y}$ に n 次 Chebyshev 多項式の定数倍を合成したものとして式 (1) で与えられる．

$$\left\{ \begin{array}{lcl} \mathcal{F} & = & g_{\rm s}\, T_n(\,\mathcal{Y}\,)\,, \\ \mathcal{Y} & = & c_\infty\, \mathcal{I} + \displaystyle\sum_{j=1}^{K} \gamma_j\, \mathcal{R}(\rho_j)\,. \end{array} \right. \tag{1}$$

レゾルベントの線形結合 $\mathcal{Y}$ を実作用素にするために，定数 c_∞ は実数とし，シフト ρ_j が実数であるレゾルベントに対する線形結合係数 γ_j は実数で，虚数であるシフトは複素共役対をなして現れて，シフトが互いに複素共役であるレゾルベントにたいする線形結合係数は互いに複素共役とする．ただし，固有値が下端付近や上端付近ではない中間固有対を求める場合には，すべてのシフトは実数を避けて虚数だけを用いる必要がある．

任意の固有対 $(\lambda, \mathbf{v})$ に対しては，$\mathcal{F}\,\mathbf{v} = f(\lambda)\,\mathbf{v}$ となる．ここで $f(\lambda)$ はフィルタ $\mathcal{F}$ の伝達関数であり，実有理関数 $y(\lambda)$ の n 次 Chebyshev 多項式の定数 g_s 倍であり，式 (2) で与えられる．

$$\left\{ \begin{array}{lcl} f(\lambda) & \equiv & g_{\rm s}\, T_n(\,y(\lambda)\,)\,, \\ y(\lambda) & \equiv & c_\infty + \displaystyle\sum_{j=1}^{K} \frac{\gamma_j}{\lambda - \rho_j}\,. \end{array} \right. \tag{2}$$

下端側の固有対を求める場合には，区間 $[a, b]$ の幅を $\mu\,(> 1)$ 倍に拡げた区間を $[\,a, b']$ として，伝達関数 $f(\lambda)$ が以下の式 (3) の各条件を満たすように，実数 c_∞ とレゾルベントのシフト ρ_j と結合係数 γ_j，$j = 1, 2, \ldots, K$，および Chebyshev 多項式の次数 n をうまく選ぶ．

$$\left\{ \begin{array}{lll} \text{（通過域）} & \lambda \in [\,a, b\,] \text{ では} & g_{\rm p} \le f(\lambda) \le 1\,, \\ \text{（遷移域）} & \lambda \in (\,b, b'\,) \text{ では} & g_{\rm s} < f(\lambda) < g_{\rm p}\,, \\ \text{（阻止域）} & \lambda \in [\,b', \infty) \text{ では} & |f(\lambda)| \le g_{\rm s}\,. \end{array} \right. \tag{3}$$

ここで 3 つのパラメタ μ，$g_{\rm s}$，$g_{\rm p}$ は，伝達関数 $f(\lambda)$ のグラフの形状を代表する値であり，2 つの閾値 $g_{\rm p}$ と $g_{\rm s}$ は条件 $0 < g_{\rm s} \ll g_{\rm p} < 1$ を満たすものとする．

必要な固有対の固有値の指定区間 $[a, b]$ が固有値分布の下端側である場合には，求めたい固有値の区間 $\lambda \in [a, b]$ を単位区間 $t \in [0, 1]$ に同じ向きで対応させる線形変換 $t = (\lambda - a)/(b - a)$ により固有値 λ の正規化座標 t を定義する．あるいは指定区間 $[a, b]$ が固有値分布の下端側でも上端側でもない中間の領域にある場合には，区間 $\lambda \in [a, b]$ を標準区間 $t \in [-1, 1]$ に同じ向きで対応させる線形変換 $t = \frac{2\lambda - a - b}{b - a}$ により固有

値 λ の正規化座標 t を定義する．そうして t を引数とする伝達関数を関係 $g(t) \equiv f(\lambda)$ で定義する．伝達関数 $g(t)$ のグラフの概形を図 1 に示す．

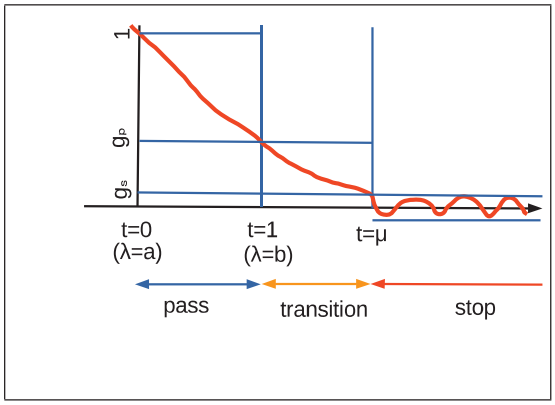


図 1: 伝達関数 $g(t) \equiv f(\lambda)$ の概形

固有値の座標 λ で表された伝達関数の中に現れる有理関数 (4) は無限遠で有限の値 c_∞ をとり，相異なる極が K 個の有理関数なので，それを座標の線形変換で正規化座標 t で表したものは式 (5) の形になる．

$$y(\lambda) \;=\; c_\infty + \sum_{j=1}^{K} \frac{\gamma_j}{\lambda - \rho_j}\,. \tag{4}$$

$$\widehat{y}(t) \;=\; c_\infty + \sum_{j=1}^{K} \frac{c_j}{t - \tau_j}\,. \tag{5}$$

座標 λ での極 ρ_j は座標 t での極 τ_j と対応しているとすると，$\gamma_j/(\lambda-\rho_j) = c_j/(t-\tau_j)$ である．そうして下端固有対を求める場合には，座標の変換が $t = (\lambda-a)/(b-a)$ であるから，$y(\lambda)$ と $\widehat{y}(t)$ の双方の極と極の係数の間には以下の関係がある．

$$\rho_j \equiv a + (b-a)\tau_j\,,\;\; \gamma_j \equiv (b-a)c_j\,. \tag{6}$$

中間固有対を求める場合にも $y(\lambda)$ と $\widehat{y}(t)$ の相互の極と極の係数の関係は座標変換から同様にして導かれる．

2.1 フィルタのベクトルの組への作用

ランダムなベクトル m 個を B-正規直交化してベクトルの組 X を作成する（つまり $X^T B X = I_m$ である）．このベクトルの組 X にフィルタ $\mathcal{F}$ を適用する．Chebyshev 多項式は 3 項漸化式 (7) を満たす．

$$\left\{\begin{array}{l} T_0(z) \;=\; 1,\quad T_1(z) \;=\; z, \\ T_\ell(z) \;=\; 2z\,T_{\ell-1}(z) - T_{\ell-2}(z),\;\; (\ell \geq 2)\,. \end{array}\right. \tag{7}$$

フィルタ $\mathcal{F}$ のベクトルの組 X への作用の結果 $\mathcal{F}X = g_{\rm s}\,T_n(\mathcal{Y})X$ は，まず $X^{(j)} \equiv T_j(\mathcal{Y})X$ とおいて $X^{(j)}$, $j = 1, 2, \ldots, n$ を漸化式 (8) で求めれば，$\mathcal{F}X = g_{\rm s}\,X^{(n)}$ で与えられる．

$$\left\{\begin{array}{l} X^{(0)} \;=\; X,\quad X^{(1)} \;=\; \mathcal{Y}\,X, \\ X^{(\ell)} \;=\; 2\mathcal{Y}\,X^{(\ell-1)} - X^{(\ell-2)},\;\; (\ell \geq 2). \end{array}\right. \tag{8}$$

ランダムなベクトルにフィルタを直接適用して得られる「必要な固有対」の精度は，伝達関数の特性の値 g_p が小さいほど数値丸め誤差の影響が拡大されて，不均一になる傾向がある．

3　Chebyshev 多項式の次数を変えたフィルタによる前処理

3.1　B-正規直交化を挟んで同一のフィルタを 2 回適用する場合

フィルタ $\mathcal{F}$ として，式 (1) で表されたレゾルベントの線形結合 $\mathcal{Y}$ と n 次 Chebyshev 多項式を合成したものを用いることにする．

以下の式 (9) で与えられるように，与えられたベクトルの組 X に対して間に B-正規直交化の操作を挟んでフィルタ $\mathcal{F}$ を 2 回適用した結果の Z を計算する処理では，作用素 $\mathcal{Y}$ をベクトルの組に適用する回数は合計 $2n$ になる．そこで計算量を 2 倍にまでは増やさずに固有対の近似精度をある程度向上させることを考える．

$$\left\{\begin{array}{lcl} X & \leftarrow & \text{ランダムな } B\text{−正規直交ベクトルm個の組；} \\ Z & \leftarrow & g_s\, T_n(\mathcal{Y})\, X\text{；※ 1 回目の多項式の次数が } n \text{ のフィルタ操作} \\ X & \leftarrow & Z \text{ の } B\text{−正規直交化；} \\ Z & \leftarrow & g_s\, T_n(\mathcal{Y})\, X\,.\ \text{※ 2 回目の多項式の次数が } n \text{ のフィルタ操作} \end{array}\right. \tag{9}$$

3.2　B-正規直交化を挟んで多項式の次数を変えたフィルタを適用する場合

以下の式 (10) で表されるように，ベクトルの組 X に対して，まず先に Chebyshev 多項式の次数を ν（$\nu < n$）に変更した「前処理用のフィルタ」を適用して，その結果に対して B-正規直交化を施した後に，Chebyshev 多項式の次数が n の「本来のフィルタ」を適用して Z を作る操作を行うと，レゾルベントの線形結合 $\mathcal{Y}$ の適用回数の合計は $\nu + n$ になる．そうして $T_\nu(\mathcal{Y})$ と $T_n(\mathcal{Y})$ で必要なレゾルベントの組は同一である．

$$\left\{\begin{array}{lcl} X & \leftarrow & \text{ランダムな } B\text{−正規直交ベクトルm個の組；} \\ Z & \leftarrow & \widetilde{g}_s\, T_\nu(\mathcal{Y})\, X\text{；※ 前処理用の多項式の次数が } \nu \text{ のフィルタ操作} \\ X & \leftarrow & Z \text{ の } B\text{−正規直交化；} \\ Z & \leftarrow & g_s\, T_n(\mathcal{Y})\, X\,.\ \text{※ 本来の多項式の次数が } n \text{ のフィルタ操作} \end{array}\right. \tag{10}$$

