数理解析研究所講究録
第1785巻 2012年 131-145

レゾルベントの線形結合によるフィルタの伝達特性の調整

村上 弘
HIROSHI MURAKAMI
首都大学東京 数理情報科学専攻
DEPARTMENT OF MATHEMATICS AND INFORMATION SCIENCES,
Tokyo Metropolitan University *

要約

係数行列 A, B が実対称で B が正定値である一般固有値問題 $A\mathbf{v}=\lambda B\mathbf{v}$ を近似的に解くために，固有値が指定された区間にある固有ベクトルだけを選択的に通過させるフィルタを利用する．フィルタとして複数のシフトを持つレゾルベントの線形結合を用いると，フィルタの伝達関数は各シフトを単純な極に持つ有理関数になる．指定区間の特性関数（区間内で1，区間外で0）をよく近似するように伝達関数の極の位置と極の係数をうまく選ぶ．固有値が固有値分布の下端付近にある固有対だけを求める場合に，特に各シフトを最小固有値よりも小さい実数にとると，レゾルベントの作用を実現するための連立一次方程式の係数行列は実対称正定値になる．

このフィルタの極の位置と極の係数を決める設計は，現状では数値的手法を用いてもある程度はできる．しかしまだ極の組の配置をどのようにすれば伝達関数の形状が最適になるかがよくわからず，不完全な状況である．今後はこの伝達関数の形状の最適化問題が例えばQEで解ける可能性にも期待したい．

1 フィルタにレゾルベントの線形結合を用いた対角化法

いま係数行列 A, B が実対称で B は正定値である一般固有値問題 $A\mathbf{v}=\lambda B\mathbf{v}$ を扱うことにする．この問題に対応するレゾルベントの定義を $\mathcal{R}(\tau)\equiv(A-\tau B)^{-1}B$ として，フィルタとして用いる作用素をレゾルベントの線形結合 $\mathcal{F}=c_\infty I+\sum_p\gamma_p\mathcal{R}(\tau_p)$ で固有値が実区間 $[a,b]$ から離れた固有ベクトルを強く減衰させるように適切に構成する．

いま乱数に基づいて作った m 個の縦ベクトルを B-正規直交化した組を $X=\{\mathbf{x}^{(1)},\mathbf{x}^{(2)},\ldots,\mathbf{x}^{(m)}\}$ とする．この X を入力ベクトルの組として，フィルタ $\mathcal{F}$ を作用させて出力ベクトルの組 $Y=\mathcal{F}X$ を作る．いま固有値が $[a,b]$ 近傍にある固有ベクトルで張られた不変部分空間を $\mathcal{S}$ とする．そのとき入力ベクトルの組 X，出力ベクトルの組 Y および使用したフィルタの伝達特性を参照して，基底 Z が張る空間 $\mathcal{S}'$ が不変部分空間 $\mathcal{S}$ の近似となるような B-正規直交基底 Z を Y の張る空間の内からうまく選び出す [9]．そのようにして得られた，近似不変部分空間 $\mathcal{S}'$ の基底 Z に Rayleigh-Ritz 法を適用して必要な固有対の近似を得る．

2 フィルタ作用素とその伝達関数

N 次の実対称定値一般固有値問題（A, B が実対称で B は正定値）：$A\mathbf{v}=\lambda B\mathbf{v}$ に対するレゾルベントを $\mathcal{R}(\tau)\equiv(A-\tau B)^{-1}B$ とする．フィルタ作用素をレゾルベントの線形結合 $\mathcal{F}=c_\infty I+\sum_p\gamma_p\mathcal{R}(\tau_p)$ であ

*mrkmhrsh@tmu.ac.jp

るとする．そのとき固有値がλである固有ベクトル$\mathbf{v}$に対しては$\mathcal{F}\mathbf{v} = f(\lambda)\cdot\mathbf{v}$が成り立つ（このことは，元の固有値問題の固有値が$\lambda$の固有ベクトルは同時にまた作用素$\mathcal{F}$の固有ベクトルにもなっていて，$\mathcal{F}$の固有値が$\phi = f(\lambda)$であることを示している）．ただしここで，$f(\lambda) = c_\infty + \sum_p \frac{\gamma_p}{\lambda-\tau_p}$である．有理関数$f(\lambda)$は$\mathcal{F}$による固有値$\lambda$の固有ベクトルの伝達率を与える伝達関数である．逆に上の形式で$f(\lambda)$を与えると，$\mathcal{F}$はレゾルベントの線形結合として実現できる．レゾルベントの個数，実係数c_∞，複素数の係数とシフト量γ_p，τ_pは，伝達関数$f(\lambda)$が関数値の振舞いに対する制約を満たすように設計する．

3 下端付近の固有値に対するフィルタの構成法

フィルタ$\mathcal{F}$が有界な作用素であることと，その伝達関数$f(\lambda)$が固有値と一致する極を持たないことは同値になる．実対称定値一般固有値問題の固有値はすべて実数なので，伝達関数の極がすべて虚数ならばフィルタ作用素は有界になる．しかし固有値と一致あるいは極端な接近をしていない実数の極があっても特に支障はない．たとえば伝達関数のすべての極が最小固有値よりも小さい実数であればフィルタ作用素は有界になる．

そこで固有値分布の下端付近の区間$[a,b]$を通過域とするフィルタを構成する（図1）．ただし，aは「最小固有値に対するある下界」であるとする．まず線形変換$\lambda = \mathcal{L}(t) = a + (b-a)\,t$により$\lambda \in [a,b]$を（正規座標）$t \in [0,1]$に対応させて，引数を$t$に変えた伝達関数$g(t) \equiv f(\lambda)$を定義する．すると$f(\lambda)$のすべての極が$a$よりも小さい実数であれば$g(t)$のすべての極は負の実数であるので，極を$-a_p$，$p$=1,2,...,$n$とする．そうして引数$t$の伝達関数$g(t) = c_\infty + \sum_{p=1}^{n} \frac{c_p}{t+a_p}$に含まれるパラメタ$c_\infty$，$a_p$，$c_p$を決定したとすると，関係式$g(t) = f(\lambda)$と$\lambda = \mathcal{L}(t) = a + (b-a)\,t$から引数$\lambda$の伝達関数$f(\lambda) = c_\infty + \sum_{p=1}^{n} \frac{\gamma_p}{\lambda-\tau_p}$の係数$\gamma_p$，極の位置$\tau_p$は$\gamma_p/c_p = \mathcal{L}' = (b-a)$，$\tau_p = \mathcal{L}(-a_p) = a - (b-a)a_p$となる．そうして$f(\lambda)$に対応してレゾルベントの線形結合であるフィルタが$\mathcal{F} = c_\infty I + \sum_{p=1}^{n} \gamma_p \mathcal{R}(\tau_p)$と決まる．

