Vol.2014-HPC-146 No.13
2014/10/3

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

村上 弘[1,a)]

概要：与えられた実対称定値一般固有値問題に対して，フィルタ対角化法を用いて指定した実区間に固有値がある固有対だけを解く．本論文で考察するフィルタ作用素は，ある一つの虚数をシフトとするレゾルベントの多項式として構成されたものである．

キーワード：フィルタ対角化，固有値問題，レゾルベント，多項式

On diagonalization method which uses a polynomial of a resolvent as a filter

Hiroshi Murakami[1,a)]

Abstract: For a given real symmetric definite generalized eigenproblem, by the filter diagonalization method we solve only those eigenpairs whose values are in a specified real interval. The filter operator we study in this paper is constructed as a polynomial of resolvent whose shift is an imaginary number.

Keywords: fiter diagonalization, eigenproblem, resolvent, polynomial

1. はじめに

行列の実対称定値一般固有値問題の解法としてフィルタ対角化法がある [2]．これまではフィルタ作用素として異なるシフトを持つ「レゾルベントの線形結合」を用いる手法を調べてきた [1]（たとえばシフトは複素共役な虚数の対で 6〜16 対程度）．レゾルベントの作用は，固有値問題の係数の行列にシフトを加えた連立 1 次方程式を解いて実現するのであるが，シフト付きの連立 1 次方程式を LU 分解による直接法または不完全 LU 分解前処理付きの反復法で解くと，計算量の少なくない割合を分解が占め，また特に L や U を格納するための記憶量が計算を実行する上での制約となりがちである．

フィルタを構成するレゾルベントをベクトルの組に作用させる作業のための行列分解に注目して記憶量を低減させる方法としては，1) 低次のフィルタ（少ない個数のレゾルベントの線形結合）を数回反復することで阻止域にある固有値を持つ固有ベクトルの削減性能をべき乗で高めて実質的に高次のフィルタとする方法や，2) レゾルベントを全部一度に並行的に計算せずに、一度に扱う個数を制限して何回かに分割して作用を合計していく方法などが考えられる．

1.1 低次のフィルタの反復を行う方法

低次のフィルタの数回の反復をフィルタに用いる方法である．反復により阻止域での減衰率を確保し，低次であることから必要なレゾルベントの個数が減るので，L や U を保持する記憶量が減り，LU 分解の計算量も減る．

例：フィルタとしてレゾルベント 6 個の低次フィルタを 3 回反復したものを使うと，フィルタの特性はレゾルベント 18 個の高次フィルタに比べて悪い．最短の経過時間は 3 倍になり，6 個分の LU 分解が保持できるならば分解は

[1] 首都大学東京・数理情報科学専攻
Department of Mathematics and Information Sciences, Tokyo Metropolitan University
[a)] mrkmhrsh@tmu.ac.jp

6 通り行うだけでよくて高次フィルタの場合の 18 個の 1/3 で済むほか，L や U を保持する記憶量も 1/3 になる．

1.2 同時に扱うレゾルベントの個数を制限する方法

レゾルベントを全部一斉に適用せずに分割し順次適用する．これにより LU を保持する記憶量のピークを抑えることができる．ただし経過時間は分割数に比例して延びるし，LU 分解の演算はレゾルベントの全個数分だけ必要である．

例： レゾルベント 15 個を 5 個ずつの 3 組に分けて一度には 5 個ずつの作用を計算する．フィルタの特性は変化しない．LU の記憶量のピークは 1/3 になるが，経過時間は延びる．LU 分解が 15 通り必要であることは変わらない．もしも一度には 1 個ずつにすると，LU 分解は全部で 15 通り必要で，経過時間も延びるが，LU の記憶量のピークは 1/15 となる．

今回は，記憶量の制約が強い場合を想定し，用いるレゾルベントの個数はもっとも少ない 1 個だけとして，フィルタを「レゾルベントの多項式」により構成してみる．必要なレゾルベントが 1 個なので必要な LU 分解や不完全 LU 分解も 1 回で済む．「レゾルベントの n 次多項式」の作用は，「レゾルベント」の作用を n 回順次に適用することで実現ができ，レゾルベントの作用の適用するときには最初に行った LU 分解が再利用できる．