ここで $\widetilde{g}_s$ は ν 次のフィルタの伝達率の最大値を 1 に規格化するための定数であり（通過域での伝達率の最小値は $\widetilde{g}_p$），式 (11) により計算できる．

$$\left\{\begin{array}{lcl} \widetilde{g}_s & \equiv & 1 \big/ \cosh\big(\dfrac{\nu}{n} \cosh^{-1} \dfrac{1}{g_s}\big)\,, \\ \widetilde{g}_p & \equiv & \widetilde{g}_s \cosh\big(\dfrac{\nu}{n} \cosh^{-1} \dfrac{g_p}{g_s}\big)\,. \end{array}\right. \tag{11}$$

4　実験

4.1　計算環境

実験に用いた計算機システムは東京大学情報基盤センター Oakbridge-CX の 1 ノード（`PRIMERGY CX2550 M5` あるいは `CX2560 M5`）である．ノードあたりの CPU の数は 2 つで CPU は `Intel Xeon Platinum 8280(CascadeLake)`(2.7GHz) である．コア数は 28/CPU (56/ノード) で，L3 キャッシュの容量は 38.5MiB/CPU，拡張命令セットのレベルは AVX-512，理論ピーク性能の値は 4.84TFLOPS(DP)/ノード．主記憶 DDR4 メモリの容量は 192GiB/ノード，ノードのメモリバンド幅は 281.6GB/s である．Fortran は intel `ifort`，バージョン `"19.1.3.304 20200925"`．計算は 1 ノード（2CPU）で，OpenMP を用いてノード内のコア数に等しい 56 スレッドで並列に実行した．

4.2　固有値問題の例題と，近似固有対の精度の評価法

例題とした一般固有値問題 $A\mathbf{v} = \lambda B\mathbf{v}$ は，1 辺の長さが π の立方体領域において零境界条件を課したラプラシアン $-\Delta$ の有限要素法（FEM）による離散化で生じたものである．FEM の各要素は立方体を各辺方向に N_1+1，N_2+1，N_3+1 に等分して得られる直方体とし，要素内の基底関数は最低次の 3 重線形関数とした．そうして通常の倍精度による計算では要素分割は $(N_1, N_2, N_3) = (40, 50, 60)$ とした．その場合には係数行列 A と B の次数は $N = 120,000$（12 万）で下帯幅は $w_L = 2,041$ である（四倍精度による計算の場合には，計算に時間がかかりすぎるため，要素分割の数を減らし $(N_1, N_2, N_3) = (30, 40, 50)$ にして問題の規模を小さくした．その場合には $N = 60,000$（6 万），$w_L = 1,231$ である）．

固有値分布の下端区間 $[3, 30]$ にある 54 個の固有対を近似して求めることにして，最初に与えるランダムなベクトルの数は 110 とした（四倍精度の場合は固有値問題としては係数行列が異なるが，固有値が区間 $[3, 30]$ に含まれる固有対の数は同じく 54 で，最初に与えるランダムなベクトルの数も 110 としている）．

特に断らない場合には数値と演算には倍精度（IEEE 754，binary64）を用いた（intel Fortran は四倍精度として IEEE 754 の binary128 を使用している）．今回の計算ではレゾルベントを与える（右辺が多数の）連立 1 次方程式の係数行列はすべて正定値対称なので，倍精度計算の場合には，intel MKL ライブラリから，LAPACK の実対称帯行列に対する Cholesky 分解を用いたルーチンを使用した（四倍精度計算の場合には，自作の実対称帯行列に対する改訂 Cholesky 分解を用いたルーチンを使用した）．

得られた近似固有対 $(\lambda, \mathbf{v})$ の精度は，式 (12) により与えられる「相対残差の大きさ」Θ を用いて評価した．

$$\Theta \equiv \frac{||A\mathbf{v} - \lambda B\mathbf{v}||_2}{||\lambda B\mathbf{v}||_2} \,. \tag{12}$$

4.3　フィルタの構成の詳細

今回用いたフィルタの構成は式 (1) で与えられ，対応する伝達関数は式 (2) で与えられる．

以下に具体的な構成を示したそれぞれのフィルタは，過去の文献 [1] においてその構成法も含めて示したものである（ただしたとえば同文献でのフィルタ Ex1-1 は本論文ではフィルタ F1-1 などと名称の先頭の”Ex”をすべて”F”に変更している）．例としてこれらのフィルタを取り上げたのは，それらが特に優れたものであることを意味するものではなくて，今回の前処理操作の有効性を示すためである．

> 文献 [1] で構成したフィルタは，精度改良のために反復して適用したりあるいは今回のような前処理を併用することをまったく想定せずに設計されたものである．フィルタ伝達関数の g_p をなるべく大きくして 1 に近づけ，また g_s をなるべく微小にするように努力をしたが，計算量に直結する Chebyshev 多項式の次数 n についてはあまり小さくしようとしていなかった．レゾルベントのシフトをすべて実数にすると，計算はすべて数値や演算も実数だけで実現でき，係数行列が複素対称な連立 1 次方程式を扱わずに済むが，使用するレゾルベントの数を同じに揃えた場合のフィルタとしての性能は虚数をシフトに利用する場合に比べると悪い．

レゾルベントのシフト ρ_j はすべて実数で，線形変換 $t = (\lambda - a)/(b - a)$ により固有値分布の下端の区間 $\lambda \in [a, b]$ を単位区間 $t \in [0, 1]$ に対応させる．そうして以下では使用した各フィルタごとに $y(\lambda) \equiv \widehat{y}(t)$ の関係から，正規化座標による $\widehat{y}(t)$ の部分分数分解の形の式 (5) に現れる係数と極の値 c_∞，τ_j，c_j，$j = 1, 2, \ldots, K$ をそれぞれ示す．フィルタの構成で必要となるレゾルベントの線形結合 $\mathcal{Y}$ を具体的に作る際には，係数 c_∞ 以外のシフト ρ_j と結合係数 γ_j は式 (6) を計算することで求める．

4.3.1 実数シフトのレゾルベントを 1 つ用いたフィルタの構成

実験で用いた実数シフトのレゾルベントを 1 つ用いたフィルタのそれぞれについて，Chebyshev 多項式の次数 n と伝達関数の 3 つの形状パラメタ μ，g_s，g_p の値を表 1 に示す．さらに各フィルタを構成するレゾルベントの線形結合 $\mathcal{Y}$ に対応する関数 $y(\lambda)$ を λ の正規化座標 t を用いて表した有理関数 $\widehat{y}(t)$ を与える式に現れる各パラメタ c_∞，τ_1，c_1 の値をそれぞれ示す．

表 1: レゾルベントを 1 つ用いたフィルタの次数と形状パラメタ

フィルタ名	n	μ	g_s	g_p
F1-1	27	2.0	1E-12	1E-4
F1-2	28	1.75	1E-15	1E-6
F1-3	33	1.5	1E-15	1E-7
F1-4	25	1.5	1E-16	1E-8
F1-5	109	1.3	1E-15	1E-8

フィルタ F1-1：
$c_\infty = -0.96358\,45465\,07514\,72$, $\tau_1 = -6.51708\,92693\,11496\,6$, $c_1 = 16.72402\,48704\,45035$.
フィルタ F1-2：
$c_\infty = -0.99245\,32413\,90484\,24$, $\tau_1 = -3.86666\,39819\,61619\,7$, $c_1 = 11.19094\,03566\,60613$.
フィルタ F1-3：
$c_\infty = -0.98699\,30684\,12698\,48$, $\tau_1 = -4.75998\,30140\,15106\,7$, $c_1 = 12.43854\,28572\,29249$.
フィルタ F1-4：
$c_\infty = -0.91180\,78252\,55441\,48$, $\tau_1 = -2.11584\,43350\,13528\,4$, $c_1 = \ \ 6.91279\,94945\,84421\,8$.
フィルタ F1-5：
$c_\infty = \ -0.98752\,04260\,92374\,90$, $\tau_1 = -49.03287\,74799\,46448$, $c_1 = \ 100.03762\,20953\,9847$.

4.3.2 実数シフトのレゾルベントを 2 つ用いたフィルタの構成

実験で用いた実数シフトのレゾルベントを 2 つ用いたフィルタのそれぞれについて，Chebyshev 多項式の次数 n と 3 つの形状パラメタの値を表 2 に示す．さらにレゾルベントの線形結合に対応する有理関数 $\widehat{y}(t)$ の式に現れる各パラメタの値をそれぞれ示す．

表 2: レゾルベントを 2 つ用いたフィルタの次数と形状パラメタ

フィルタ名	n	μ	g_s	g_p
F2-1	23	1.5	1E-10	1E-3
F2-2	38	1.5	1E-12	1E-3
F2-3	38	1.5	1E-15	1E-5
F2-4	40	1.5	1.1E-15	1E-4

フィルタ F2-1：
$c_\infty = -0.99230\,41602\,40466\,38$, $\tau_1 = -0.76$, $\tau_2 = -1.2$,
$c_1 = \ -6.90373\,44873\,22084\,4$, $c_2 = \ \ 13.62704\,56201\,57944$.
フィルタ F2-2：
$c_\infty = -0.94058\,93038\,64582\,17$, $\tau_1 = -1.25$, $\tau_2 = -2.03$,
$c_1 = -7.55520\,41413\,37089\,6$, $c_2 = \ 16.54841\,50131\,58312$.

フィルタ F2-3：

$c_\infty = -0.97197\,32868\,06298\,33$, $\tau_1 = -0.8,\ \tau_2 = -2.5$,

$c_1 = -1.37229\,04037\,26818\,0$, $c_2 = \ 10.27448\,51537\,06616$.

フィルタ F2-4：

$c_\infty = -0.99864\,90114\,17620\,76$, $\tau_1 = -1.25,\ \tau_2 = -1.28$,

$c_1 = -207.65212\,64969\,0007$, $c_2 = \ \ 215.47366\,66740\,6178$.

