

# 数値代数方程式のフィルタ対角化法による解法

村上 弘

MURAKAMI HIROSHI

首都大学東京 数理情報科学専攻

DEP. OF MATHEMATICS AND INFORMATION SCIENCES, TOKYO METROPOLITAN UNIVERSITY \*

**要約:** 随伴行列法で数値係数の代数方程式 (系) を解くには, 疎で非対称な数値行列の固有値を求める必要がある. フィルタ対角化法を用いて, 随伴行列の指定した位置付近にある比較的少数個の固有値だけを求める効率的な方法について説明し, 一変数の場合の簡単な例を 1 つ示す.

## 1 導入

「随伴行列法」により数値係数の代数方程式に対応する随伴行列を構成し, その固有値を数値的に解いて方程式の零点を求めることにする [2, 8, 9, 10, 11, 12]. 一変数の場合が最も典型的であるが, 随伴行列法は多変数の方程式系にも適用可能である [5]. 高次方程式や多変数の場合には特に, 行列の次数が極めて大きくなる. 随伴行列は一変数の場合は Hessenberg 形で, 多変数の場合にも類似の構造を持つ極めて疎な行列になり, 方程式の係数が疎であればさらに疎性が高まる.

### 1.1 非対称行列の固有値解法 (QR 反復法)

非対称行列の固有値を求める標準的な解法には, まず行列を Householder 変換または Gauss 消去法により Hessenberg 形へ変換し (最初から Hessenberg 形の場合には省略), その後に Hessenberg 形を QR 反復法で収束させて三角化 (Schur 分解) する方法がある.

QR 反復法では通常は全部の固有値を求めるので計算量が  $O(N^3)$  と多いことや, 疎行列でも fill-in の発生により非零要素が増えて  $O(N^2)$  の記憶量が必要となること, などが難点として挙げられる.

大規模行列の場合には特定の位置付近の少数の固有値だけが必要な場合が多い. 少数個の必要な固有値だけを求めるための QR 反復法のシフト戦略についての研究が今後行なわれることが望まれる.

### 1.2 非対称行列の固有値解法 (双 Lanczos 法)

双 Lanczos 法 (Two-Sided Lanczos Method), または Lanczos 双直交化 (Lanczos Biorthogonalization) は, 元の行列  $A$  を変形せずに疎性を最大限に活かして,  $A$  と相似な非対称三重対角行列  $T$  を作る方法で,  $A$  と  $A^T$  の Krylov 部分空間の双直交基底  $U, V$  を生成する:  $V^T U = I, AU = UT, VA = TV$ . 現実には, 丸め誤差や漸化式の不安定性から, 対称の場合の Lanczos 法と同様,  $T$  は安定には求められない. しかし Lanczos 法は冪乗法 (Power Method) の収束性の改良版とみなせ, 両端付近の固有値は計算を途中で打ち切った三重対角行列の固有値により良く近似できる性質がある. Lanczos 法は魅力的ではあるが, 今回の考察対象にはしない. 他に, 非対称固有値問題の疎性を活かせる反復解法として Arnoldi 法 [7] がある.

### 1.3 固有値（零点）の一部だけがが必要な場合

方程式（系）の指定位置の付近の一部の零点だけがが必要な場合は、通常は Newton 法などの反復法に初期値を与えて解くが、与えた初期値で求まる零点が決まるので、必要な零点を落とす可能性がある。

「指定位置の付近」で方程式（系）の多項式（系）を低次の多項式（系）で近似し、得られた近似方程式（系）を解いて得た零点の近似値を元の方程式（系）の反復法の初期値とする方法 [11, 12] も考えられる。しかし、随伴行列法で得た行列の指定位置付近にある固有値を直接求める方が安全で確実であろう。