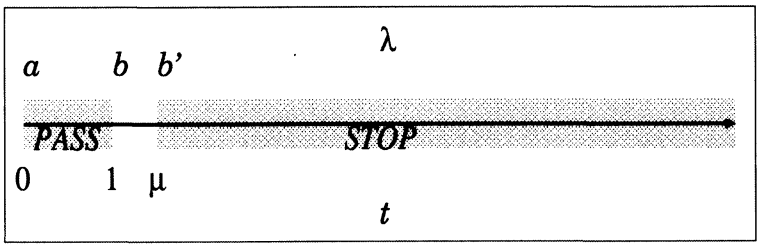

図 1: 固有値分布とフィルタの設定（下端付近の固有値に対する）
aは最小固有値の下界（最小固有値以下のある値）．

4 伝達関数の形状への制約

フィルタの伝達関数$g(t)$はn次の実有理関数であるとする．その形状についての制約は，まず$t \in [0,1]$では$g(t)$の値が1からなるべく離れず，またtが1より大きいところでは$|g(t)|$の値が非常に小さくなるようにする．そのように構成された$g(t)$の特性から形状パラメタg_{pass}，g_{stop}，μを決めると考える（図2）．

- 通過域（passband）は$t \in [0,1]$で，通過域では$g(t) \geq g_{\mathrm{pass}}$となり，またその逆も成立する．
- 阻止域（stopband）は$t \geq \mu$（> 1）で，阻止域では常に$|g(t)| \leq g_{\mathrm{stop}}$となる．

また，中間の$1 < t < \mu$を遷移域（transitionband）という．

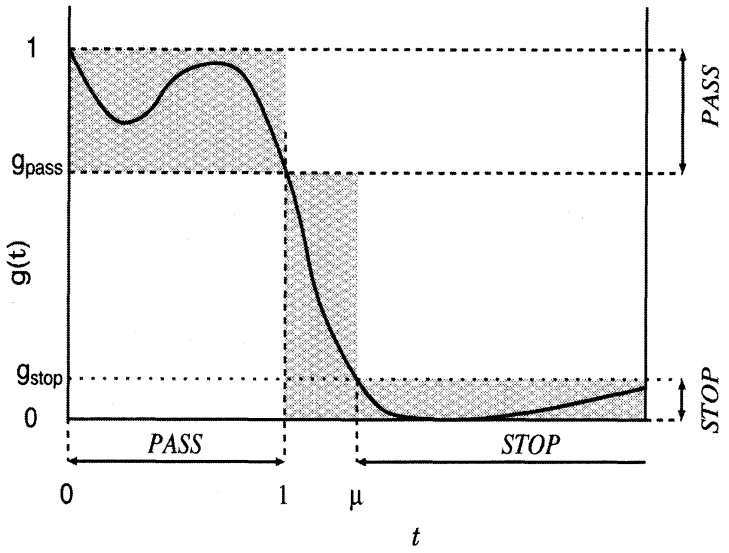

図 2: フィルタの伝達関数 $g(t)$ の概形図（下端付近の固有値に対する）

5 伝達関数 $g(t)$ の構成

フィルタがレゾルベントの線形結合ならば，$g(t)$ の極はすべて単純である．いま簡単のために定数項 c_∞ は 0 であるとして，$g(t) = \sum_{p=1}^{n} \frac{c_p}{t+a_p}$ とする．そうして n 個の負数 $-a_p$ と実係数 c_p，p=1, 2, ..., n を調節して，$t \in [0,1]$ では $g(t)$ の値は 1 付近で，遠方 $t > \nu(>1)$ では $|g(t)|$ の値が非常に小さくなるようにする．

いまのところ「シフト量 $-a_p$ の組」を最適に決めることはできていない．今回は，主に予め n 個のシフト量 $-a_p$ の組が与えられているとして c_p だけを最適に決定する方法に話を限定する．

5.1 構成法の例 I) 阻止域 $[\nu, \infty)$ での最小自乗法

ν を 1 よりも少し大きい値として，阻止域 $[\nu, \infty)$ での $g(t)$ の 2 乗積分を最小化することを考える．ただし，Lagrange の未定乗数法を用いて拘束条件 $g(0) = 1$ を課す．$2J = I^{[s]} + 2\lambda \cdot \{1 - \sum_{p=1}^{n}(c_p/a_p)\}$ $I^{[s]} \equiv \int_\nu^\infty \{g(t)\}^2 w^{[s]}(t)\, dt = \sum_{p,q} S_{p,q}\, c_p c_q$．ここで，$S_{p,q} \equiv \int_\nu^\infty \frac{1}{(t+a_p)(t+a_q)}\, w^{[s]}(t)\, dt$ であり，$w^{[s]}(t)$ は非負の積分の重み関数である．

最小点の必要条件を得るために J の偏微分を 0 とおくと $\partial J/\partial c_p = \sum_{q=1}^{n} S_{p,q}\, c_q - \lambda/a_p = 0$, p=1, 2, ..., n, $\partial J/\partial \lambda = 1 - \sum_{p=1}^{n}(c_p/a_p) = 0$ となる．つまり，$S\mathbf{c} = \lambda \mathbf{b}$, $\mathbf{b}^T\mathbf{c} = 1$．ただし $\mathbf{b}$ は第 p 要素が $1/a_p$ のベクトルである．これを解くにはまず線形方程式 $S\mathbf{x} = \mathbf{b}$ を解いて，$\mathbf{c} \Leftarrow 1/(\mathbf{b}^T\mathbf{x}) \cdot \mathbf{x}$ とする．対称行列 S を係数とする連立一次方程式を解く際には，S の極めて小さい固有値を切断することで，解 $\mathbf{c}$ のノルムの増大を抑制する．

5.2 構成法の例 II) 阻止域 $[\nu, \infty)$ と通過域 $[0, 1]$ での最小自乗法

いま $\nu > 1$ とし，ある微小な正定数 ω を用いて，最小化すべき目的関数の値を $2J = I^{[s]} + \omega\, I^{[p]}$ とする．ここで，$I^{[s]}$ は阻止域 $[\nu, \infty)$ での $\{g(t)\}^2$ の積分値で，$I^{[p]}$ は通過域 $[0, 1]$ での $\{1 - g(t)\}^2$ の積分値である．

$w^{[s]}(t)$, $w^{[p]}(t)$ は非負の積分の重み関数で, $\int_0^1 w^{[p]}(t)\,dt = 1$ とする.

$$I^{[s]} \equiv \int_\nu^\infty \{g(t)\}^2\, w^{[s]}(t)\,dt = \sum_{p,q} S^{[s]}_{p,q}\, c_p c_q\,.$$

$$I^{[p]} \equiv \int_0^1 \{1-g(t)\}^2\, w^{[p]}(t)\,dt = \sum_{p,q} S^{[p]}_{p,q}\, c_p c_q - 2\sum_p h_p c_p + 1\,.$$

ここで, $S^{[s]}_{p,q} \equiv \int_\nu^\infty \frac{1}{(t+a_p)(t+a_q)} w^{[s]}(t)dt$, $S^{[p]}_{p,q} \equiv \int_0^1 \frac{1}{(t+a_p)(t+a_q)} w^{[p]}(t)dt$, $h_p \equiv \int_0^1 \frac{1}{t+a_p} w^{[p]}(t)dt$.

$S = S^{[s]} + \omega S^{[p]}$ とおくとき, 最小点を与える条件は連立一次方程式 $S\mathbf{c} = \omega\,\mathbf{h}$ となり, これを解けば係数 c_p が求まる (必要ならば係数を再規格化して $g(0) = 1$ にすることができる). この連立一次方程式を解く場合にも, 対称行列 S の極めて小さい固有値を切断することで, 解 $\mathbf{c}$ のノルム増大を抑制する.