4.3.3　実数シフトのレゾルベントを 3 つ用いたフィルタの構成

実験で用いた実数シフトのレゾルベントを 3 つ用いたフィルタのそれぞれについて，Chebyshev 多項式の次数 n と 3 つの形状パラメタの値を表 3 に示す．さらにレゾルベントの線形結合に対応する有理関数 $\widehat{y}(t)$ の式に現れる各パラメタの値をそれぞれ示す．

表 3: レゾルベントを 3 つ用いたフィルタの次数と形状パラメタ

フィルタ名	n	μ	g_s	g_p
F3-I-1	20	1.5	1E-10	1E-3
F3-I-2	30	1.5	1E-10	1E-2
F3-I-3	30	1.5	1E-12	1E-3
F3-I-4	40	1.5	1E-12	5E-3
F3-I-5	20	1.5	1.2E-12	1E-3
F3-I-6	30	1.5	2E-14	1E-3
F3-II-1	30	1.5	1E-10	1E-2
F3-II-2	34	1.5	1E-11	1E-2
F3-II-3	41	1.5	1E-12	1E-2

フィルタ F3-I-1：

$c_\infty = 1.0,\ \tau_1 = -1.5,\ \tau_2 = -5.0,\ \tau_3 = -7.5$,

$c_1 = \ \ -10.54347\,20670\,71244$, $c_2 = \ \ \ 230.91814\,77483\,9765$, $c_3 = -288.10240\,37581\,0610$.

フィルタ F3-I-2：

$c_\infty = -0.65,\ \tau_1 = -2.5,\ \tau_2 = -4.0,\ \tau_3 = -8.0$,

$c_1 = \ \ -46.94773\,12232\,16199$, $c_2 = \ \ \ 138.44834\,07901\,2331$, $c_3 = -111.96218\,15278\,0179$.

フィルタ F3-I-3：

$c_\infty = 0.2,\ \tau_1 = -2.5,\ \tau_2 = -4.0,\ \tau_3 = -8.0$,

$c_1 = \ \ -61.72673\,49701\,25999$, $c_2 = \ \ \ 194.11248\,93001\,2096$, $c_3 = -181.08421\,32370\,6878$.

フィルタ F3-I-4：

$c_\infty = 1.0$, $\tau_1 = -3.65,\ \tau_2 = -5.5,\ \tau_3 = -10.5$,

$c_1 = -130.27325\,48867\,5910,\ c_2 = \ \ \ 374.74537\,92283\,5478$, $c_3 = -338.87131\,83307\,7839$.

フィルタ F3-I-5：

$c_\infty = 1.0,\ \tau_1 = -\,1.73,\ \tau_2 = -\,2.65,\ \tau_3 = -\,2.87$,

$c_1 = \ \ -207.17748\,65340\,6107$, $c_2 = \ \ \ 2173.81997\,44641\,517$, $c_3 = -2008.75967\,13432\,417$.

フィルタ F3-I-6：

$c_\infty = 1.0,\ \tau_1 = -3.3,\ \tau_2 = -3.4,\ \tau_3 = -3.6$,

$c_1 = -27705.17451\,93703\,20$, $c_2 = \ \ \ 43785.11285\,87155\,20$, $c_3 = -16135.51239\,55055\,96$.

フィルタ F3-II-1：

$c_\infty = \ -0.99901\,74798\,13294\,42$, $\tau_1 = -0.60,\ \tau_2 = -0.69,\ \tau_3 = -0.81$,

$c_1 = \ \ \ \ 89.32133\,62718\,75544$, $c_2 = -210.39740\,96156\,3449$, $c_3 = \ \ 128.29029\,52793\,5843$.

フィルタ F3-II-2：

$c_\infty = -0.98131\,60097\,38869\,53$, $\tau_1 = -0.63$, $\tau_2 = -0.64$, $\tau_3 = -0.74$,

$c_1 = \quad 1805.99131\,75101\,005$, $c_2 = -2052.65714\,16016\,740$, $c_3 = \quad 253.75537\,27298\,5273$.

フィルタ F3-II-3：

$c_\infty = -0.99988\,50444\,27308\,67$, $\tau_1 = -0.722$, $\tau_2 = -0.820$, $\tau_3 = -0.837$,

$c_1 = \quad 236.05420\,17668\,8206$, $c_2 = -2101.86563\,74776\,586$, $c_3 = \quad 1873.66974\,20694\,337$.

4.3.4 実数シフトのレゾルベントを 4 つ用いたフィルタの構成

実験で用いた実数シフトのレゾルベントを 4 つ用いたフィルタのそれぞれについて，Chebyshev 多項式の次数 n と 3 つの形状パラメタの値を表 4 に示す．さらにレゾルベントの線形結合に対応する有理関数 $\widehat{g}(t)$ の式に現れる各パラメタの値をそれぞれ示す.

表 4: レゾルベントを 4 つ用いたフィルタの次数と形状パラメタ

フィルタ名	n	μ	g_s	g_p
F4-I-1	23	1.5	1E-12	1E-2
F4-I-2	40	1.5	1E-15	1E-2
F4-I-3	50	1.5	1E-16	1E-2
F4-II-1	63	1.5	1E-12	1E-1

フィルタ F4-I-1：

$c_\infty = 0.72830\,15709\,44898\,84$, $\tau_1 = -0.46$, $\tau_2 = -1.3$, $\tau_3 = -1.6$, $\tau_4 = -2.2$,

$c_1 = \quad 11.10360\,90168\,58801$, $c_2 = -747.44207\,46055\,1708$,

$c_3 = 1358.76323\,95652\,740$, $c_4 = -654.01397\,05430\,8215$.

フィルタ F4-I-2：

$c_\infty = 0.73890\,25275\,28577\,41$, $\tau_1 = -0.8$, $\tau_2 = -0.85$, $\tau_3 = -3.4$, $\tau_4 = -4.4$,

$c_1 = \quad 175.18672\,76796\,1545$, $c_2 = -206.28364\,42406\,7430$,

$c_3 = \quad 413.29017\,80931\,2039$, $c_4 = -427.58306\,04243\,9581$.

フィルタ F4-I-3：

$c_\infty = 0.93709\,12122\,38981\,41$, $\tau_1 = -0.5$, $\tau_2 = -1.4$, $\tau_3 = -4.9$, $\tau_4 = -5.9$,

$c_1 = 1.95463\,78129\,48879\,3$, $c_2 = \quad -26.92005\,26898\,41116$,

$c_3 = \quad 639.14392\,55792\,3146$, $c_4 = -677.08425\,05864\,2284$.

フィルタ F4-II-1：

$c_\infty = -0.97891\,04015\,29457\,29$, $\tau_1 = -0.420$, $\tau_2 = -0.617$, $\tau_3 = -0.995$, $\tau_4 = -1.075$,

$c_1 = \quad -8.42245\,36308\,00969\,2$, $c_2 = \quad 49.95291\,83198\,95732$,

$c_3 = -390.72185\,12457\,4242$, $c_4 = \quad 358.88151\,30449\,7295$.

4.4 前処理を施した実験例

- **実数シフトのレゾルベントを 1 つ用いたフィルタによる例**

  図 2〜6 のそれぞれに，左側にはフィルタの伝達関数の大きさ $|g(t)|$ の対数値をプロットしたグラフを，右側には前処理用のフィルタの次数 ν を 0 から 25 まで 5 刻みでとって求めた近似固有対の相対残差の大きさの対数値を縦軸に近似固有対の固有値を横軸にとってプロットしたグラフを示す.

- **実数シフトのレゾルベントを 2 つ用いたフィルタによる例**

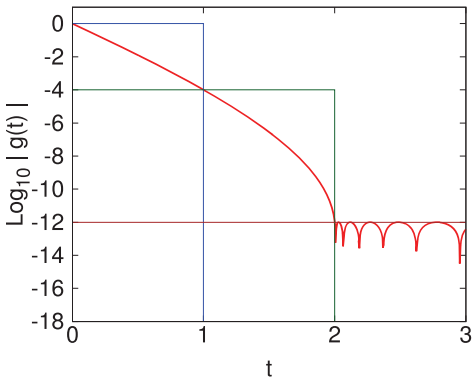
  図 7〜10 のそれぞれに，左側にはフィルタの伝達関数の大きさ $|g(t)|$ の対数値をプロットしたグラフを，右側には前処理用のフィルタの次数 ν を 0 から 25 まで 5 刻みでとって求めた近似固有対の相対残差の大きさの対数値を縦軸に近似固有対の固有値を横軸にとってプロットしたグラフを示す．

- **実数シフトのレゾルベントを 3 つ用いたフィルタによる例**

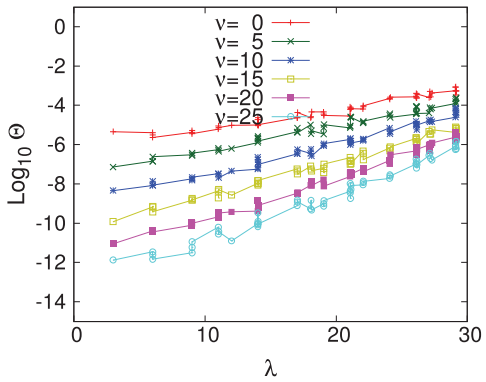
  図 11〜19 のそれぞれに，左側にはフィルタの伝達関数の大きさ $|g(t)|$ の対数値をプロットしたグラフを，右側には前処理用のフィルタの次数 ν を 0 から 25 まで 5 刻みでとって求めた近似固有対の相対残差の大きさの対数値を縦軸に近似固有対の固有値を横軸にとってプロットしたグラフを示す．

- **実数シフトのレゾルベントを 4 つ用いたフィルタによる例**

  図 20〜23 のそれぞれに，左側にはフィルタの伝達関数の大きさ $|g(t)|$ の対数値をプロットしたグラフを，右側には前処理用のフィルタの次数 ν を 0 から 25 まで 5 刻みでとって求めた近似固有対の相対残差の大きさの対数値を縦軸に近似固有対の固有値を横軸にとってプロットしたグラフを示す．