### 1.4 逆反復法（Inverse Iteration）

逆反復法は固有値問題の少数個の「指定位置の付近にある固有値」を求めるのに良く用いられる [14, 16]。固有値の近似値  $\rho$  を与えて、一般的な初期ベクトル  $\mathbf{x}$  に対し、線形方程式  $(A - \rho I)\mathbf{y} = \mathbf{x}$  を解き（この線形方程式が速く解けることが利用の前提である）、 $\mathbf{y}$  を適当にスケーリングした後に  $\mathbf{x}$  と置き直して反復する。これは逆行列  $(A - \rho I)^{-1}$  による冪乗法で、固有値は  $\rho$  に近い固有ベクトルへ収束する。十分に収束したベクトル  $\mathbf{y}$  から固有値は Rayleigh 商  $(\mathbf{y}^* A \mathbf{y}) / (\mathbf{y}^* \mathbf{y})$  で近似する。

### 1.5 同時逆反復法（Rayleigh-Ritz 付き）

同時逆反復法では複数の解を同時に扱い、縮重や近接の固有値の収束を改善する [14, 16, 20]。

- 1)  $\rho$  の初期値を与える。初期ベクトルの組  $X$  を与える。
- 2) 組  $X$  に  $(A - \rho I)^{-1}$  を作用させて組  $Y$  を作る。
- 3)  $Y$  を正規直交化  $Y^* Y = I$  の後、縮小行列（Rayleigh 商の行列） $C \equiv Y^* A Y$  を作る。
- 4)  $C$  を正則行列  $W$  で Jordan 形  $J$  に変形： $CW = WJ$ 。
- 5)  $J$  の対角と対応するベクトル  $Z \equiv YW$  の組を作る（Ritz 対）。

$Y$  から Ritz 対を作る操作（3～5）が "Rayleigh-Ritz procedure" である。

より安定な計算の為に、 $Z$  が Schur ベクトルを近似するよう 4,5 を変更する：

- 4')  $C$  を Unitary 行列  $U$  で三角行列  $T$  に変形： $CU = UT$ 。
- 5')  $T$  の対角とそれに対応するベクトル  $Z \equiv YU$  の組を作る。

以降の反復では、まず  $Z$  を  $X$  と置き、それから上記 1,2 の箇所を変えて（II は収束の加速を狙ったもの）、

方式 I)  $\rho$  を固定して、 $X$  の全ての列ベクトルに  $(A - \rho I)^{-1}$  を作用させて  $Y$  を作る。

方式 II)  $\rho$  を  $X$  の各列ベクトル毎に異なる Ritz 値にして、 $(A - \rho I)^{-1}$  を作用させて  $Y$  を作る。

とし、 $Y$  の組から Rayleigh-Ritz procedure の 3,4,5 あるいは 3,4',5' で、最新の Ritz 値と Ritz ベクトル  $Z$  を求める。以上を繰り返す。

逆反復法も同時逆反復法も、 $\rho$  が固定では 1 次収束で、固有値に十分に近い値を  $\rho$  に与えないと収束は遅い。 $\rho$  からの距離が最も近い固有値を持つ固有成分が優越して成長するので、同等の距離の固有成分との分離は良くない。Ritz 値（Rayleigh 商）を用いて  $\rho$  の値を動かすと、固有値に縮重や極端な近接がなければ最終的には 2 次収束となるが、初期ベクトルが悪いと  $\rho$  の値が跳ぶので注意して制御を行なう。同時逆反復では  $\rho$  の値を共用すると LU-分解が節約できるが収束が遅いので、線形方程式が容易に解ける Hessenberg 形などの場合には、節約の効果よりも反復回数が増してむしろ損になることがある。

以降で説明する「フィルタ対角化法」[19] を用いて、指定した位置付近に固有値を持つ固有対（複数）に対する一様に品質の良い近似対を求めて、その近似対を逆反復法や同時逆反復法の初期値とするならば、ごく少ない反復回数で収束すると期待できる。