2. 「レゾルベントの多項式」によるフィルタ

実対称定値一般固有値問題 $A\,\mathbf{v} = \lambda\,B\,\mathbf{v}$ を扱う．シフトが τ のレゾルベントは $\mathcal{R}(\tau) \equiv (A-\tau B)^{-1}B$ とする．固有対 $(\lambda, \mathbf{v})$ へのレゾルベント $\mathcal{R}(\tau)$ の作用は $\mathcal{R}(\tau)\,\mathbf{v} = \frac{1}{\lambda-\tau}\,\mathbf{v}$ である．フィルタをレゾルベントの n 次多項式（の実部）（c_∞ は実数）として $\mathcal{F} = c_\infty I + \mathrm{Re}\sum_{k=1}^{n}\gamma_k\,\{\mathcal{R}(\tau)\}^k$ とする．固有対 $(\lambda, \mathbf{v})$ に対するフィルタの作用は $\mathcal{F}\,\mathbf{v} = f(\lambda)\,\mathbf{v}$ となる．ここで $\mathcal{F}$ による固有値 λ のベクトルの伝達率を与える $f(\lambda)$ は λ の実有理関数 $f(\lambda) = c_\infty + \mathrm{Re}\,\sum_{k=1}^{n}\frac{\gamma_k}{(\lambda-\tau)^k}$ となり，（伝達関数）と呼ぶ．

区間 $[a, b]$ に固有値 λ がある固有対を求めたいとする．いま $\lambda{\in}[a, b]$ から $t{\in}[-1, 1]$ への 1 次変換 $\lambda = \frac{b+a}{2} + \frac{b-a}{2}\,t$ により λ の正規化座標 t を定義する．通過域は $t \in [-1, 1]$，阻止域は $1 < \mu \le |t|$，遷移域は中間の $1 < |t| < \mu$ である．

正規化座標 t での伝達関数 $g(t)$ を $g(t) \equiv f(\lambda)$ で定義する．そうして伝達関数 $g(t)$ の振る舞いに制約を課す：1) 通過域での $g(t)$ の最小値は g_{pass} であり，通過域にあるときに限って $g(t) \ge g_{\mathrm{pass}}$ となる．2) 阻止域での $|g(t)|$ の上限値は g_{stop} である．3) 最後に $g(t)$ を偶関数に限定する．すると（1 対の複素共役な）極は純虚数である．

いま虚部が正の極を $i\sigma$（$\sigma > 0$）とすると，$g(t) = c'_\infty + \mathrm{Re}\,\sum_{k=1}^{n}\frac{c_k}{(t-i\sigma)^k}$．また $f(\lambda) = c_\infty + \mathrm{Re}\,\sum_{k=1}^{n}\frac{\gamma_k}{(\lambda-\tau)^k}$ なので，$f(\lambda) = g(t)$ により，両者の間には，定数項：$c_\infty = c'_\infty$，極の位置：$\tau = \frac{a+b}{2} + i\left(\frac{b-a}{2}\right)\sigma$，多重極の係数：$\gamma_k = \left(\frac{b-a}{2}\right)^k c_k$，$k{=}1, 2, \ldots, n$ の関係が成立する．そうして $g(t)$ が決まると，$f(\lambda)$ が決まり，$f(\lambda)$ からはそれに対応する作用素 $\mathcal{F}$ が決まる．

さらに，計算の簡便化のために実数 $\sigma > 0$ による尺度変換で，$t \equiv \sigma x$ と置き，$h(x) \equiv g(t)$ と定義する．$h(x) = c''_\infty + \mathrm{Re}\,\sum_{k=1}^{n}\frac{\alpha_k}{(1+ix)^k}$．係数間の関係は $c''_\infty = c_\infty$，$c_k = (-i\sigma)^k\alpha_k$，$k{=}1, 2, \ldots, n$ である．

$g(t)$ が偶関数であると $h(x)$ も偶関数となり，係数 α_k はすべて実数となる．いま $x \equiv \tan\theta$ とおけば，$h(x) = c_\infty + \sum_{k=1}^{n}\alpha_k\cos(k\theta)\,(\cos\theta)^k$．また $\widehat{h}(\theta) = h(x)$ とする．また以降では，無限遠での伝達率を零に限定して $c_\infty = 0$ とする．