5.3 積分計算の例

阻止域に対応する行列 $S^{[s]}$ の定義は $\int_\nu^\infty \{g(t)\}^2 w^{[s]}(t)dt = \sum_{p,q} S^{[s]}_{p,q}\, c_p c_q$ である. ここで, $g(t) \equiv \sum_p \frac{c_p}{t+a_p}$ で, $w^{[s]}(t)$ は非負の積分の重み関数であり, 要素 $S^{[s]}_{p,q}$ は:

$$S^{[s]}_{p,q} \equiv \int_\nu^\infty \frac{1}{t+a_p}\,\frac{1}{t+a_q}\, w^{[s]}(t)\,dt = \int_1^\infty \frac{1}{s+\alpha}\,\frac{1}{s+\beta}\,\frac{1}{\nu} w^{[s]}(\nu s)\,ds$$

で与えられる. ただし $\alpha = a_p/\nu$, $\beta = a_q/\nu$ とおいた. 以下に, 重み関数 $w^{[s]}(t)$ の選択の 8 通りの例と, それに対応する $S^{[s]}$ の行列要素の式を掲げる (最初のものと第 6 番目のものが良い性質を持つようである).

1. $w^{[s]}(t) = 1$ の場合 ($a_p \neq a_q$ と仮定する. 以下同様):

$$\nu S^{[s]}_{p,p} = \frac{1}{1+\alpha},\quad \nu S^{[s]}_{p,q} = \frac{1}{\alpha-\beta}\log\left(\frac{1+\alpha}{1+\beta}\right).$$

2. $w^{[s]}(t) = \frac{1}{t}$ の場合:

$$\nu^2 S^{[s]}_{p,p} = \frac{\log(1+\alpha)}{\alpha^2} - \frac{1}{\alpha(1+\alpha)},\quad \nu^2 S^{[s]}_{p,q} = \frac{-1}{\alpha-\beta}\left\{\frac{\log(1+\alpha)}{\alpha} - \frac{\log(1+\beta)}{\beta}\right\}.$$

3. $w^{[s]}(t) = \sqrt{t}$ の場合:

$$\sqrt{\nu} S^{[s]}_{p,p} = \frac{1}{1+\alpha} + \frac{\arctan(\sqrt{\alpha})}{\sqrt{\alpha}},\quad \sqrt{\nu} S^{[s]}_{p,q} = \frac{2}{\alpha-\beta}\left\{\sqrt{\alpha}\arctan(\sqrt{\alpha}) - \sqrt{\beta}\arctan(\sqrt{\beta})\right\}.$$

4. $w^{[s]}(t) = \frac{\nu}{t^2}$ の場合:

$$\nu^2 S^{[s]}_{p,p} = \frac{1}{\alpha^2}\left\{1 + \frac{1}{1+\alpha} - \frac{2\log(1+\alpha)}{\alpha}\right\},\quad \nu^2 S^{[s]}_{p,q} = \frac{1}{\alpha\beta} + \frac{1}{\alpha-\beta}\left\{\frac{\log(1+\alpha)}{\alpha^2} - \frac{\log(1+\beta)}{\beta^2}\right\}.$$

5. $w^{[s]}(t) = \lambda e^{-\lambda(t-\nu)}$ の場合:

$$\nu^2 S^{[s]}_{p,p} = \frac{r}{1+\alpha} - r^2 e^{r(1+\alpha)} E_1(r(1+\alpha)),\quad \nu^2 S^{[s]}_{p,q} = \frac{-r}{\alpha-\beta}\left\{e^{r(1+\alpha)} E_1(r(1+\alpha)) - e^{r(1+\beta)} E_1(r(1+\beta))\right\}.$$

ここで, $E_1(x) \equiv \int_x^\infty (e^{-t}/t)\,dt$, $r \equiv \lambda\nu$.

6. $w^{[s]}(t)=\frac{\sqrt{\nu}}{\pi}\frac{1}{t\sqrt{t-\nu}}$ の場合：

$$\nu^2 S^{[s]}_{p,p}=\frac{1+\frac{1}{2\sqrt{1+\alpha}}}{(1+\alpha+\sqrt{1+\alpha})^2},\quad \nu^2 S^{[s]}_{p,q}=\frac{1+\frac{1}{\sqrt{1+\alpha}+\sqrt{1+\beta}}}{(1+\alpha+\sqrt{1+\alpha})(1+\beta+\sqrt{1+\beta})}.$$

7. $w^{[s]}(t)=\frac{\nu}{2}\frac{1}{t\sqrt{t(t-\nu)}}$ の場合：

$$\nu^2 S^{[s]}_{p,p}=\frac{1}{\alpha^2}+\frac{1}{2\alpha^2(1+\alpha)}\times\left\{1-\frac{(\frac{3}{2}+2\alpha)\log\{1+2\alpha+2\sqrt{\alpha(1+\alpha)}\}}{\sqrt{\alpha(1+\alpha)}}\right\},$$
$$\nu^2 S^{[s]}_{p,q}=\frac{1}{\alpha\beta}+\frac{1}{2(\alpha-\beta)}\left\{\frac{\log\{1+2\alpha+2\sqrt{\alpha(1+\alpha)}\}}{\alpha\sqrt{\alpha(1+\alpha)}}-\frac{\log\{1+2\beta+2\sqrt{\beta(1+\beta)}\}}{\beta\sqrt{\beta(1+\beta)}}\right\}.$$

8. $w^{[s]}(t)=\frac{8\nu^{3/2}}{\pi}\frac{\sqrt{t-\nu}}{t^3}$ の場合：

$$\frac{\nu^2}{4}S^{[s]}_{p,p}=\frac{6+5\alpha}{\alpha^4\sqrt{1+\alpha}}+\frac{1}{4\alpha^2}-\frac{2}{\alpha^3}-\frac{6}{\alpha^4},$$
$$\frac{\nu^2}{4}S^{[s]}_{p,q}=\frac{\frac{2}{\alpha^2\beta^2}}{\frac{\sqrt{1+\alpha}}{\alpha^3}+\frac{\sqrt{1+\beta}}{\beta^3}}\left[(\alpha+\beta)\left\{\frac{1+\alpha}{\alpha^4}+\frac{1+\beta}{\beta^4}+\frac{1}{\alpha^2\beta^2}\right\}+\frac{1}{\alpha\beta}\right]+\frac{1}{4\alpha\beta}-\frac{\alpha+\beta}{\alpha^2\beta^2}-\frac{2(\alpha^2+\alpha\beta+\beta^2)}{\alpha^3\beta^3}.$$

同様に，通過域に対応する行列 $S^{[p]}$ とベクトル $\mathbf{h}$ の定義は $\int_0^1\{1-g(t)\}^2 w^{[p]}(t)dt=\sum_{p,q}S^{[p]}_{p,q}c_pc_q-2\sum_p h_pc_p+1$ である．ここで，$g(t)\equiv\sum_p\frac{c_p}{t+a_p}$ で，$w^{[p]}(t)$ は積分の重み関数である．以下に，重み関数 $w^{[p]}(t)$ の選択の 2 通りの例と，それに対応する行列 $S^{[p]}$ とベクトル $\mathbf{h}$ の要素の式を掲げておく．