## 2 フィルタ対角化法

### 2.1 フィルタ対角化法の概要

フィルタ対角化法の概要を説明する。行列  $A$  の「選択したい固有値」を持つ固有ベクトル成分は通すが、その他の固有ベクトル成分は相対的に強く減衰させる「線形フィルタ」 $\mathcal{F}$  を構成する。そのフィルタに「一般的なベクトル」 $\mathbf{x}$  を通すと、性質から出力  $\mathbf{y} = \mathcal{F}\mathbf{x}$  は「選択したい固有成分」の近似的な線形結合となり、それ以外の固有成分の割合は微小となる。いま  $S$  を「選択したい固有値」を持つ固有ベクトルの張る部分空間とする。フィルタの出力を十分集めて分析して、 $S$  を近似する部分空間  $S'$  の基底の組を構成する。複数の固有部分空間の直和である  $S$  は  $A$  の不変部分空間である。Rayleigh-Ritz 法を  $S$  に適用すれば、 $S$  内の  $A$  の固有ベクトルが全て求まる。実際には  $S'$  に適用するので、近似固有ベクトルが得られる。

### 2.2 不変部分空間内の固有ベクトルの決定 (Rayleigh-Ritz 法)

$Y$  が  $A$  の右不変部分空間を張る基底の組ならば、 $AY = YC$  を満たす正方行列  $C$  がある。  $C$  は  $AY$  の各ベクトルを  $Y$  の各ベクトルで再展開した係数である。予め基底の組  $Y$  が正規直交  $Y^*Y = I$  になっていれば、 $C = Y^*AY$  として  $C$  (Rayleigh 商の行列) が求まる。(残差  $AY - YC$  は、(右) 不変部分空間からの  $Y$  のズレを表す。)

いま  $Y^*Y = I$  で、 $Y$  が  $A$  の不変部分空間の基底の組  $AY = YC$  とするとき：

Jordan 分解で  $C$  を Jordan 形  $J$  へ相似変換して  $CW = WJ$  とし、 $Z \equiv YW$  と置くと  $AZ = ZJ$  だから、 $J$  の対角の値と対応する  $Z$  の列ベクトルは  $A$  の固有値と (一般) 固有ベクトルである。

計算の安定性からは、以下の固有値と Schur ベクトルを求める方法の方が良い：

Schur 分解で  $C$  を三角行列  $T$  へユニタリ変換して  $CU = UT$  とし、 $Z \equiv YU$  と置くと  $AZ = ZT$  だから、 $T$  の対角の値と対応する  $Z$  の列ベクトルは  $A$  の固有値と Schur ベクトルである。

小規模な非対称行列  $C$  の相似変換による対角化する場合には (数値的には Jordan 標準形は安定には求められないが)、複素 Jacobi 回転を用いる [1] の方法が使用できる。  $C$  をユニタリ変換で三角化する場合は、複素 Householder 変換で Hessenberg 化の後に、複素 QR 反復法を用いることができる。  $C$  が実行列であれば、Householder 変換後の実 Hessenberg 行列には double-QR 法を用いるのが効率的である。

### 2.3 諸注意

フィルタの入力の組  $X$  に乱数を用いる場合には、組  $X$  の独立性が低いと組  $Y$  の独立性も低くなり無駄が多いので、 $X$  をあらかじめ正規直交化 ( $X^*X = I$ ) しておく。但し、高い直交化精度は必要としない。

フィルタを通した直後の出力の組  $Y$  の各ベクトルのノルムが (途中の計算で生じる数値誤差程度で) どれも極端に小さい場合は、1) 行列  $A$  の固有値がフィルタの通過域には存在しない、2) 選出したい固有ベクトル成分を入力ベクトルの組  $X$  が欠いている、のいずれかである。場合 1) に対してはフィルタ特性を変えてみる。場合 2) については、初期ベクトルの組  $X$  としてある程度十分な個数の乱数ベクトルをとればまず生じない (必要なら  $X$  を取り換えてみる)。

$Y$  の正規直交化には特異値分解 (SVD) の使用が理論上最も明解である (列交換を伴う Householder-QR 分解 [15] も利用できる)。SVD の特異値 (列交換 Householder-QR 分解のピボット値) は、「近似的な  $A$  の不変部分空間」 $S'$  を張る直交基底でフィルタの出力を説明する尺度である。直交化の際には相対的な閾値を設けて特異値 (ピボット値) の切断を行い、数値独立性の小さい成分を棄てて誤差の拡大を防ぎ計算の安定性を保つ。フィルタが不用な固有成分を明瞭に遮断できる場合は、フィルタの出力  $Y$  の特異値を大きい側から順に並べると、ある所から先で急に小さい値 (相対的な零) になる。フィルタの能力が不十分で不用