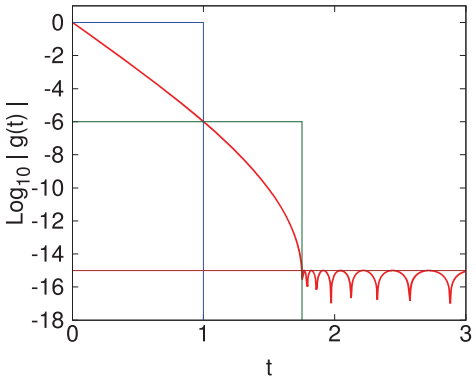


フィルタの伝達関数の大きさ $|g(t)|$ の対数

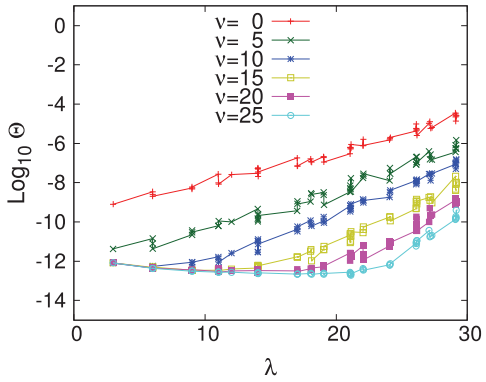


近似固有対の相対残差の大きさ Θ の対数

図 2: フィルタ F1-1 （$n = 27$，$\mu = 2.0$，g_s =1E-12，g_p =1E-4 ）

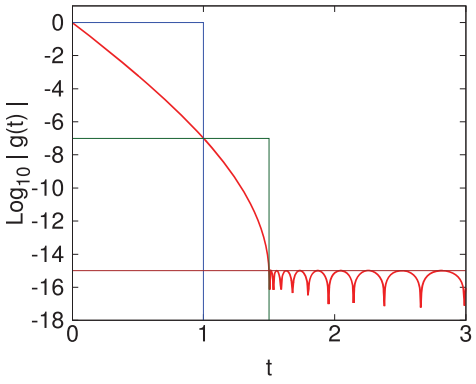


フィルタの伝達関数の大きさ $|g(t)|$ の対数

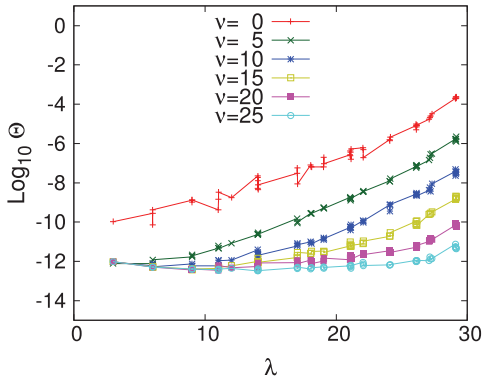


近似固有対の相対残差の大きさ Θ の対数

図 3: フィルタ F1-2 （$n = 28$，$\mu = 1.75$，g_s =1E-15，g_p =1E-6 ）

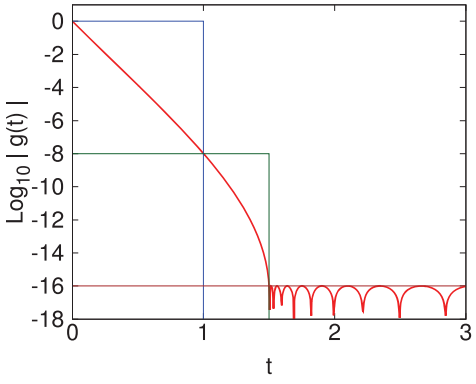


フィルタの伝達関数の大きさ $|g(t)|$ の対数

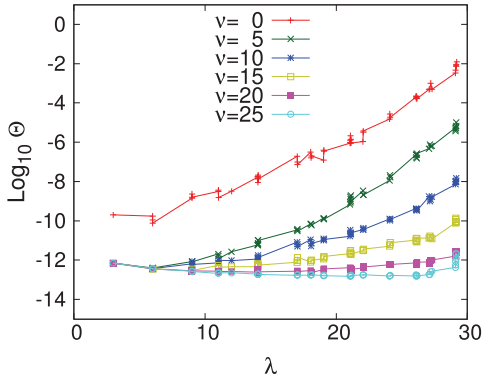


近似固有対の相対残差の大きさ Θ の対数

図 4: フィルタ F1-3 （$n = 33$, $\mu = 1.5$, g_s =1E-15, g_p =1E-7 ）

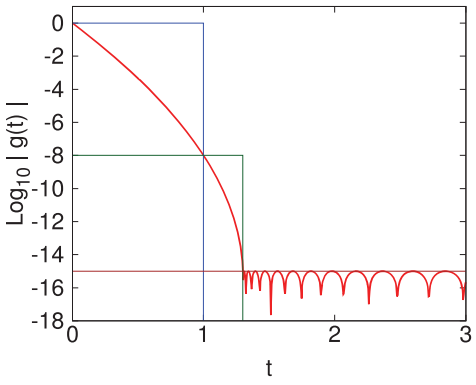


フィルタの伝達関数の大きさ $|g(t)|$ の対数

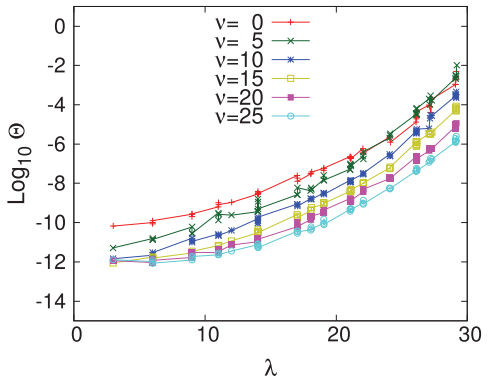


近似固有対の相対残差の大きさ Θ の対数

図 5: フィルタ F1-4 （$n = 25$, $\mu = 1.5$, g_s =1E-16, g_p =1E-8 ）

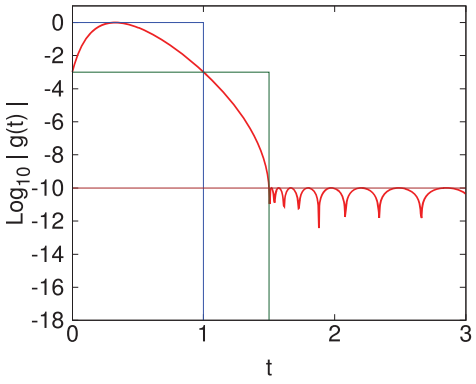


フィルタの伝達関数の大きさ $|g(t)|$ の対数

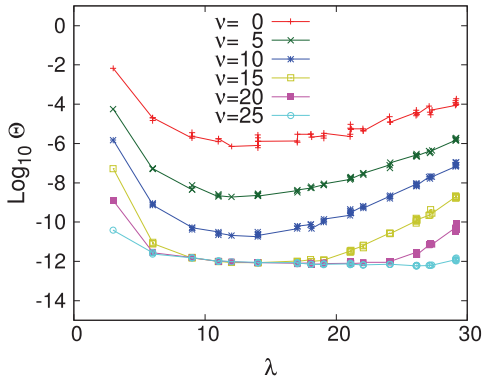


近似固有対の相対残差の大きさ Θ の対数

図 6: フィルタ F1-5 （$n = 109$, $\mu = 1.3$, g_s =1E-15, g_p =1E-8 ）

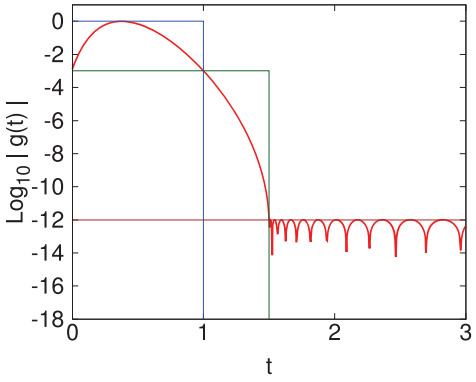


フィルタの伝達関数の大きさ $|g(t)|$ の対数

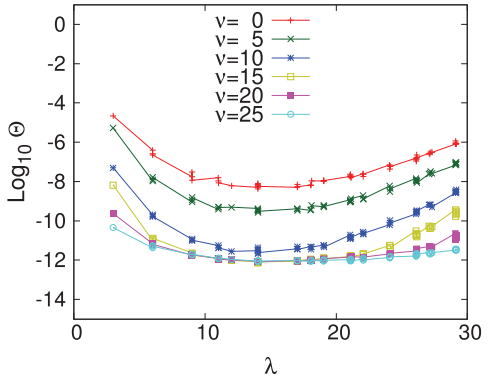


近似固有対の相対残差の大きさ Θ の対数

図 7: フィルタ F2-1 （$n = 23$, $\mu = 1.5$, g_s =1E-10, g_p =1E-3 ）

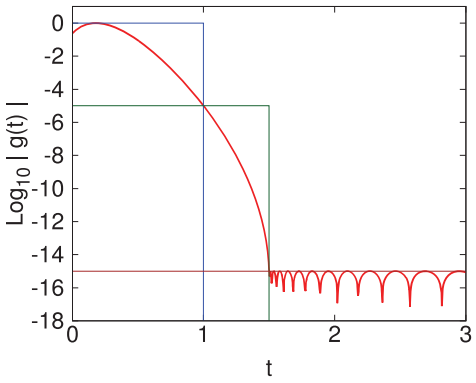


フィルタの伝達関数の大きさ $|g(t)|$ の対数

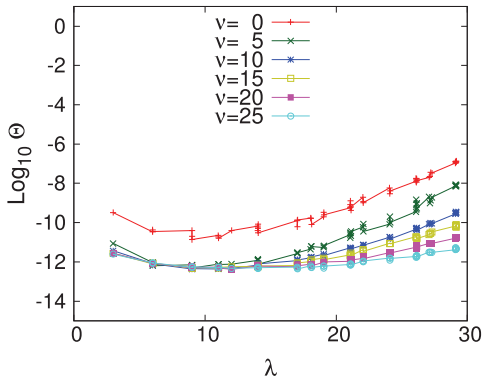


近似固有対の相対残差の大きさ Θ の対数

図 8: フィルタ F2-2 （$n = 38$, $\mu = 1.5$, g_s =1E-12, g_p =1E-3 ）

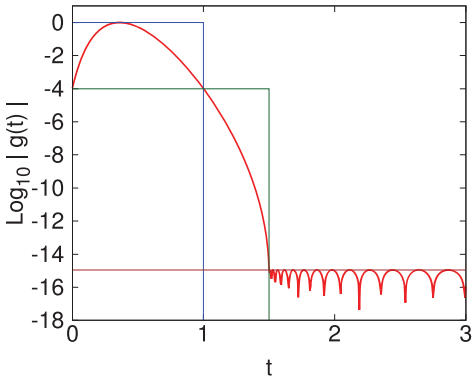


フィルタの伝達関数の大きさ $|g(t)|$ の対数