以上をまとめると，以下のようになる（ただし少し簡単化して $\sigma = 1$ とする）．1 次変換 $\lambda = \frac{a+b}{2} + \frac{b-a}{2}\,x$ で $\lambda \in [a, b]$ と $x \in [-1, 1]$ を対応させる．x の偶関数である実有理関数 $h(x)$ が，極を $x = \pm i$ だけに n 重に持ち，（簡単のため）無限遠での値が零とすると，$h(x) = \mathrm{Re}\,\sum_{k=1}^{n}\frac{\alpha_k}{(1+ix)^k}$ となる．そうして実数パラメタ α_k，$k{=}1, 2, \ldots, n$ をうまく調整して通過域 $[-1, 1]$ 付近から離れるとき $|h(x)|$ が急減少であるようにする．

$h(x)$ の座標を λ に変換した実有理関数を $f(\lambda) \equiv h(x)$ とする．伝達関数が $f(\lambda)$ である作用素 $\mathcal{F}$ はレゾルベントの多項式を用いて $\mathcal{F} = \mathrm{Re}\sum_{k=1}^{n}\gamma_k\,\{\mathcal{R}(\tau)\}^k$ となる．ただし，シフト $\tau \equiv \frac{a+b}{2} + i\left(\frac{b-a}{2}\right)$，係数 $\gamma_k \equiv (-i)^k\left(\frac{b-a}{2}\right)^k\alpha_k$ である．

次数 n を与え，極を $\pm i$ だけに持つ偶関数の n 次実有理関数 $h(x) = \mathrm{Re}\,\sum_{k=1}^{n}\frac{\alpha_k}{(1+ix)^k}$ を実数パラメタ α_k，$k{=}1, 2, \ldots, n$ をうまく調整して，1) 通過域 $[-1, 1]$ では $|h(x)|$ は 1 の付近．2) 通過域からある程度離れた阻止域では $|h(x)|$ は微小．となるようにする．

計算の便宜上，$x \equiv \tan\theta$ と置換し，$\widehat{h}(\theta) \equiv h(x)$ と置くと $\widehat{h}(\theta) = \sum_{k=1}^{n}\alpha_k\cos(k\theta)\,(\cos\theta)^k$ である．$h(x)$ は偶関数だから，θ は非負の範囲で考えれば良くて，例えば通過域は $\theta \in [\,0, \frac{\pi}{4}\,]$ とできる．

パラメタ α_k の調整は今回は最小 2 乗法（の系統）で行った．

3. 最小 2 乗法（解法 I）

伝達関数 $\widehat{h}(\theta)$ の理想特性からのずれの 2 乗を阻止域と通過域に於いて（簡単のため重み 1 で）θ で積分したものをそれぞれ J_{stop} と J_{pass} とする．そうして微小な正数 η を適切に選んで，目的関数 $J \equiv J_{\mathrm{stop}} + \eta J_{\mathrm{pass}}$ を最小化する．

通過域の中央と端点 $x{=}0$，$x{=}1$，阻止域の両端点 $x{=}\mu$，$x{=}\infty$ に対応する θ の値をそれぞれ $\theta_0{=}0$，$\theta_1{=}\frac{\pi}{4}$，$\theta_\mu{=}\arctan\mu$，$\theta_\infty{=}\frac{\pi}{2}$ とすると，$J_{\mathrm{stop}}{\equiv}\int_{\theta_\mu}^{\theta_\infty}\{\widehat{h}(\theta)\}^2 d\theta = \sum_{p,q=1}^{n}\alpha_p A_{p,q}\alpha_q$，$J_{\mathrm{pass}}{\equiv}\int_{\theta_0}^{\theta_1}\{1-$

$\widehat{h}(\theta)\}^2 d\theta \;=\; \sum_{p,q=1}^{n} \alpha_p B_{p,q} \alpha_q \;-\; 2\sum_{p=1}^{n} \alpha_p B_{p,0} \;+\; \mathrm{const}$ となる．ただし，$A_{p,q} \equiv \int_{\theta_\mu}^{\theta_\infty} \cos(p\theta)\cos(q\theta)(\cos\theta)^{p+q} d\theta$，$B_{p,q} \equiv \int_{\theta_0}^{\theta_1} \cos(p\theta)\cos(q\theta)(\cos\theta)^{p+q} d\theta$ であり，これらの積分が値を解析式で計算できることが容易に示せる（注：ここでの行列 A や B は一般固有値問題の係数行列とは無関係）．$J \equiv J_{\mathrm{stop}} + \eta J_{\mathrm{pass}}$ の最小化条件は，$b_p = B_{p,0}$ とおくと，対称正定値な連立 1 次方程式 $(A + \eta B)\,\alpha = \eta\,\mathbf{b}$ になり，これを解くとパラメタ α_k，$k{=}1, 2, \ldots, n$ が決まる．