な固有ベクトル成分が十分に遮断できない場合は、特異値分布は緩やかに減少する。切断後の特異値の個数が（近似）不変部分空間  $S'$  の階数となり、不用な固有ベクトル成分の混入した個数分だけ階数が増加する。

行列  $A$  が実の場合に、フィルタ  $\mathcal{F}$  を線形作用素として実 ( $\bar{\mathcal{F}} = \mathcal{F}$ ) となるように選ぶならば、入力組  $X$  を実にとれば出力組  $Y = \mathcal{F}X$  もまた実になる。あるいは入力  $X$  を複素に選んで得た出力  $Y$  の実部と虚部はそれぞれ  $X$  の実部と虚部に対する出力組になる。このことを利用すると、フィルタの作用を計算する演算量が半減できる。他にも利点として、出力ベクトルの基底を実にとると、 $A$  が実なので Rayleigh-Ritz 法の Rayleigh 商の行列  $C$  も実行列となり、複素共役対称性を保つ算法を Jordan 分解（固有値分解）や Schur 分解に用いれば、フィルタ対角化法の出力も複素共役対称性を保ったものになる。

### 3 レゾルベント (resolvent) の線形結合によるフィルタ

#### 3.1 準備

$N$  次行列  $A$  の Jordan 標準形を  $J$  とし、 $A$  を  $J$  へ相似変換する正則な行列を  $U$  とすると  $AU = UJ$  である。そのとき、シフト量  $\rho$  の  $A$  のレゾルベント  $R(\rho) \equiv (A - \rho I)^{-1}$  についても  $R(\rho)U = U(J - \rho I)^{-1}$  が成立する。同様に  $k$  個のシフト量  $\rho_\ell$  の  $A$  のレゾルベントの線形結合  $\mathcal{F} \equiv \sum_{\ell=1}^k \omega_\ell R(\rho_\ell)$  についても、線形性から  $\mathcal{F}U = U\mathcal{F}$  が成立する。但し、 $\mathcal{F} \equiv \sum_{\ell=1}^k \omega_\ell \cdot (J - \rho_\ell I)^{-1}$  は、 $\mathcal{F}$  を  $U$  で相似変換した行列、 $U$  の列ベクトル  $u^{(\mu)}$ ,  $\mu = 1, 2, \dots, N$  は  $A$  と  $\mathcal{F}$  に共通の右（一般）固有ベクトルである。

$A$  の Jordan 標準形  $J$  が番号  $\nu$  で識別される Jordan 細胞  $J_{p_\nu}(\lambda_\nu)$  (固有値  $\lambda_\nu$ , サイズ  $p_\nu$  を持つ) の対角ブロックに分かれるとき、逆行列  $(J - \rho I)^{-1}$  も対応して対角ブロックに分かれ、 $(J - \rho I)^{-1} = \bigoplus_{\nu} \{J_{p_\nu}(\lambda_\nu - \rho)\}^{-1}$  となる。Jordan 細胞  $J_p(a)$  の逆行列の例を  $p=1, 2, 3, 4$  について示す。但し  $b \equiv -a^{-1}$ 。

$$\begin{aligned} [a]^{-1} &= -[b], \quad \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}^{-1} = -\begin{bmatrix} b & b^2 \\ 0 & b \end{bmatrix}, \quad \begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}^{-1} = -\begin{bmatrix} b & b^2 & b^3 \\ 0 & b & b^2 \\ 0 & 0 & b \end{bmatrix}, \quad \begin{bmatrix} a & 1 & 0 & 0 \\ 0 & a & 1 & 0 \\ 0 & 0 & a & 1 \\ 0 & 0 & 0 & a \end{bmatrix}^{-1} \\ &= -\begin{bmatrix} b & b^2 & b^3 & b^4 \\ 0 & b & b^2 & b^3 \\ 0 & 0 & b & b^2 \\ 0 & 0 & 0 & b \end{bmatrix}. \end{aligned}$$