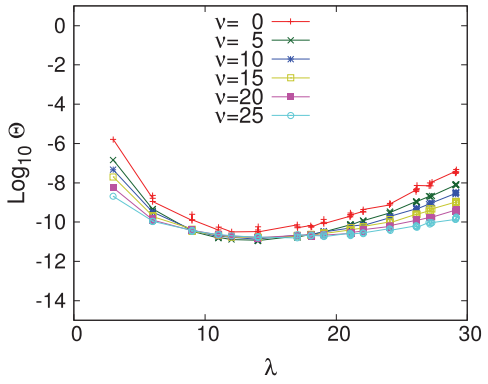


近似固有対の相対残差の大きさ Θ の対数

図 9: フィルタ F2-3 （$n = 38$, $\mu = 1.5$, g_s =1E-15, g_p =1E-5 ）

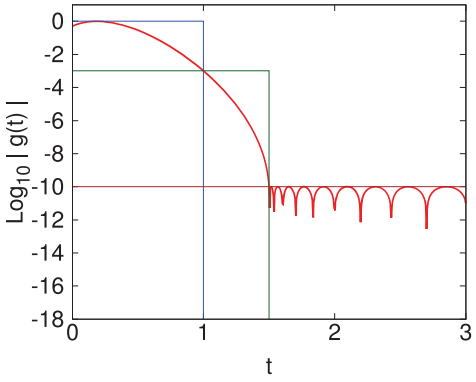


フィルタの伝達関数の大きさ $|g(t)|$ の対数

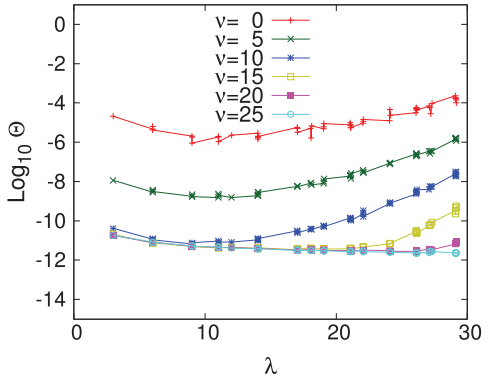


近似固有対の相対残差の大きさ Θ の対数

図 10: フィルタ F2-4 （$n = 40$， $\mu = 1.5$， g_s =1.1E-15， g_p =1E-4 ）

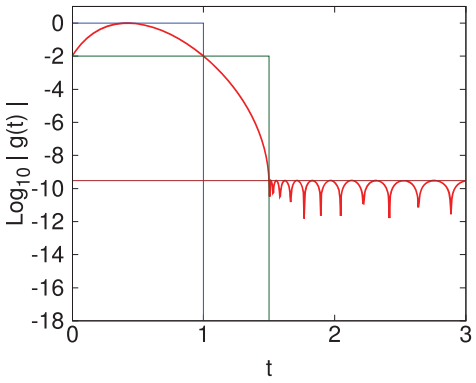


フィルタの伝達関数の大きさ $|g(t)|$ の対数

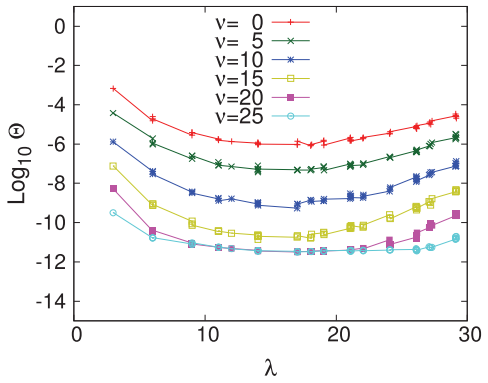


近似固有対の相対残差の大きさ Θ の対数

図 11: フィルタ F3-I-1 （$n = 20$， $\mu = 1.5$， g_s =1.1E-10， g_p =1E-3 ）

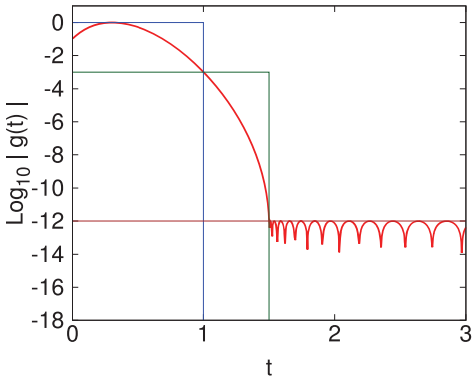


フィルタの伝達関数の大きさ $|g(t)|$ の対数

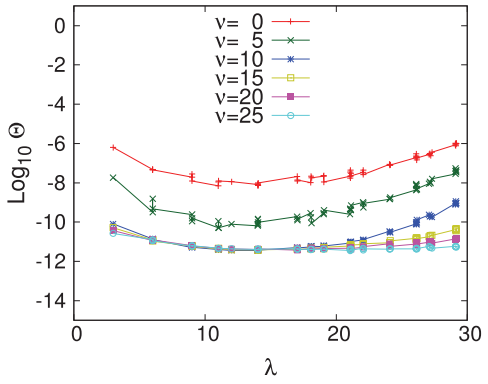


近似固有対の相対残差の大きさ Θ の対数

図 12: フィルタ F3-I-2 （$n = 30$， $\mu = 1.5$， g_s =3E-10， g_p =1E-2 ）

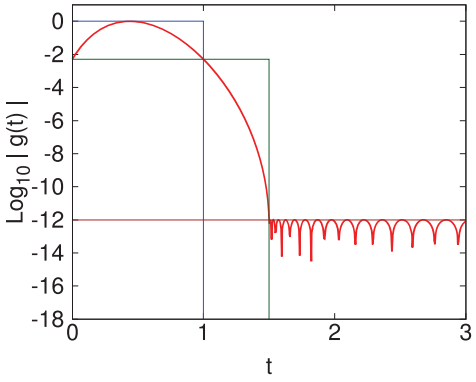


フィルタの伝達関数の大きさ $|g(t)|$ の対数

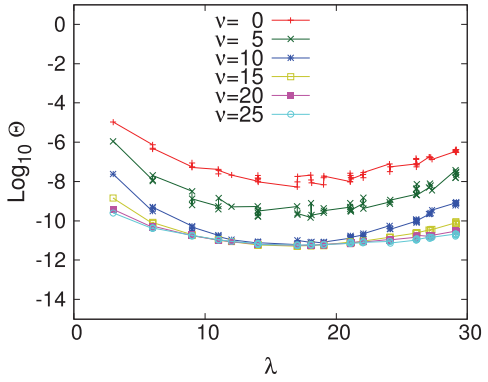


近似固有対の相対残差の大きさ Θ の対数

図 13: フィルタ F3-I-3 （$n = 30$, $\mu = 1.5$, g_s =3E-12, g_p =1E-3 ）

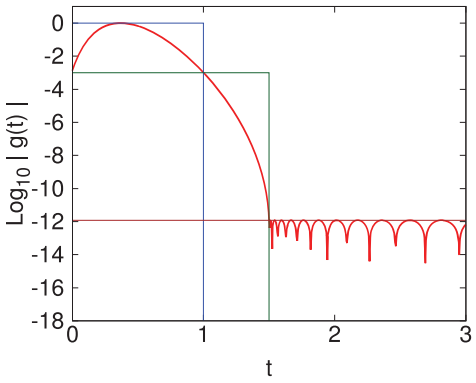


フィルタの伝達関数の大きさ $|g(t)|$ の対数

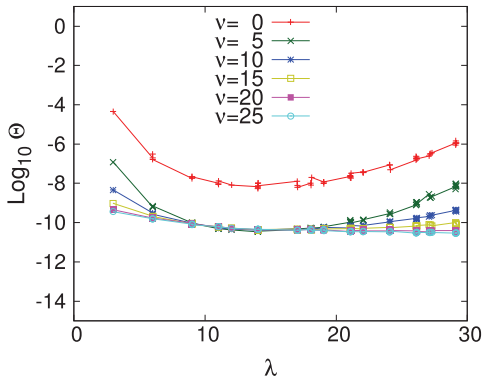


近似固有対の相対残差の大きさ Θ の対数

図 14: フィルタ F3-I-4 （$n = 40$, $\mu = 1.5$, g_s =1E-12, g_p =5E-3 ）

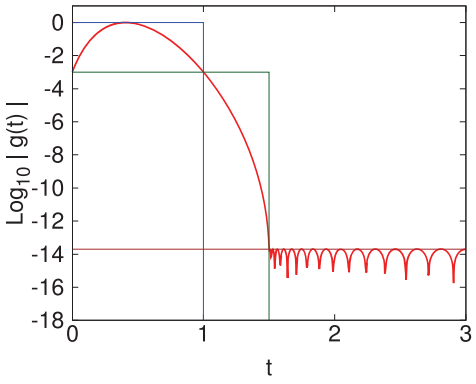


フィルタの伝達関数の大きさ $|g(t)|$ の対数

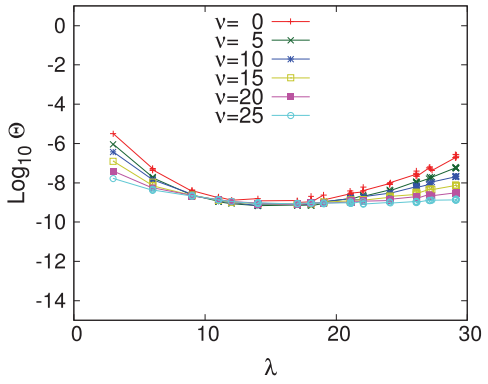


近似固有対の相対残差の大きさ Θ の対数

図 15: フィルタ F3-I-5 （$n = 20$, $\mu = 1.5$, g_s =1.2E-12, g_p =1E-3 ）

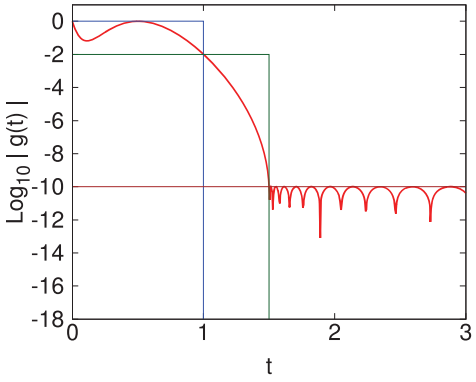


フィルタの伝達関数の大きさ $|g(t)|$ の対数

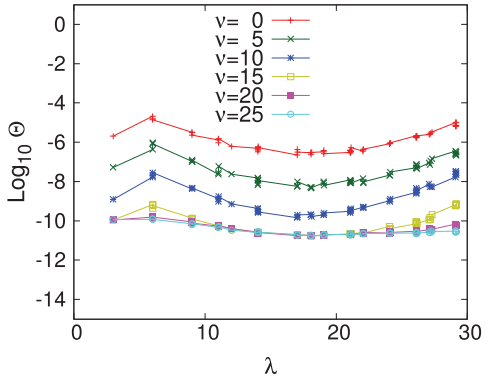