1. $w^{[p]}(t)=1$ の場合（$a_p\neq a_q$ と仮定する．以下同様）：

$$S^{[p]}_{p,p}=\frac{1}{a_p(1+a_p)},\quad S^{[p]}_{p,q}=\frac{-1}{a_p-a_q}\log\left(\frac{1+\frac{1}{a_p}}{1+\frac{1}{a_q}}\right);\quad h_p=\log\left(1+\frac{1}{a_p}\right).$$

2. $w^{[p]}(t)=\frac{1}{\pi}\frac{1}{\sqrt{t(1-t)}}$ の場合：

$$S^{[p]}_{p,p}=\frac{\frac{1}{2}+a_p}{a_p(1+a_p)\sqrt{a_p(1+a_p)}},\quad S^{[p]}_{p,q}=\frac{1+a_p+a_q}{a_p(1+a_p)\sqrt{a_q(1+a_q)}+a_q(1+a_q)\sqrt{a_p(1+a_p)}};$$
$$h_p=\frac{1}{\sqrt{a_p(1+a_p)}}.$$

数式処理システムによる初等関数の積分には，現状ではまだ不満足な点がいろいろある．特に関数がパラメタを含むときに積分結果の式が（実際には初等関数の範囲で存在しても）求められないことがよくある．あるいは，$w(x)$ を具体的な関数で与えたときに，不定積分あるいは定積分として $\int\frac{1}{x+\alpha}w(x)\,dx$ が数式としてうまく求められる場合であっても，なぜか $\int\frac{1}{(x+\alpha)(x+\beta)}w(x)\,dx$ や $\int\frac{1}{(x+\alpha)^2}w(x)\,dx$ が求められないことがある（前者は恒等式 $\frac{1}{(x+\alpha)(x+\beta)}=\frac{-1}{\alpha-\beta}(\frac{1}{x+\alpha}-\frac{1}{x+\beta})$ を用いて，後者は α に関する微分と積分の順序交換を用いればよい）．また，不定積分が複素解析関数の意味で正しく求められた場合でも，それが分岐点を持つ多価関数である場合には数式の表現として，実区間を積分路とする定積分の値を求めるためには不適切であったり，または実関数として値を求めるためには不都合なものである場合が多い．今回の上記の例でもシステムの出力した数式に対して，人手による整理や簡易化などの作業が必要であった．積分 $\int\frac{1}{x+\alpha}w(x)\,dx$ の式は数式処理システムで求められたのに，$\int\frac{1}{(x+\alpha)(x+\beta)}w(x)\,dx$ や $\int\frac{1}{(x+\alpha)^2}w(x)\,dx$ が求められなかった場合にも，人間の作業が必要であった．上記の数式による定積分の計算例に対して最も有効かつ便利に使えた数式処理システムが Mathematica や Maple ではなく古典的な Maxima であったことは若干驚きであった．

6 シフトが実数だけからなるフィルタの作成例

以下で構成する正規化座標 t でのフィルタの伝達関数は，定数項 c_∞ を含まない形，$g(t)=\sum_{p=1}^{n}\frac{c_p}{t+a_p}$ とする．次数は $n=16$ として，シフト量を $-a_p=-3(1+z_p)$，p=1,2,...,n と設定する．但し，z_p は $[-1,1]$ を直交区間とする n 次の第一種 Chebyshev 多項式 $T_n(z)$ の零点である（注：このシフト量の値の組の設定はこの時点でのとりあえずのものであって，より良い設定が可能であろう）．

6.1 フィルタ（その 1）

いま $\nu=3$ として，制約条件 $g(0)=1$ 付きの積分 $J=\int_\nu^\infty\{g(t)\}^2\,dt$ の値の最小化から係数 c_p を求めた．係数の絶対値の最大値は 1.63 であった．この係数 c_p を求める計算には四倍精度演算を用いた．構成したフィルタのシフト量 $-a_p$ とそれに対応する線形結合の係数 c_p を表 1 に掲げる．得られたフィルタの特性は，$\mu=3$ とすると，$g_{\rm pass}$ =1.07E-6，$g_{\rm stop}$ =1.16E-14 となった．得られたフィルタの伝達関数の絶対値 $|g(t)|$ の対数プロットを図 3 に示す．

6.2 フィルタ（その 2）

いま $\nu=3$ として，$J=\int_\nu^\infty\{g(t)\}^2\,dt+\omega\int_0^1\{1-g(t)\}^2\,dt$ を目的関数とした．ただし $\omega=10^{-18}$ である．J の最小化条件から c_p についての対称行列を係数とする連立一次方程式が得られる．その行列の 10^{-23} 以下の固有値を切断して c_p を求めた．係数 c_p の絶対値の最大値は 30.3 であった．この係数 c_p を求める計算には四倍精度演算を用いた．構成したフィルタのシフト量 $-a_p$ と線形結合の係数 c_p を表 2 に掲げる．得られたフィルタの特性は，$\mu=3$ とすると，$g_{\rm pass}$ =4.25E-4，$g_{\rm stop}$ =4.72E-10 となった．得られたフィルタの伝達関数の絶対値 $|g(t)|$ の対数プロットを図 4 に示す．

6.3 若干の考察

実数のシフト τ を最小固有値よりも小さくとると，レゾルベントを実現する連立一次方程式の係数 $C=A-\tau B$ は実対称正定値になり，疎行列用の算法やライブラリの選択の面からは利点がある．また A，B が帯行列であれば C も帯行列となり，連立一次方程式は係数 C が正定値なので，ピボット選択をしない帯用の修正 Cholesky 法を用いて数値的に安定に解ける．ただしシフトは最小固有値よりも小さい実数値に制限されていて遷移域の付近にはないため（シフトに複素数を許した場合に比べると）伝達関数の遷移域の幅は狭くできない．

いま $g(t)$ の最大値は 1 に規格化されているとする．値 $g_{\rm stop}$ は丸め誤差単位（マシンイプシロン）程度あるいはそれ以下の小さい値であることが理想である．値 $g_{\rm pass}$ が 1 から離れて小さい（通過域での伝達率の一様性が悪い）と近似固有対の精度が下がる．なぜならば，必要な固有値を持つ固有ベクトルの組の間の相対的なノルムの比（一種の条件数）がフィルタを通すことで $1/g_{\rm pass}$ 倍に拡大されるので丸め誤差の影響も大きくなり，拡大率の分だけ伝達率の小さい固有ベクトルの有効精度が失なわれるからである．また $g_{\rm stop}/g_{\rm pass}$ の値が十分に小さくないと，不要な固有ベクトルの混入率が十分に小さいことが保証されないので，近似固有対の精度が低下する．また遷移域の幅が広いとそれだけ遷移域に含まれる固有値の個数も多くなる可能性があり，その個数の分だけフィルタで濾過するベクトルの個数を多くする必要が出てくる．

表 1: フィルタ（その 1）の正規化座標 t でのシフト量 $-a_p$ と結合係数 c_p

p	$-a_p$	c_p
1	-5.9855541800165907	-0.70679914619474910
2	-5.8708210071966266	1.3067487312909856
3	-5.6457637930450651	-0.43083970283861383
4	-5.3190313600882109	-0.92035417653569363
5	-4.9031798524909365	1.3618472707060915
6	-4.4141902104779929	-0.57891735108368220
7	-3.8708540317633871	-0.65897102634236280
8	-3.2940514209886818	1.4881281890531120
9	-2.7059485790113182	-1.6319872643458175
10	-2.1291459682366129	1.3177569441745646
11	-1.5858097895220071	-0.87954171110468911
12	-1.0968201475090635	0.51664763945141700
13	-0.68096863991178912	-0.27885684304419086
14	-0.35423620695493491	0.14145585974181497
15	-0.12917899280337341	-0.065074410781865841
16	-0.014445819983409341	0.018756997853631990