$\mathcal{F}$  も  $J$  の Jordan 細胞分割  $J_p(\lambda)$  と対応して対角ブロック  $F_p(\lambda)$  に分かれて、

$$F_p(\lambda) \equiv \sum_{\ell=1}^k \omega_\ell \cdot \{J_p(\lambda - \rho_\ell)\}^{-1} \text{ で, } F_{p=1}(\lambda) = \sum_{\ell=1}^k \frac{\omega_\ell}{\lambda - \rho_\ell} [1], \quad F_{p=2}(\lambda) = \sum_{\ell=1}^k \frac{\omega_\ell}{\lambda - \rho_\ell} \begin{bmatrix} 1 & \frac{-1}{\lambda - \rho_\ell} \\ 0 & 1 \end{bmatrix},$$

$$F_{p=3}(\lambda) = \sum_{\ell=1}^k \frac{\omega_\ell}{\lambda - \rho_\ell} \begin{bmatrix} 1 & \frac{-1}{\lambda - \rho_\ell} & \frac{1}{(\lambda - \rho_\ell)^2} \\ 0 & 1 & \frac{-1}{\lambda - \rho_\ell} \\ 0 & 0 & 1 \end{bmatrix}, \text{ などとなる. } f(\lambda) \equiv \sum_{\ell=1}^k \frac{\omega_\ell}{\lambda - \rho_\ell} \text{ と置くと, } f'(\lambda) =$$

$$\sum_{\ell=1}^k \frac{-\omega_\ell}{(\lambda - \rho_\ell)^2}, \quad \frac{f''(\lambda)}{2!} = \sum_{\ell=1}^k \frac{\omega_\ell}{(\lambda - \rho_\ell)^3}, \quad \frac{f'''(\lambda)}{3!} = \sum_{\ell=1}^k \frac{-\omega_\ell}{(\lambda - \rho_\ell)^4}, \text{ などから, } F_{p=1}(\lambda) = \begin{bmatrix} f(\lambda) \end{bmatrix}, \quad F_{p=2}(\lambda) =$$

$$\begin{bmatrix} f(\lambda) & f'(\lambda) \\ 0 & f(\lambda) \end{bmatrix}, \quad F_{p=3}(\lambda) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} \\ 0 & f(\lambda) & f'(\lambda) \\ 0 & 0 & f(\lambda) \end{bmatrix}, \quad F_{p=4}(\lambda) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \frac{f'''(\lambda)}{3!} \\ 0 & f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} \\ 0 & 0 & f(\lambda) & f'(\lambda) \\ 0 & 0 & 0 & f(\lambda) \end{bmatrix}.$$

#### 3.2 フィルタ $\mathcal{F}$ の構成

以降では簡単の為、 $A$  の Jordan 標準形が対角になる場合だけを説明する（Jordan 標準形が対角にならない場合の解析は式が多少複雑となるが、分割されたブロックに基づき同様の結果を導ける）。

$A$  の右固有対を  $(\lambda^{(\mu)}, \mathbf{u}^{(\mu)})$  とする。いま  $k$  個の相異なるシフト量  $\rho_i$  の  $A$  のレゾルベント  $R(\rho_i)$  を係数  $\omega_i$  で線形結合した作用素  $\mathcal{F} \equiv \sum_{i=1}^k \omega_i R(\rho_i)$  と、有理関数  $f(\lambda) \equiv \sum_{i=1}^k \omega_i / (\lambda - \rho_i)$  を定義すると、 $\mathcal{F}\mathbf{u}^{(\mu)} = \mathbf{u}^{(\mu)} f(\lambda^{(\mu)})$  で、 $\mathbf{u}^{(\mu)}$  は  $\mathcal{F}$  の右固有ベクトルでもあり、対応する  $\mathcal{F}$  の固有値は  $f(\lambda^{(\mu)})$  となる。