近似固有対の相対残差の大きさ Θ の対数

図 16: フィルタ F3-I-6 （$n = 30$, $\mu = 1.5$, g_s =2E-14, g_p =1E-3 ）

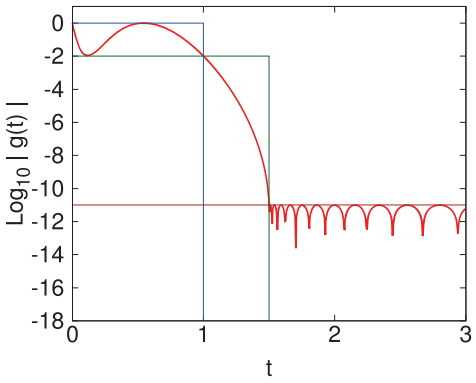


フィルタの伝達関数の大きさ $|g(t)|$ の対数

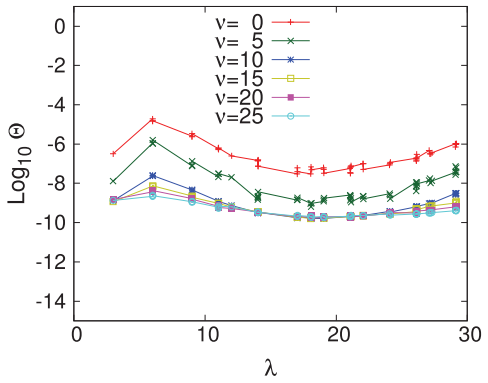


近似固有対の相対残差の大きさ Θ の対数

図 17: フィルタ F3-II-1 （$n = 30$, $\mu = 1.5$, g_s =1E-10, g_p =1E-2 ）

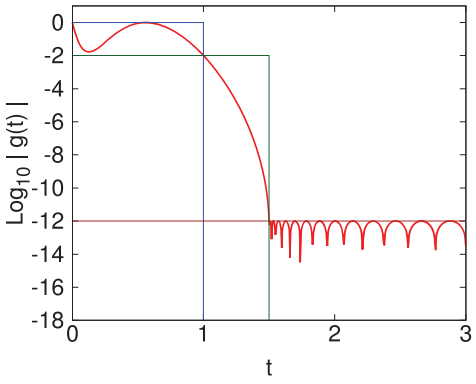


フィルタの伝達関数の大きさ $|g(t)|$ の対数

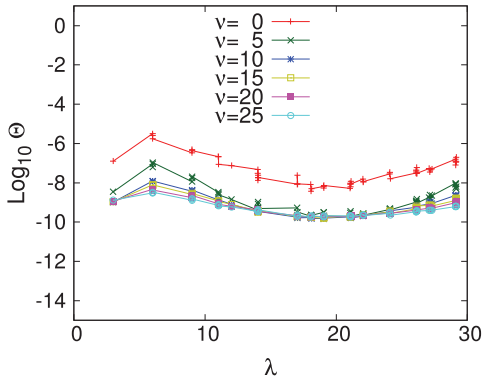


近似固有対の相対残差の大きさ Θ の対数

図 18: フィルタ F3-II-2 （$n = 34$, $\mu = 1.5$, g_s =1E-11, g_p =1E-2 ）

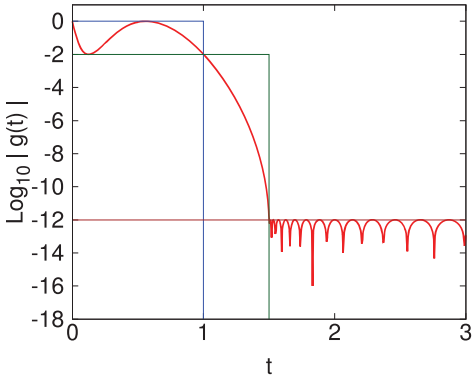


フィルタの伝達関数の大きさ $|g(t)|$ の対数

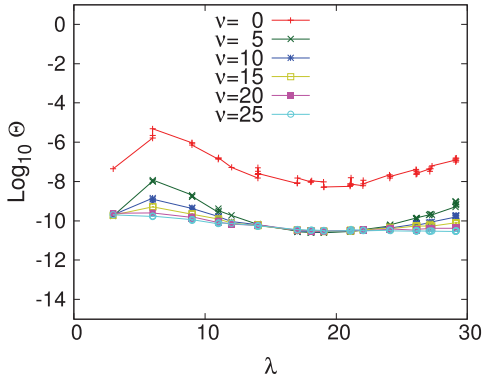


近似固有対の相対残差の大きさ Θ の対数

図 19: フィルタ F3-II-3 （$n = 41$, $\mu = 1.5$, g_s =1E-12, g_p =1E-2 ）

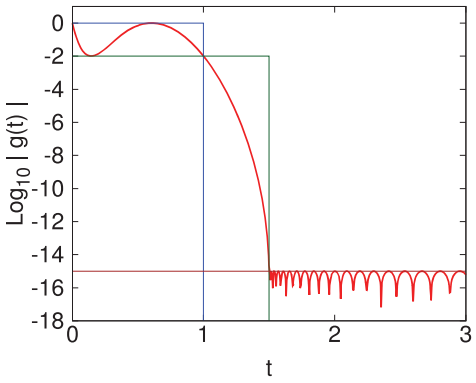


フィルタの伝達関数の大きさ $|g(t)|$ の対数

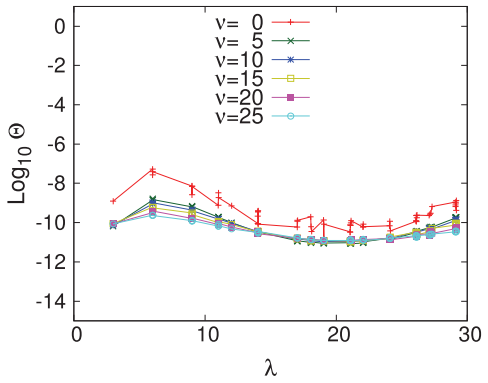


近似固有対の相対残差の大きさ Θ の対数

図 20: フィルタ F4-I-1 （$n = 23$, $\mu = 1.5$, g_s =1E-12, g_p =1E-2 ）

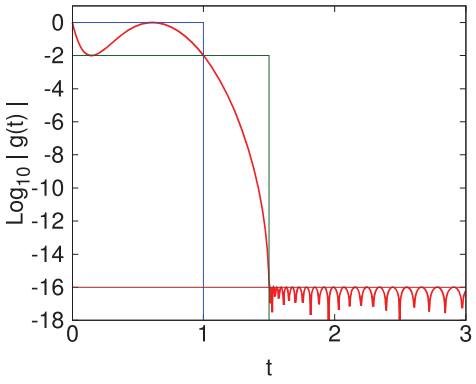


フィルタの伝達関数の大きさ $|g(t)|$ の対数

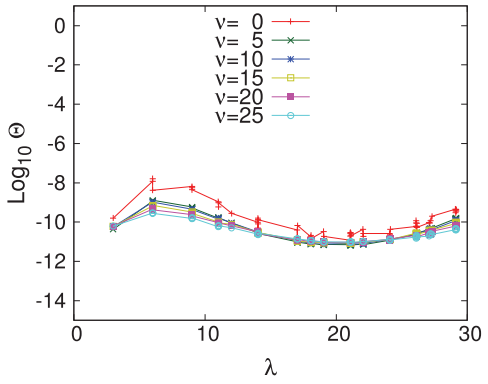


近似固有対の相対残差の大きさ Θ の対数

図 21: フィルタ F4-I-2 （$n = 40$, $\mu = 1.5$, g_s =1E-15, g_p =1E-2 ）

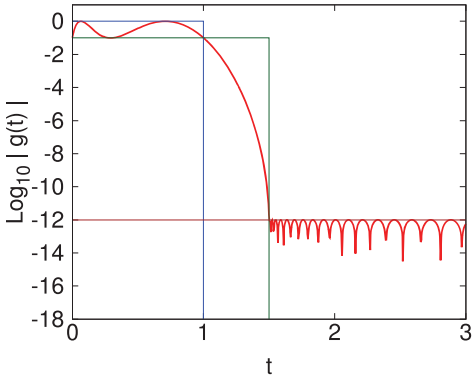


フィルタの伝達関数の大きさ $|g(t)|$ の対数

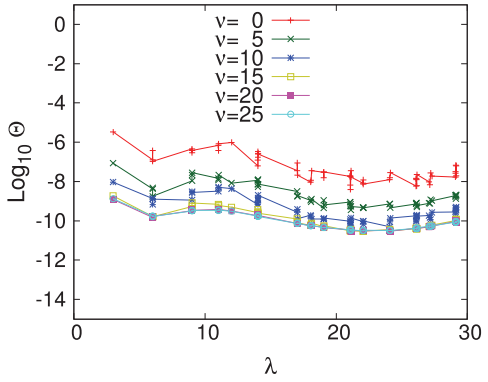


近似固有対の相対残差の大きさ Θ の対数

図 22: フィルタ F4-I-3 （$n = 50$, $\mu = 1.5$, g_s =1E-16, g_p =1E-2 ）



フィルタの伝達関数の大きさ $|g(t)|$ の対数



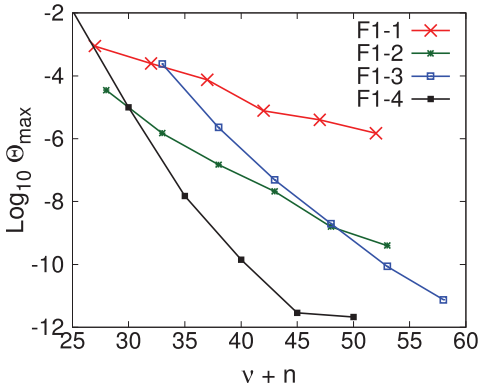
近似固有対の相対残差の大きさ Θ の対数

図 23: フィルタ F4-II-1 （$n = 63$, $\mu = 1.5$, g_s =1E-12, g_p =1E-1 ）