4. 最小 2 乗法（解法 II）

α を 2-ノルム一定のベクトルとみなし，まず最初に阻止域での積分 $J_{\mathrm{stop}} = \sum_{p,q=1}^{n} \alpha_p A_{p,q} \alpha_q = \alpha^T A \alpha$ の最小化を考える．対称正定値行列 A の固有値分解を $A \to UDU^T$ とする．D は A の固有値を順に並べた対角行列で，U は固有ベクトルを固有値の順に並べた直交行列である．いま α を固有値が最小の固有ベクトルに採るならば，J_{stop} の値は最小になるが，$h(x)$ の通過域での振舞いを調整する自由度が無い．そこで自由度を追加する工夫を入れる．それにはまず微小な正数 ϵ を適切に決める．そうして固有値が ϵ 以下の A の固有対は ℓ 個であるとし，それら ℓ 個の固有ベクトルの張る部分空間を $\mathcal{S}^{(\ell)}$ とする．すると $\alpha \in \mathcal{S}^{(\ell)}$ であれば J_{stop} は上から $\epsilon\,||\alpha||_2^2$ で抑えられる．ベクトル α を部分空間 $\mathcal{S}^{(\ell)}$ に制限して J_{pass} の値を最小化すると，それは ℓ 次に縮小された対称正定値な連立 1 次方程式に帰着するので，それを解けば α が決まる．

部分空間を拡げると，伝達関数の理想特性からのずれが通過域では減って改善されるが，阻止域では伝達率が増えて悪化しまうので，両者の釣り合いを考える必要がある．

5. 実験例

解法 II を用いて $n = 15$，$\mu = 2.0$ のフィルタを求めた例を示す．途中の数値計算には intel Fortran の IEEE 754 四倍精度の数値と演算を用いた．行列 A の固有値分解には閾 Jacobi 法を使用した．この例で固有値の閾値を $\epsilon = 10^{-30}$ と設定すると $\ell = 2$ となった．得られた係数 α_k，$k{=}1, 2, \ldots, n$ を IEEE 754 の倍精度に丸めた値を表 1 に示す．係数の大きさは 1 程度で良く揃っている．

偶関数性から片側 $x{>}0$ での伝達関数 $h(x) = \mathrm{Re}\sum_{k=1}^{n} \frac{\alpha_k}{(1+ix)^k}$ の大きさ $|h(x)|$ を（倍精度に丸めた係数 α_k を用いて倍精度演算により）両対数でプロットしたグラフを図 1 に示す．図 2 は図 1 と同じ内容を片対数グラフで表示したものである．

伝達率 $h(x)$ は最大値 1.325 を $x = 0.216$ 付近でとり，$h(0) = 7.2{\times}10^{-1}$，$h(1) = 2.3{\times}10^{-4}$，$h(1.5) = 1.9{\times}10^{-8}$，$h(2) = 1.1{\times}10^{-15}$ である．

$h(x)$ の通過域での最小値は $y_{\min} = 0.71862$ で，遷移域で $h(x) = y_{\min}$ となる x の値は $x_0 = 0.388220$ である．そ

表 1　係数 α（解法 II，$n = 15$，$\mu = 2.0$，$\epsilon = 10^{-30}$）

k	α_k
1	3.10422 91727 23495 10^{-1}
2	3.10422 91727 25609 10^{-1}
3	2.85453 67519 83506 10^{-1}
4	2.35515 19113 67395 10^{-1}
5	1.64913 99494 59607 10^{-1}
6	8.22631 58940 55446 10^{-2}
7	-6.57520 79352 44120 10^{-4}
8	-7.11802 27019 60262 10^{-2}
9	-1.18756 19212 14338 10^{-1}
10	-1.37828 28527 33139 10^{-1}
11	-1.29654 88587 73316 10^{-1}
12	-1.01680 66293 50991 10^{-1}
13	-6.60360 83956 00963 10^{-2}
14	-3.26587 11429 62141 10^{-2}
15	-1.19174 53737 97113 10^{-2}