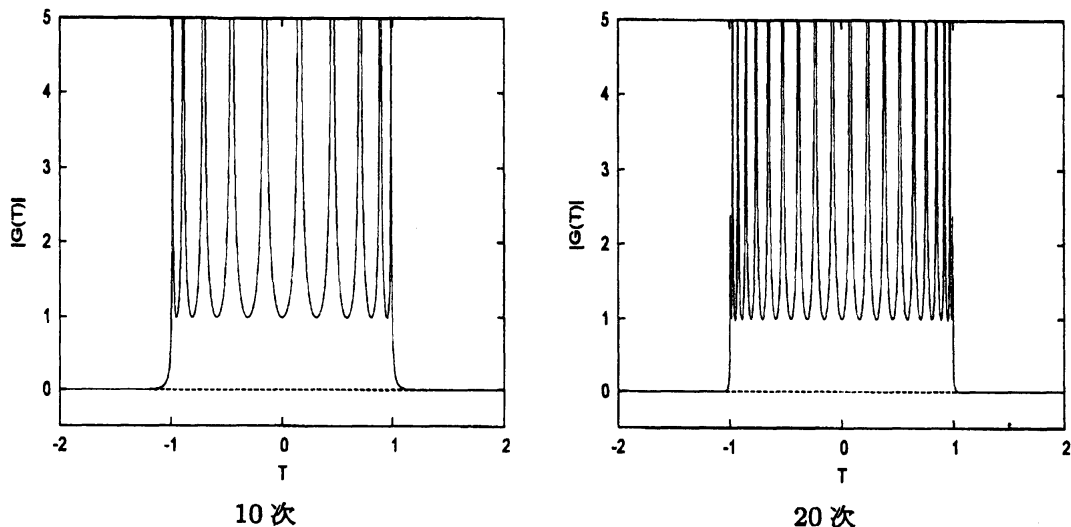
### 3.3 フィルタ $\mathcal{F}$ の透過特性 $f(\lambda)$

$\mathcal{F}$  をフィルタとみなす場合、 $f(\lambda)$  はフィルタの透過特性で、 $\mathbf{u}^{(\mu)}$  成分が  $f(\lambda^{(\mu)})$  倍になることを表す。有理関数  $f(\lambda)$  を多項式の商で表すと、分母は丁度  $k$  次で  $c_k$  を定数として  $\varphi(\lambda) \equiv c_k \prod_{i=1}^k (\lambda - \rho_i)$  となり、分子の次数は分母よりも常に低い ( $k' < k$ ) 次とする。よって  $f(\lambda)$  は遠方  $|\lambda| \rightarrow \infty$  では減衰し、 $f(\lambda) = O(\lambda^{-(k-k)})$  である。 $(f(\lambda)$  の  $\ell (> 0)$  階の導関数では  $O(\lambda^{-(k-k'+\ell)})$  で、減衰がより速くなる。)  $f(\lambda)$  の遠方での減衰が最も速いのは分子が定数の場合である。そこで以下では、分子を定数 1 に限定し、 $f(\lambda) \equiv 1/\varphi(\lambda)$  とする。 $1/\varphi(z) = \sum_{i=1}^k \omega_i / (z - \rho_i)$  から、 $\omega_i = 1/\varphi'_k(\rho_i)$ 。よって  $k$  個の相異なる分点  $\{\rho_i\}$  を指定すると、定数倍を除いて  $\mathcal{F}$  が完全に決まる。

フィルタを用いて「固有値が領域  $\mathcal{D}$  の外にある固有ベクトル」の成分だけをなるべく減衰させたい。そのためには  $|f(\lambda)|$  の値は領域  $\mathcal{D}$  の内では小さくなく、外ではなるべく小さくなるようにする。言い替えれば  $|\varphi(\lambda)|$  の値は領域  $\mathcal{D}$  内では大きくなく、外ではなるべく大きくなるようにする。 $f(\lambda)$  の分母に「絶対値が領域内  $\mathcal{D}$  でだけなるべく小さい」性質を持つ多項式  $\varphi(\lambda)$  を採用すれば良い。実区間の近傍でそのような性質を持つ典型的な多項式としては、関数の最良近似理論で良く知られた Chebyshev 多項式がある。