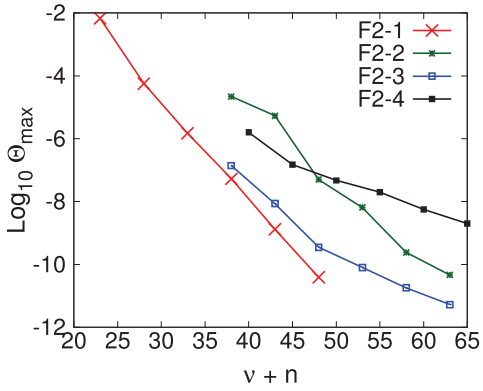
上記の各フィルタのどれについても，正しい数 54 個の近似固有対が得られた．そうして前処理用のフィルタの Chebyshev 多項式の次数である ν が増えるほど各近似固有対の相対残差の大きさは減少したことがわかる．

図 24 の 6 枚 1 組の図は，それぞれのフィルタを用いた前処理付きの実験について, 横軸に前処理用と本来のフィルタそれぞれの Chebyshev 多項式の次数の和 $\nu + n$ をとり縦軸に求めた近似固有対の相対残差の大きさの最大値を対数でプロットしたグラフである（これらは倍精度による計算である）．次数 ν は 0 から 25 まで 5 刻みでとっている（ただし $\nu = 0$ は前処理をまったく行わない場合である）．

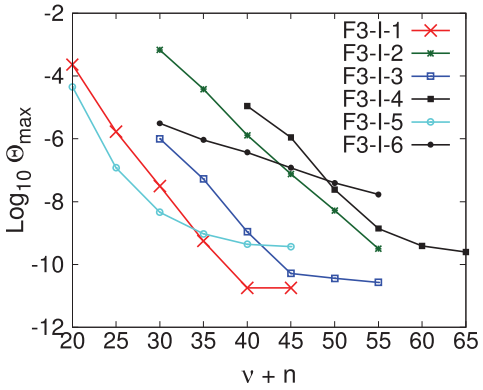
さらに図 25 の 6 枚 1 組の図は，四倍精度の数値と演算を用いて規模を小さくした問題を解いた結果について，同様にプロットをしたものである．解いている問題が異なるので直接の比較にはならないが，丸めによる数値誤差が小さい四倍精度を用いた計算の場合には片対数によりプロットされたグラフの折れ線は曲がらずまっすぐ延びていることが見てとれる．



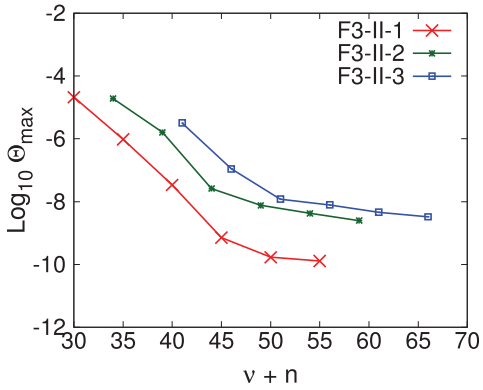
レゾルベントを 1 つ用いたフィルタ



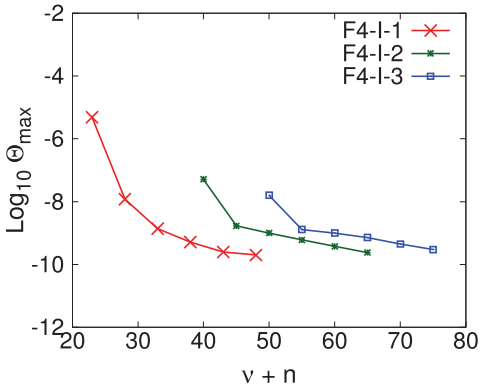
レゾルベントを 2 つ用いたフィルタ



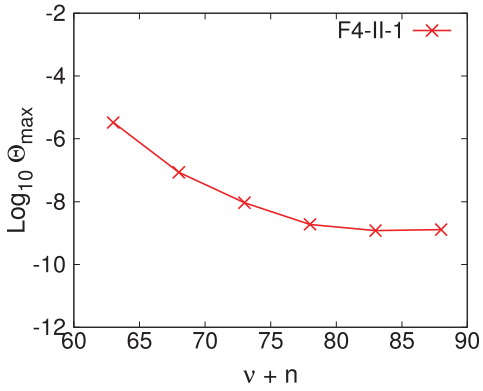
レゾルベントを 3 つ用いたフィルタ（方式 I）



レゾルベントを 3 つ用いたフィルタ（方式 II）

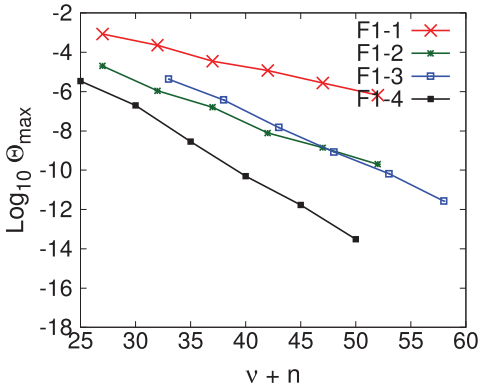


レゾルベントを 4 つ用いたフィルタ（方式 I）

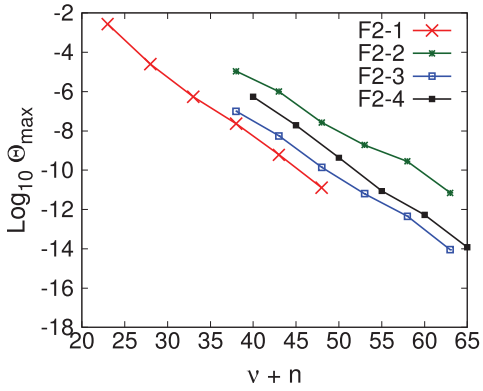


レゾルベントを 4 つ用いたフィルタ（方式 II）

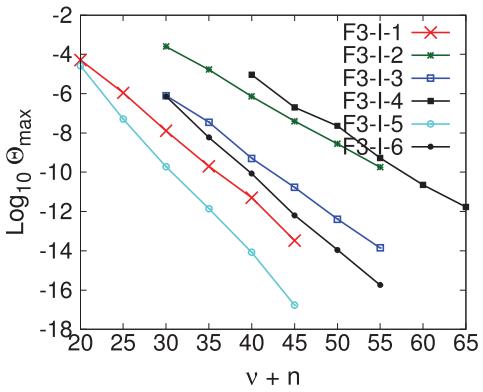
図 24: Chebyshev 多項式の次数の和 $\nu + n$ と相対残差の大きさの最大値（倍精度計算）（全 6 枚）



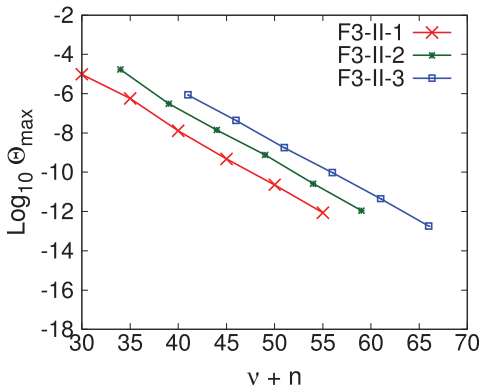
レゾルベントを 1 つ用いたフィルタ



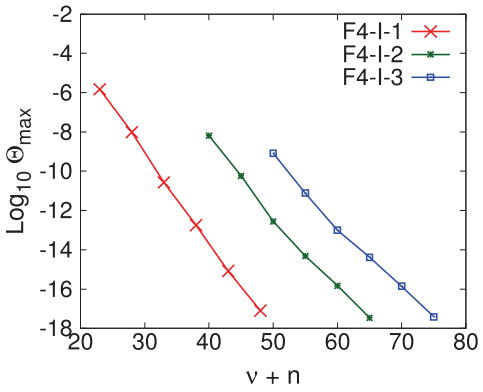
レゾルベントを 2 つ用いたフィルタ



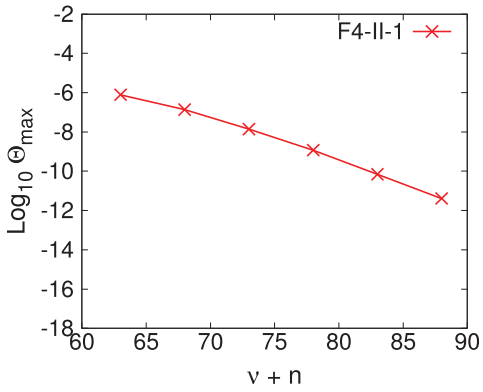
レゾルベントを 3 つ用いたフィルタ（方式 I）



レゾルベントを 3 つ用いたフィルタ（方式 II）



レゾルベントを 4 つ用いたフィルタ（方式 I）



レゾルベントを 4 つ用いたフィルタ（方式 II）

図 25: Chebyshev 多項式の次数の和 $\nu + n$ と相対残差の大きさの最大値（四倍精度計算，規模の小さい問題 $(N_1, N_2, N_3) = (30, 40, 50)$）（全 6 枚）