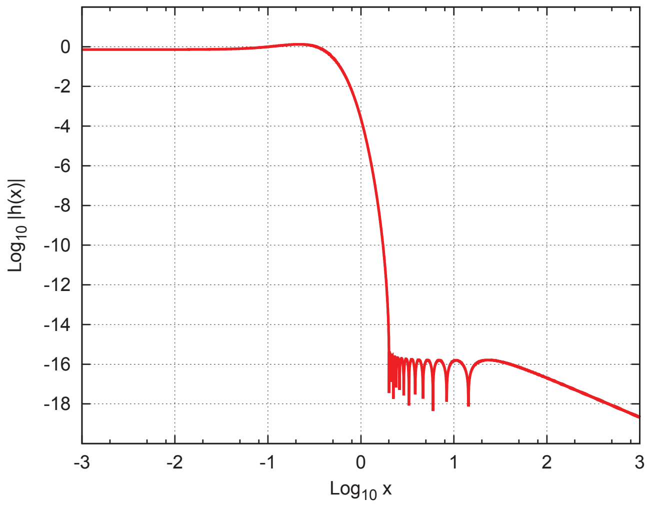

図 1　伝達関数の大きさ $|h(x)|$（解法 II, $n = 15, \mu = 2.0, \epsilon = 10^{-30}$.（両対数））

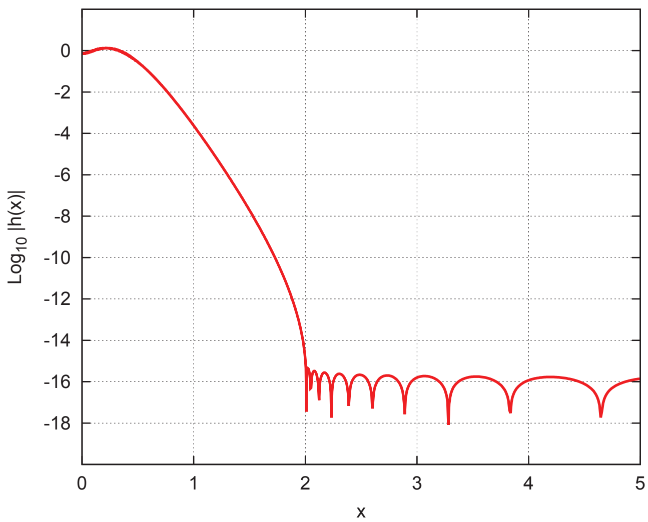

図 2　伝達関数の大きさ $|h(x)|$（解法 II, $n = 15, \mu = 2.0, \epsilon = 10^{-30}$.（片対数））

こで再スケール処理 $h'(x') \equiv h(x' * x_0)/y_{\min}$ を行なった（ここでの h' は h の微分の意味ではない）．偶関数性から片側 $x'>0$ での伝達関数の大きさ $|h'(x')|$ を両対数でプロットしたグラフを図 3 に示す．図 4 は図 3 と同じ内容を片対数グラフで表示したものである．$x' \in [-1,1]$ では $h'(x') \geq 1$ で，最大値 1.844 を $x' = 0.557$ でとり，$h'(1)$=1，$h'(2)$=1.18×10^{-2}，$h'(3)$=1.84×10^{-5}，$h'(4)$=8.68×10^{-9}，$h'(5)$=1.18×10^{-13}，$h'(6)$=2.94×10^{-16} などである．

μ=3 とすると，$g_{\rm stop}$=1.8×10^{-5}，μ=4 とすると，$g_{\rm stop}$=8.7×10^{-9}，μ=5 とすると，$g_{\rm stop}$=1.2×10^{-13}，μ=6 とすると，$g_{\rm stop}$=2.9×10^{-16}，となる．

μ を大きくして遷移域が広がると，不要な固有値を持つベクトルが遷移域に多数含まれると，フィルタで濾過するベクトルの個数をそれだけ増やす必要がでてくる．$g_{\rm stop}$ が微小でなければ，フィルタにより構成する不変部分空間の近似度がそれだけ悪くなる．また，通過域内での伝達率の最大最小比が大きいと，その影響で得られる固有対の精度がバラつく可能性がある．