表 2: フィルタ（その 2）の正規化座標 t でのシフト量 $-a_p$ と結合係数 c_p

p	$-a_p$	c_p
1	-5.9855541800165907	13.045045787111767
2	-5.8708210071966266	-12.377137100549698
3	-5.6457637930450651	-13.731924800311170
4	-5.3190313600882109	12.923686673412777
5	-4.9031798524909365	13.294980769536150
6	-4.4141902104779929	-16.095808160252892
7	-3.8708540317633871	-9.2616885206944389
8	-3.2940514209886818	21.908177036026508
9	-2.7059485790113182	-4.6095492006766534
10	-2.1291459682366129	-20.378505869411738
11	-1.5858097895220071	30.319364279617194
12	-1.0968201475090635	-24.380824341765085
13	-0.68096863991178912	13.308061318273295
14	-0.35423620695493491	-4.9584481056814305
15	-0.12917899280337341	1.0151514592585018
16	-0.014445819983409341	-0.020581221997454257

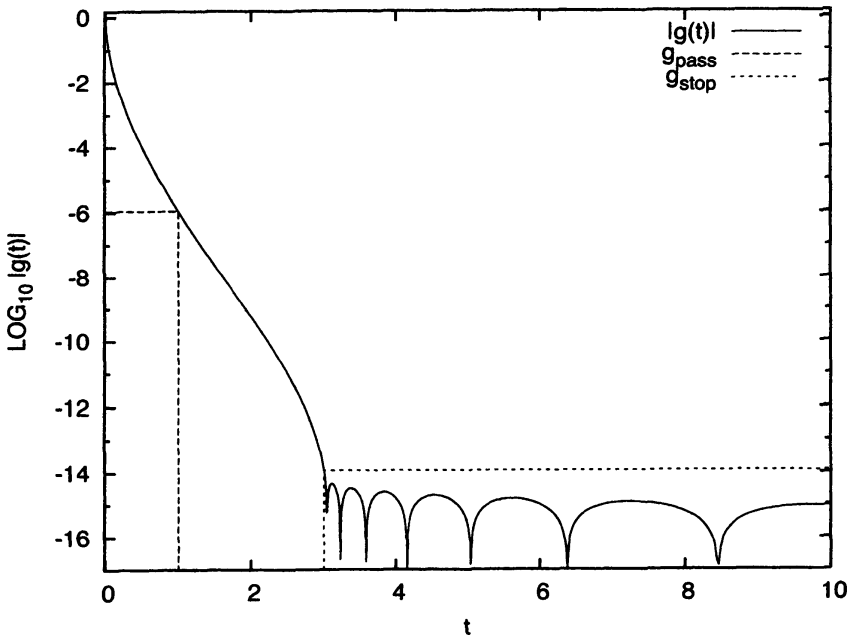

図 3: フィルタ（その 1）の伝達関数 $|g(t)|$（対数値）（$n = 16$, $\mu = 3$, g_{pass} =1.07E-6, g_{stop} =1.16E-14）

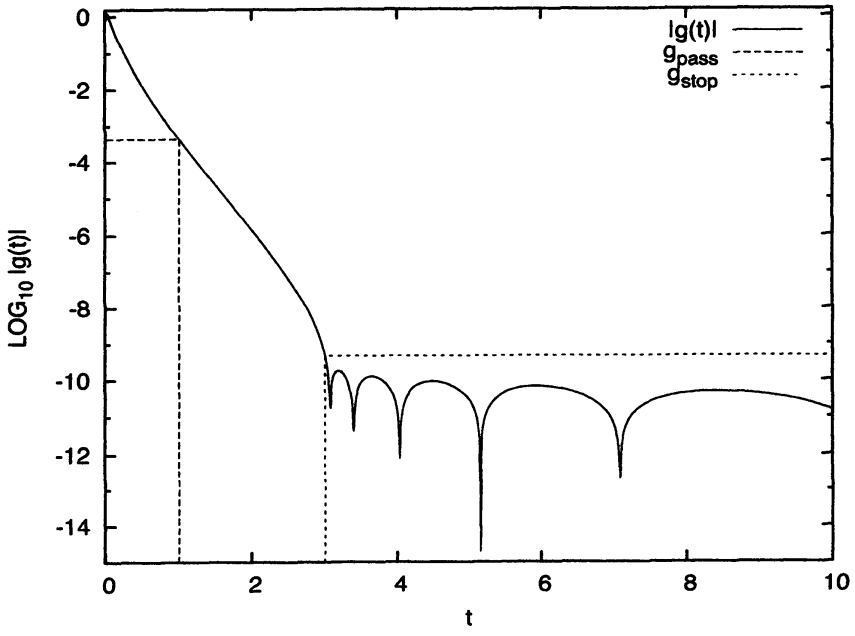

図 4: フィルタ（その 2）の伝達関数 $|g(t)|$（対数値）（$n = 16$, $\mu = 3$, g_{pass} =4.2E-4, g_{stop} =4.7E-10）

7 固有対を求める実験

7.1 例題の一般固有値問題（三次元 FEM）

以下で扱う例題は，立方体領域 $[0,\pi]\times[0,\pi]\times[0,\pi]$ 上で，零 Dirichlet 境界条件を課した Laplacian の固有値問題：

$$-(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2})\,\Psi(x,y,z)=\lambda\,\Psi(x,y,z)\,.$$

を有限要素法（FEM）で離散化した，行列の（実対称定値）一般固有値問題である．

有限要素は各軸方向の辺 $[0,\pi]$ を 26 等分した区間の直積で，要素内の基底関数は各軸方向の区分線形関数の積（三重線形関数）を用いた．この有限要素法の離散化による一般固有値問題の係数行列 A，B の次数は $N=25\times25\times25=15{,}625$ で，半帯幅は（基底関数の番号を適切に付けると）$25^2+25+1=651$ となる．

有限要素法により零 Dirichlet 境界条件を課した Laplacian の固有値問題を離散化して得られる一般固有値問題の固有値は正の実数である．そこで区間 $[0,30]$ に固有値がある固有対を求めてみることにする．ちなみに今の問題設定では，すべての固有値の厳密値は明示的な数式により計算できる．

実験に使用した計算機システムの CPU は intel Core-i7 2600K（3.4GHz，4 コア，8M バイト L3 共有キャッシュ，Turbo モード 3.8GHz，Hyperthread オフ）で，メインメモリは Dual Channel 動作で 8G バイトのモジュール（DDR3-1333MHz, PC3-10600 規格）を 4 本で計 32G バイトであり，OS は Fedora 15 である．以下の実験の計算に用いた浮動小数点数とその演算はすべて IEEE-754 の 64 ビット倍精度である．