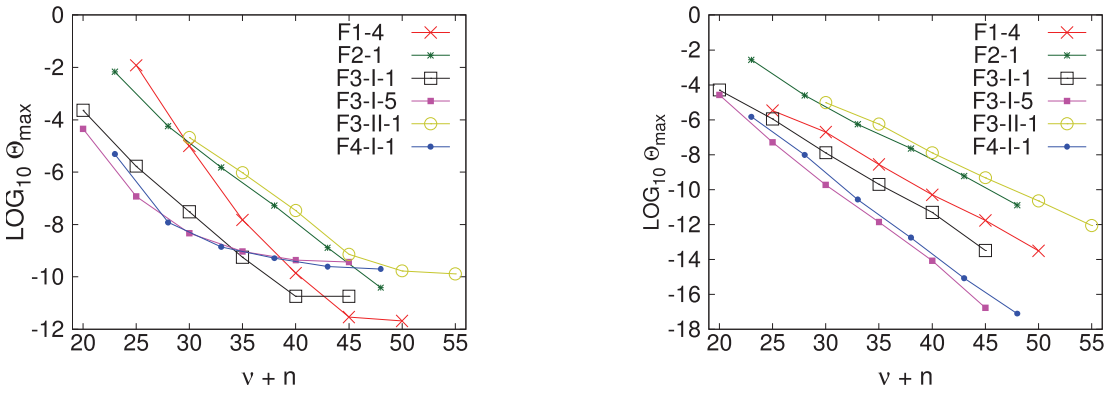


図 26: Chebyshev 多項式の次数の和 $\nu + n$ vs. 最大の相対残差の大きさ

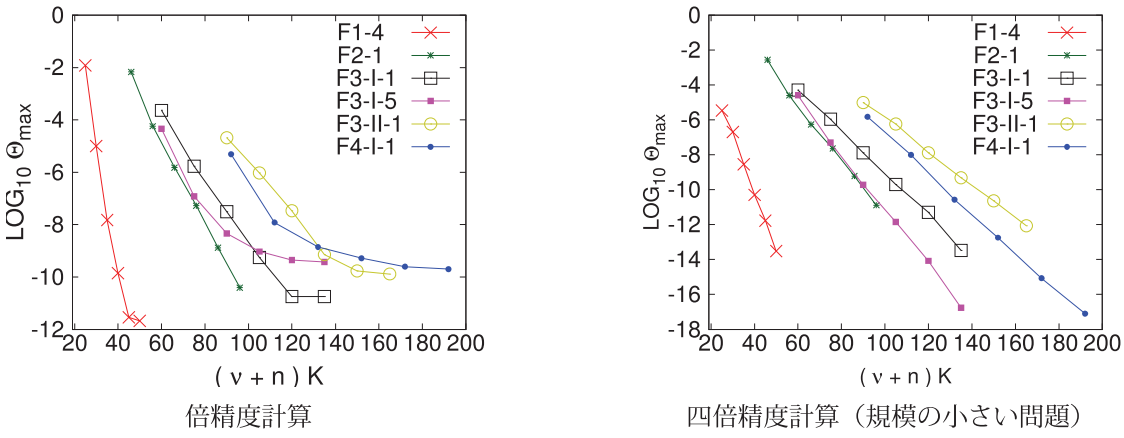


図 27: Chebyshev 多項式の次数の和 $\nu + n$ とレゾルベントの数 K の積 vs. 最大の相対残差の大きさ

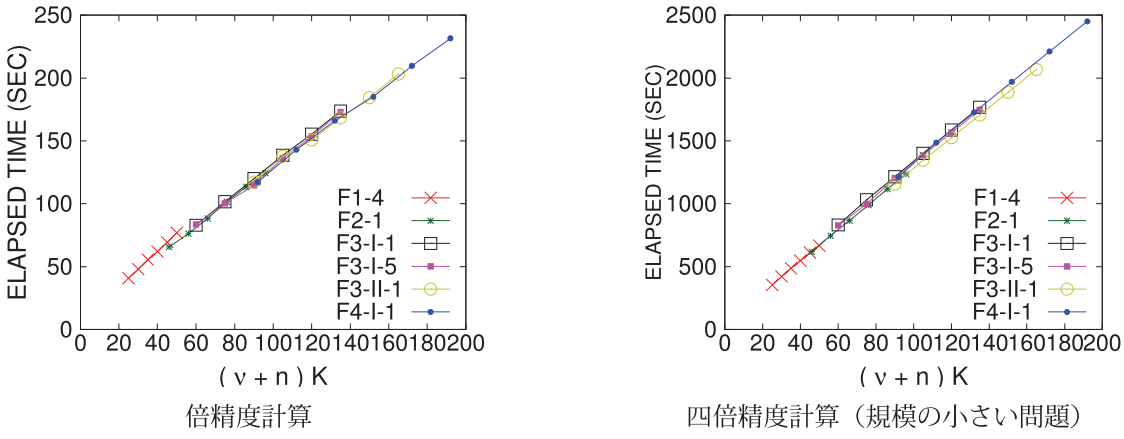


図 28: Chebyshev 多項式の次数の和 $\nu + n$ とレゾルベントの数 K の積 vs. 対角化完了までの経過時間

図 26 は，上記のそれぞれの数値精度の場合についての 6 枚組の各図のなかからグラフの折れ線が一番下側にあるものを性能の良いフィルタとして選出し，それら（F1-4，F2-1，F3-I-1，F3-I-5，F3-II-1，F4-I-1）のグラフを横軸には Chebyshev 多項式の次数の和 $\nu+n$ を縦軸には最大の相対残差の大きさ $\Theta_{\max}$ の対数値をとって，プロットしたものを 1 枚の図の中に集めたものである．右側の四倍精度を用いた場合の図では，片対数グラフ上の折れ線はほぼまっすぐに下降していて互いに交わらず，最も下側にあるのは F3-I-5 で，次は僅差で F4-I-1 となっている．左側の倍精度の場合の図では，数値丸め誤差の影響で折れ線が曲がっていてその曲がり方はフィルタごとに異なっている（レゾルベントの結合係数の大きさが大きいフィルタ，レゾルベントのシフトが近接しているフィルタは，数値相殺による桁落ちにより精度を失う傾向があり相対残差が速く停滞する），そのため順位を付けるのは簡単でなく，グラフが最も下側にあるのは $\nu+n$ が 35 以下のときには F3-I-5 で，次が僅差で F4-I-1 であり，$\nu+n$ が 35 と 40 のときには F3-I-1 であり，さらに $\nu+n$ が 45 と 50 のときには F1-4 である，という結果になった．ところで，横軸に次数の和 $\nu+n$ をとって比較することは，もしも K 個のレゾルベントの適用を完全に並行して処理をしている場合には，処理の経過時間が $\nu+n$ に比例するので意味を持ち得るが，今回の計算はレゾルベントの適用は 1 つずつ順に行っている．そのような場合には経過時間は $\nu+n$ ではなくて $(\nu+n)K$ に比例するので，そのようにグラフを描き直したものが図 27 である．この左右の 2 枚の図から読み取れることは，倍精度と四倍精度のどちらの場合も，効率が最も良いのは F1-4 であり，次にかなりの差が開いて F2-1 であることである．今回の計算全体の経過時間が実際に $(\nu+n)K$ にほぼ比例していることは，図 28 の左右の図内の横軸に $(\nu+n)K$ を縦軸に経過時間をとってプロットしたグラフを見て確認できる．その結果 K 個のレゾルベントの処理を逐次に行う場合には，今回使用したフィルタの中で効率が最も良いのはレゾルベントを 1 つだけ用いる F1-4 になる．このフィルタは前処理を行わない場合には，近似固有対の相対残差の大きさの最大値が 1E-2 程度であり精度が悪かったが，前処理での多項式の次数を $\nu=5$，$\nu=10$，$\nu=15$ などとするときそれに対応して相対残差の大きさの最大値は 1E-5，1E-8，1E-10 などとなり，工学分野での応用などの多くの用途にはほぼ満足できる精度になる．フィルタ F1-4 は $n=25$ であるから，たとえば前処理を $\nu=10$ で行うと，$n=25$ のフィルタを単純に 2 回適用するのに比べて計算の手間を減らせることになる．

5　現状でのまとめ

フィルタを適用するベクトルの組としてランダムなベクトルの組に B-正規直交化を施したものに対して Chebyshev 多項式の次数を下げたフィルタの適用と B-正規直交化を施す前処理を行うことで，固有対の近似精度を改良できる．得られる改良の程度は前処理に掛ける手間に応じたものになる．前処理用の本来の 2 つのフィルタの Chebyshev 多項式の次数の和 $\nu+n$ を一定とする場合に，次数の和の一部を前処理用の多項式の次数 ν に配分する方が前処理をしない場合 $\nu=0$ よりも良い結果が得られるであろう．

ベクトルの組に対する前処理を同じフィルタを 2 度繰り返すのではなくて，Chebyshev 多項式の次数を適度に小さく設定したフィルタを利用して行っても，近似固有対の相対残差の大きさの最大値を低減できる．実数シフトのレゾルベント 1 つだけからなるフィルタであっても，それにより得られる解の精度が応用からの要求を満たせるものであれば利用できて，計算に必要な記憶容量や演算量を少なく抑えられるであろう．

参　考　文　献

[1] 村上 弘：チェビシェフ多項式と極がすべて実数である低次有理関数の合成により実対称定値一般固有値問題の少数の下側固有対を解くためのフィルタの伝達関数を構成する方法, **情報処理学会論文誌コンピューティングシステム** *(ACS)*, Vol.15, No.3 (ACS78), pp.1–28 (2022).