**例 (Chebyshev 多項式)**  $\varphi(\lambda) = T_k(t)$  ( $k$  次 Chebyshev 多項式) にとる。但し  $t$  は  $\lambda$  の実区間  $\mathcal{I} = [\alpha, \beta]$  に対する「相対座標」 $(2\lambda - \alpha - \beta)/(\beta - \alpha)$  (区間中心で 0, 両端で  $\pm 1$ ) である。 $\lambda \in \mathcal{I}$  の場合は  $|f(\lambda)| \geq 1$  で、 $\lambda \notin \mathcal{I}$  の場合  $|t| \gg 1$  で  $|f(\lambda)| \approx 2^{-(k-1)} |t|^{-k}$  となる。零点の相対座標は  $t_\ell = \cos \theta_\ell$ ,  $\ell = 1, 2, \dots, k$ , 但し  $\theta_\ell = (2\ell - 1)\pi/(2k)$  で、その位置にフィルタの透過特性  $f(\lambda) = 1/T_k(t)$  の極がある (図 1)。

図 1: Chebyshev 多項式を分母にもつフィルタ特性  $|f(\lambda)|$  (実軸上, 相対座標)。



### 3.4 共鳴の引き起こす数値計算上の困難

分母  $\varphi(\lambda)$  が固有値と一致もしくは極めて近接する零点を持つと、フィルタの出力の中で共鳴した固有ベクトル成分だけが支配的となり、他の固有ベクトル成分の精度は丸め誤差で失われる。その結果、部分空間

$S$  の近似品質が劣化する。そこで、演算精度が限られている計算の場合には、領域  $D$  内で  $|f(\lambda)|$  の最大と最小の比（一種の条件数）を抑えることが必要になる。

実軸上に於いて、分母が Chebyshev 多項式のフィルタの透過特性は、指定した実区間の外では急減衰で、不要な成分の除去性能は良好だが、固有値と実区間内の極が共鳴するリスクがある。

実対称な固有値問題では固有値は必ず実数なので、実数の零点がない多項式をフィルタの分母に用いれば共鳴を避けられる。ところが非対称問題では行列が実でも固有値は一般には複素数なので、固有値の分布状況を知らない段階で事前に分母の多項式の零点を設定して固有値と近接がないことを保証することは難しい。実数の零点がない多項式を用いれば実固有値との共鳴は回避できても、複素固有値と一致または近接する零点が共鳴を起こすリスクは残る。以下で「実数の零点を持たない多項式」を 2 種類例示しておく。

**例 1 (円分多項式)** 円分多項式  $\varphi(\lambda) \equiv 1 + t^k$  をとる (但し  $k$  は偶数)。実区間  $[-1, 1]$  内では  $1 \leq \varphi(\lambda) \leq 2$  で最大と最小の比は 2 である。零点は複素平面上の原点を中心とする半径 1 の円周上の  $k$  等分点である。零点の相対座標:  $t_\ell = \cos \theta_\ell + i \sin \theta_\ell$ ,  $\ell=1, 2, \dots, k$ . 但し,  $\theta_\ell = (2\ell - 1)\pi/k$ . フィルタ特性を図 2,3 に示す。