7.2 例題 1：フィルタ（その 1）による実験

使用したフィルタはシフトが実数のレゾルベントの線形結合（その 1）で，シフトの個数は $n=16$ で，$\mu=3$，$g_{\rm pass}$ =1.06E-6，$g_{\rm stop}$ =1.16E-14 である．

乱数で生成した $m=300$ 個の B-正規直交ベクトルの組 X をフィルタへ入力し，フィルタの出力の組 Y から不変部分空間を近似する基底 Z を構成し，Z を Rayleigh-Ritz 法に与えて近似固有対を求めた実験の例を示す．マシンイプシロンの 100 倍の閾値 2.2E-14 で $\beta\equiv X^TBY$ の固有値を切断するとその階数は 261 となったので，入力ベクトルの個数が $m=300$ で一応飽和に達しているとみなせる．

グラフ（図 5）にフィルタ作用素 $\mathcal{F}$ の近似固有値 ϕ の分布を減少順に示す．54 番目までが通過域に対応する．ϕ が $g_{\rm pass}$ =1.066E-6 以上のものを集めると，真の固有対の個数と同じ 54 個の基底が得られる．基底を 54 個にとった場合には近似対の固有値はすべて $[0,30]$ にあった．この実験では安全のために余裕をとって閾値を半分の値に下げて 5.33E-7 としたので基底は 60 個となり，6 重に縮退した固有値 30.79 を持つ不要な近似対も求まる．

図 6 のグラフに，フィルタ対角化法だけで得られた近似対のうちで固有値が $[0,30]$ にあるものだけを折れ線 IT0 で，横軸に固有値，縦軸に残差のノルム Δ の値をとってプロットした（近似対 $(\lambda,\mathbf{v})$ のベクトルは B-正規直交条件で規格化されている．そのとき，固有対の残差ベクトル $\mathbf{r}=A\mathbf{v}-\lambda B\mathbf{v}$ のノルムを $\Delta=\sqrt{\mathbf{r}^TB^{-1}\mathbf{r}}$ と定義した）．このグラフからフィルタ対角化法による近似対の残差のノルムの大きさは 1E-8 から 1E-4 程度であること，そのことから近似固有値の相対精度は少なくとも 4 桁から 5 桁程度はあることがわかる（後に示すように，実際の精度はもっと高い）．同じグラフ中の折れ線 IT1 および IT2 は，それぞれフィルタ対角化法で得られた固有対を Rayleigh 商同時逆反復法で改良した固有対についてプロットしたものである．

この問題では行列の一般固有値問題の厳密な固有値が数式から計算できる．そこで図 7 のグラフに，フィルタ対角化法だけで求めた近似固有値の真値からの誤差の絶対値をプロットした．絶対誤差の最大値は 7.51E-10，相対誤差の最大値は 2.55E-11 で，近似固有値の精度は 10 桁から 11 桁程度であった．

OpenMP を用いた 4 スレッド計算による経過時間は，フィルタ対角化の部分が 27.1 秒，Rayleigh 商逆反復を余分な固有対も含めた 60 個すべてに対して 2 回ずつ施すのには 1,443.1 秒であった（表 3）.

7.3 例題 2：フィルタ（その 2）による実験

使用したフィルタはシフトが実数のレゾルベントの線形結合（その 2）で，シフトの個数は $n = 16$ として，$\mu = 3$，g_{pass} =4.25E-4，g_{stop} =4.72E-10 とした.

乱数を用いて生成した $m = 300$ 個の B-正規直交ベクトルの組 X をフィルタに入力し，フィルタの出力の組 Y から不変部分空間の近似基底 Z を構成して Rayleigh-Ritz 法で近似固有対を求めた実験の例を示す．マシンイプシロンの 100 倍の閾値 2.2E-14 で $\beta \equiv X^T BY$ の固有値を切断すると階数は 297 となったので，入力ベクトルの個数が $m = 300$ でちょうど飽和に達した程度であるとみなせる.

図 8 のグラフに，フィルタ作用素 $\mathcal{F}$ の近似固有値 ϕ の分布を減少順に示す．54 番目までが通過域に対応する．ϕ が g_{pass} =4.25E-4 以上のものは，真の固有対の個数と同じ 54 個であった．基底を 54 個にとった場合は近似対の固有値はすべて $[0, 30]$ にあった．この実験では安全のために余裕をとって閾値を半分の値 2.12E-4 に下げたので基底は 60 個となり，6 重に縮退した不要な固有値 30.79 を持つ近似対も求まる.

図 9 のグラフに，フィルタ対角化法により得られた近似固有対のうちで固有値が $[0, 30]$ にあるものだけを折れ線 IT0 で，横軸に固有値，縦軸に残差のノルム Δ の値をとってプロットした．このグラフからフィルタ対角化法による近似対の残差のノルムの大きさは 10^{-6} から 10^{-3} 程度であること，そのことから近似固有値の相対精度は少なくとも 4 桁程度あることがわかる．同じグラフ中の折れ線 IT1 および IT2 は，フィルタ対角化法で得られた固有対を Rayleigh 商同時逆反復法で改良した固有対についてそれぞれプロットしたものである.

図 10 のグラフに，フィルタ対角化だけで求めた近似固有値の，真値からの誤差の絶対値をプロットした．絶対誤差の最大値は 2.75E-8 で，相対誤差の最大値は 9.37E-10 であり，近似固有値の精度は 9 桁から 10 桁程度であった.

OpenMP を用いた 4 スレッド計算による経過時間は，フィルタ対角化の部分に 26.9 秒，Rayleigh 商逆反復を余分な固有対も含めた 60 個すべてに対して 2 回ずつ施すのには 1,443.4 秒であった（表 4）.

8．まとめ

実対称定値一般固有値問題に対して，a を最小固有値以下のある実数とするとき，通過域 $[a, b]$ のフィルタをシフトが実数のレゾルベントの線形結合で構成する方法の例として，積分で定義された目的関数の最小化を用いる方法を示した．但し，良い特性を持つフィルタの構成法はまだ未完成で，今後まだ追求する必要がある．良い実数シフトの組を決定することが特に重要な検討課題である.

例題として，「立方体領域での Laplacian の固有値問題」を FEM で離散化して得られる「行列の一般固有値問題」を解いてみた．下端付近の固有値とその固有ベクトルを今回の「実数シフトだけのフィルタ」を用いて求めてみて，一応有望な結果を得た.

今回の，シフトを固有値分布の下界よりも小さい実数に限定したフィルタは伝達関数の遷移域の幅を狭くできず，また通過域での伝達率の値の変動も大きい．そのため，複素数のシフトを任意に選べる場合と比べると，フィルタの出力したベクトルの組から近似固有対を良い精度で抽出することは（特に素朴な Rayleigh-Ritz 法では）かなり難しいが，それでも「フィルタの入力 X と出力 Y の組と伝達関数の形状の情報から，不変部分空間 $\mathcal{S}$ の良い近似 $\mathcal{S}'$ の B-正規直交基底 Z を構成する方法 [9]」を用いたので，ある程度の精度を持つ解が得られた.