6. まとめ

「フィルタ対角化法」は，固有値問題に付随するレゾルベントで構成された作用素をフィルタとして用いて，不変部分空間の近似基底を構成する手法である [2]．レゾルベントの作用は連立 1 次方程式を直接法もしくは反復法で解いて実現するが，直接法であれば LU 分解，反復法であっても前処理に不完全 LU 分解を用いるならば，シフトの異なるレゾルベントそれぞれについて，LU 分解の計算を行い，その結果の L と U を格納する記憶が必要がある．

レゾルベントの線形結合による高品位のフィルタは，ある程度の個数のレゾルベントから構成されているが，並列度を最大化するため多数のレゾルベントを一度に作用させると，LU 格納のための記憶容量が個数に比例して必要で，実行が阻害されがちである．そのようなときは，レゾルベントの作用を一度にまとめて計算せずに，何組かに分割して順次に計算する方法をとることが必要になる．

今回は，用いるレゾルベントの個数を 1 個にして，フィルタをレゾルベントの多項式で構成する可能性を調べた．多項式の係数を最小 2 乗法の手法でレゾルベントの多項式であるフィルタの伝達関数が，理想フィルタの特性をできるだけ良く近似するようにする（最小 2 乗法以外の近似法の適用も考えられる）．現在の方法で得られている「レゾルベントの多項式」によるフィルタの伝達関数は，遷移域に於ける変化が極を複数配置できる場合と比べると急峻ではない．

今後は，より良い特性の伝達関数を得る方法を探す試みを行うほか，構成した伝達関数に対応するフィルタ作用素を実際に使用してフィルタ対角化法により固有対の近似がうまく得られるかの検証を行うつもりである．

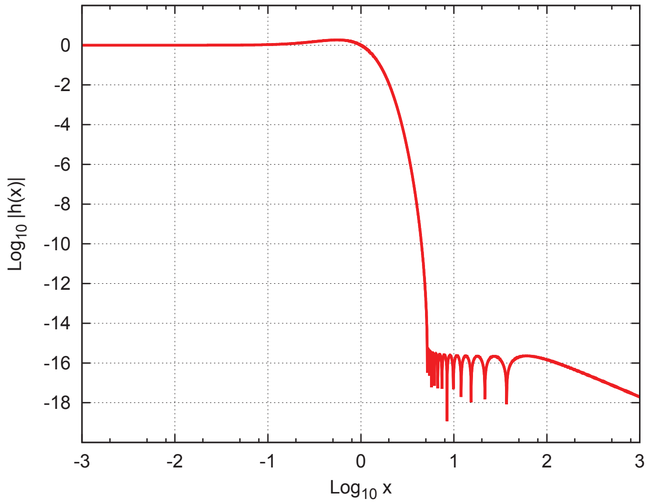

図 3 伝達関数の大きさ $|h'(x')|$ （解法 II，$n = 15$，再スケール．(両対数)）

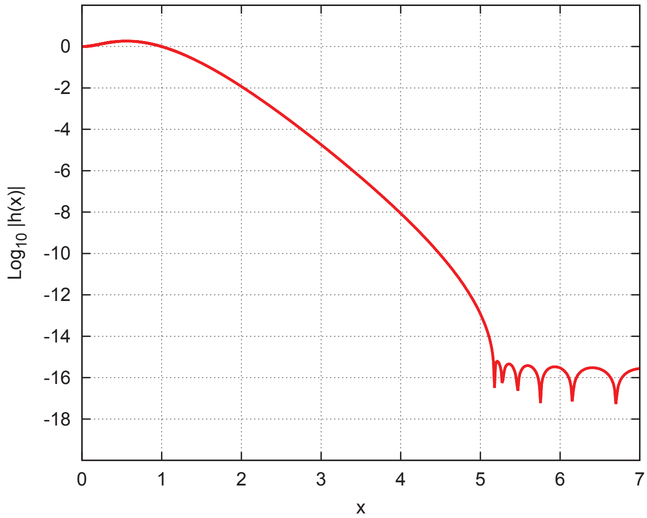

図 4 伝達関数の大きさ $|h'(x')|$ （解法 II，$n = 15$，再スケール．(片対数)）

参考文献

[1] 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 (2010), pp.1–21.

[2] 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 (2011), pp.1–14.