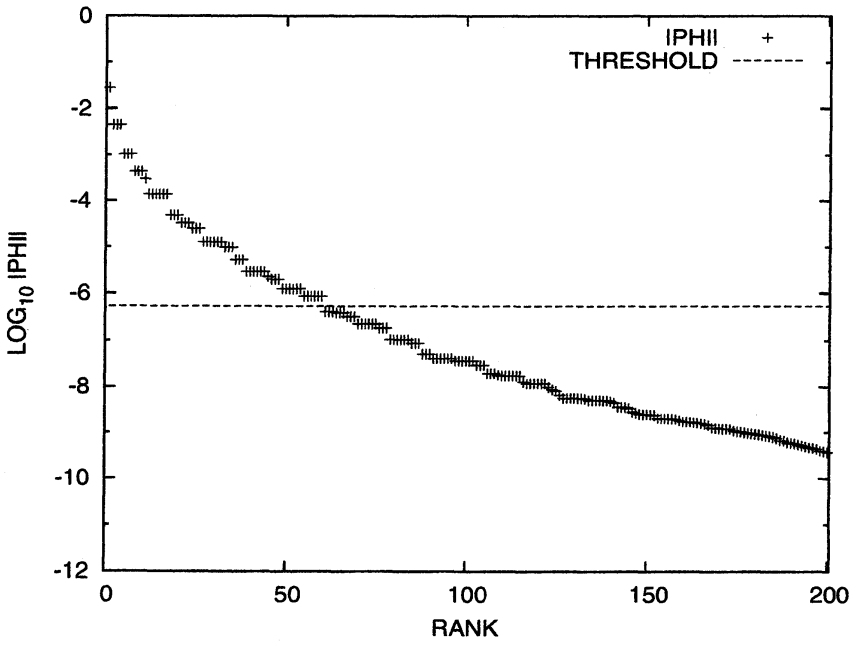

図 5: 例題 1：$|\phi|$ の値の分布（m=300, 閾値 5.33E-7）

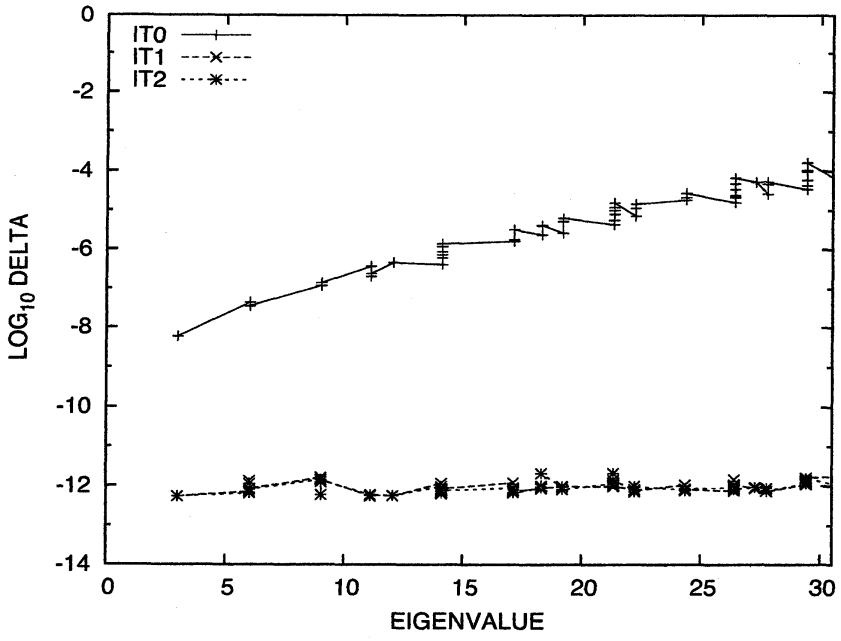

図 6: 例題 1：近似対の残差のノルム Δ（区間 $[0,30]$, m=300）

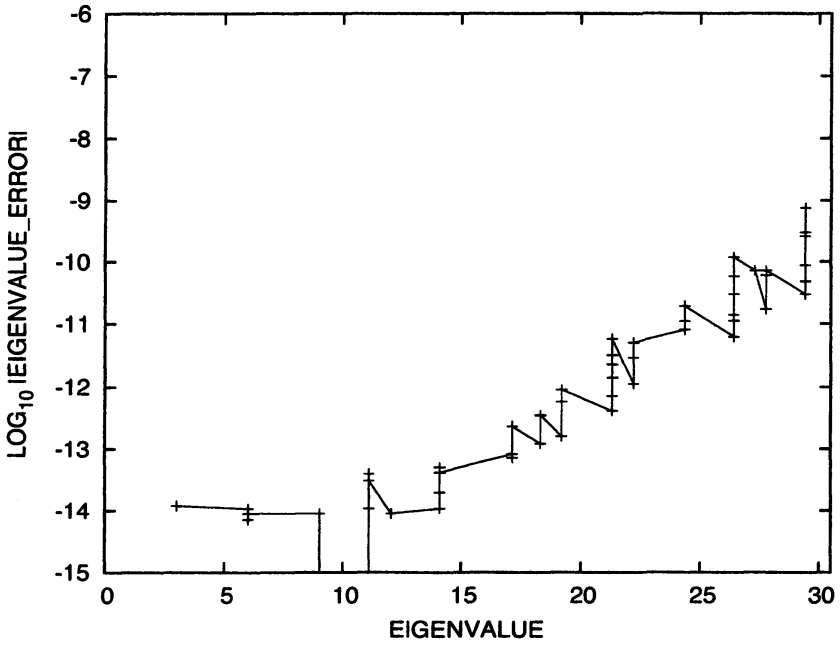

図 7: 例題 1：近似固有値の真値からの誤差の分布（m=300, 閾値 5.33E-7）

表 3: 例題 1：経過時間（秒）の内訳

	4 スレッド
フィルタ対角化全体	27.11
- 乱数ベクトル生成	0.04
- 正規直交化	1.25
- フィルタの適用	22.84
- 不変部分空間の基底作成	2.51
- Rayleigh-Ritz 法	0.47
Rayleigh 商逆反復 2 回分	1,443.14

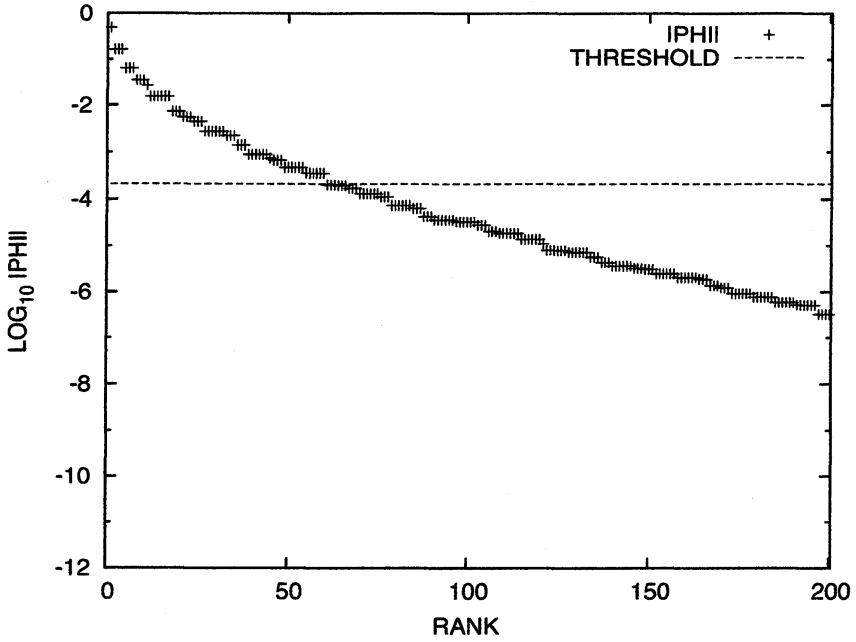

図 8: 例題 2：$|\phi|$ の値の分布（m=300, 閾値 2.12E-4）

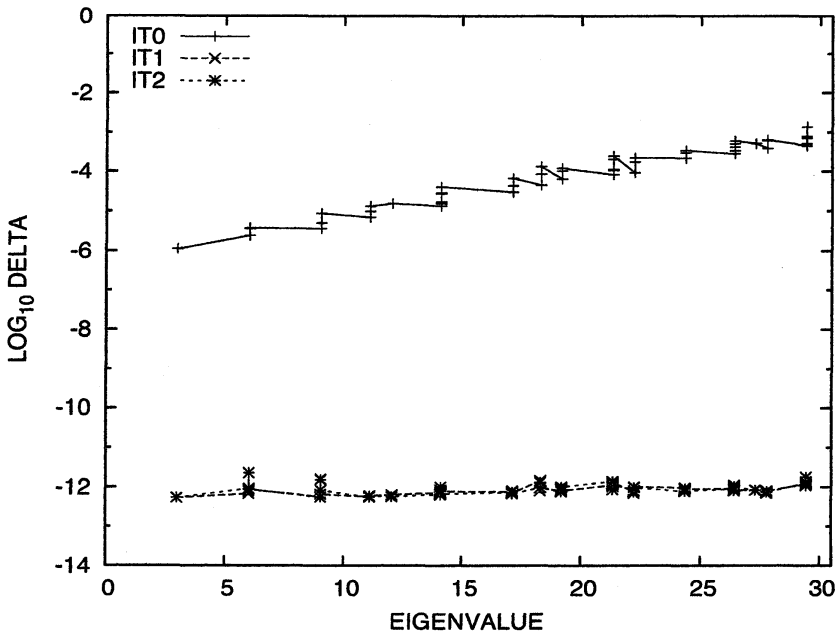

図 9: 例題 2：近似対の残差のノルム Δ（区間 $[0,30]$, m=300）

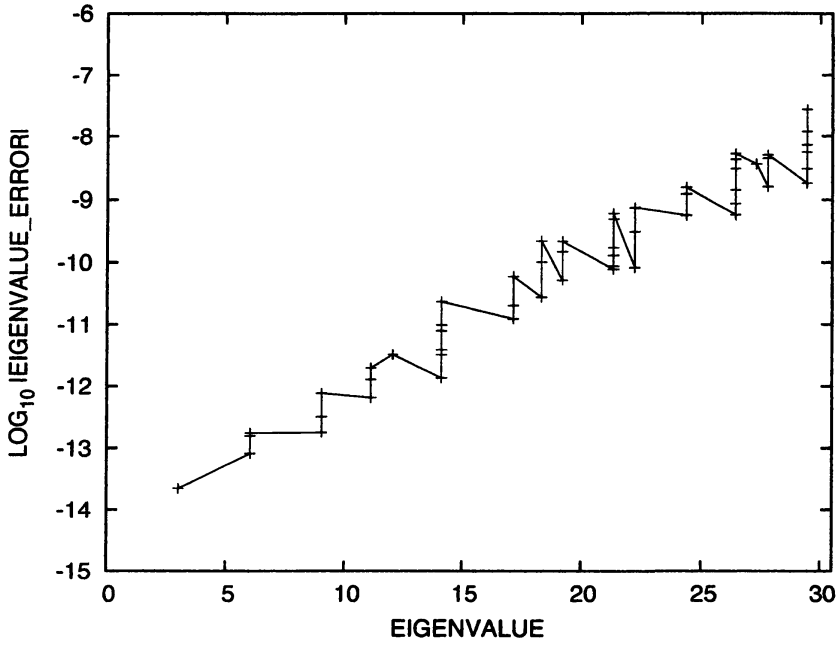

図 10: 例題 2：近似固有値の真値からの誤差の分布（m=300, 閾値 2.12E-4）

表 4: 例題 2：経過時間（秒）の内訳

	4 スレッド
フィルタ対角化全体	26.94
- 乱数ベクトル生成	0.04
- 正規直交化	1.24
- フィルタの適用	22.67
- 不変部分空間の基底作成	2.52
- Rayleigh-Ritz 法	0.47
Rayleigh 商逆反復 2 回分	1,443.40

9 QEに対する期待

今現在は，伝達関数の「極の最適配置」を目的関数の多次元の非線形最適化問題として，数値的な手段による解決を試みているところである．しかしまだ良い結果が得られる段階には至っていない．数値的最適化にはHessianを用いるNewton法や準Newton法であるDFP法（Davidon-Fletcher-Powell）とその改良であるBFGS法(Broyden-Fletcher-Goldfarb-Shanno）などの可変計量法などがある[10]．しかし各種の探索による非線形最適化の解法はどれも，与えた初期値に近い局所的な最適解が得られるだけで，大域的な最適解を得ることが困難であり，しかも通常は得られた解が大域的最適であるかの判定もできない．ところが，問題は有理関数についてであるので，(計算量の点を別とすれば）たとえばQE（Quantifier Elimination）を応用する可能性に期待する．

問題： いま自然数nと，三個の実数パラメタμ，$g_{\rm pass}$，$g_{\rm stop}$（ただし$1<\mu$で$0<g_{\rm stop}<g_{\rm pass}<1$とする）を数値で与えたときに，以下の条件：

1. $g(t)$はn次の実数値の有理関数で，n個の極はすべて負の実数．
2. 通過域$0\le t\le 1$では$g_{\rm pass}\le g(t)\le 1$．
3. 遷移域$1<t<\mu$では$g(t)<g_{\rm pass}$．
4. 阻止域$\mu\le t<\infty$では$|g(t)|\le g_{\rm stop}$．

を満たす$g(t)$の存在を判定し，存在するときには具体的例を構成せよ．

参 考 文 献

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

[2] 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).

[3] Sakurai, T. and Sugiura, H: A Projection Method for Generalized Eigenvalue Problems Using Numerical Integration, *J. Comp. Appl. Math.*, Vol.159, pp.119–128 (2003).

[4] Polizzi, E.: Density-Matrix-Based Algorithm for Solving Eigenvalue Problems, *Phys. Rev. B*, Vol.79, No.11, p.115112[6pages] (2009).

[5] 村上 弘: レゾルベントの線形結合によるフィルタ対角化法, 情報処理学会論文誌コンピューティングシステム, (ACS21), Vol.49, No.SIG2, pp.66–87 (2008).

[6] Ikegami, T., Tadano, H., Umeda, H. and Sakurai, T.: Hierarchical Parallel Algorithm to Solve Large Generalized Eigenproblems, *HPCS2010*論文集, pp.107–114 (2010).

[7] 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).

[8] 村上 弘: フィルタで濾過されたベクトルの組から不変部分空間の直交基底の組を近似構成するフィルタ対角化法, 情報処理学会研究報告, Vol.2011-HPC-129, No.1, pp.1–8 (2011).

[9] 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).

[10] 茨木俊秀: 最適化の数学, 共立出版, 2011.