図 2: 例 1 (円分多項式) フィルタ特性  $f(\lambda)$  (実軸上, 相対座標)。

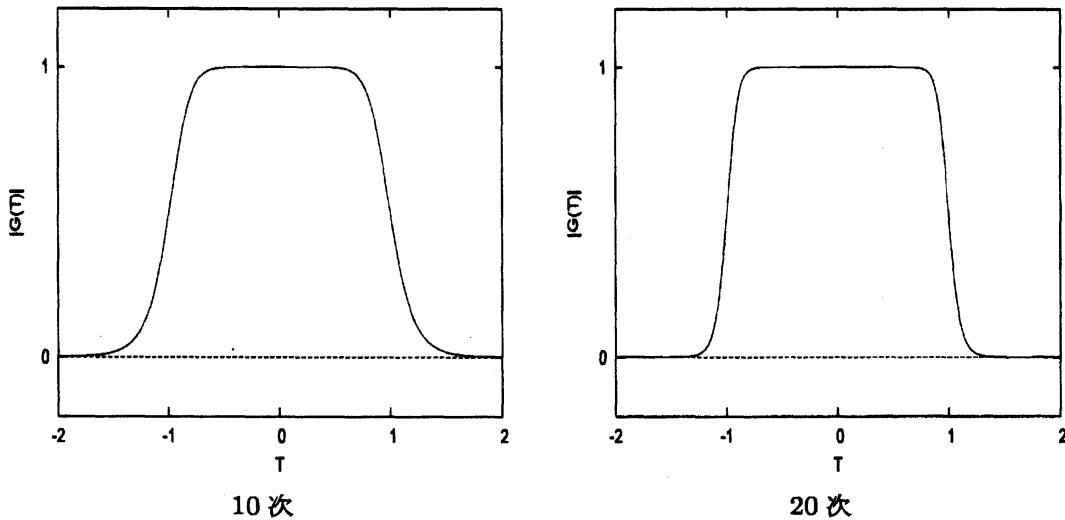
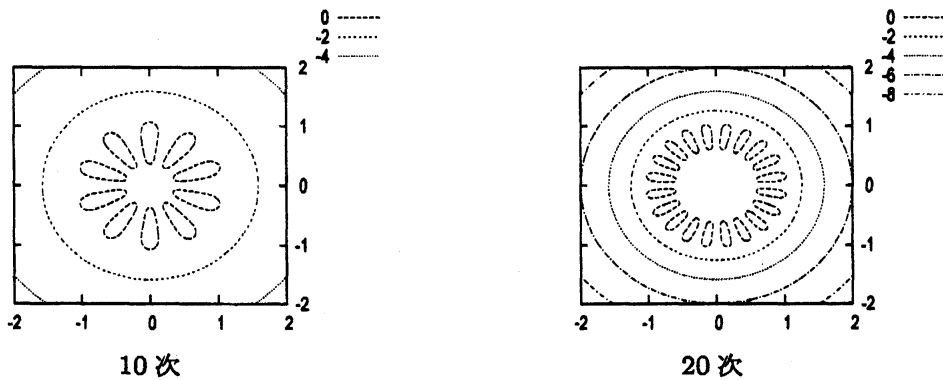


図 3: 例 1 (円分多項式)  $\log_{10} |f(\lambda)|$  の等高線 (複素平面上, 相対座標)。



**例 2 (値シフト Chebyshev 多項式)** 多項式は  $\varphi(\lambda) \equiv (T_k(t) + 1 + 2\gamma) / (2\gamma)$  ( $k$  は偶数,  $\gamma > 0$ ) で, 実区間  $[-1, 1]$  内で  $1 \leq \varphi(\lambda) \leq 1 + \gamma^{-1}$  を満たし, 最大と最小の比は  $1 + \gamma^{-1}$  である. フィルタの不要成分の漏出は  $\gamma$  に比例するので, 部分空間の構成用には  $\gamma \approx 1$  が適切である. 零点は複素平面上の実区間の両端を焦点とする楕円の周上にある. 零点の相対座標は  $t_\ell = \cosh \tau \cdot \cos \theta_\ell + i \sinh \tau \cdot \sin \theta_\ell$ ,  $\ell = 1, 2, \dots, k$ . 但し,  $\tau = (1/k) \cosh^{-1}(1 + 2\gamma)$ ,  $\theta_\ell = (2\ell - 1)\pi/k$  である.  $\gamma = 1$  の場合のフィルタ特性を図 4,5 に示す.

図 4: 例 2 (値シフト Chebyshev 多項式 ( $\gamma=1$ )) フィルタ特性  $f(\lambda)$  (実軸上, 相対座標).

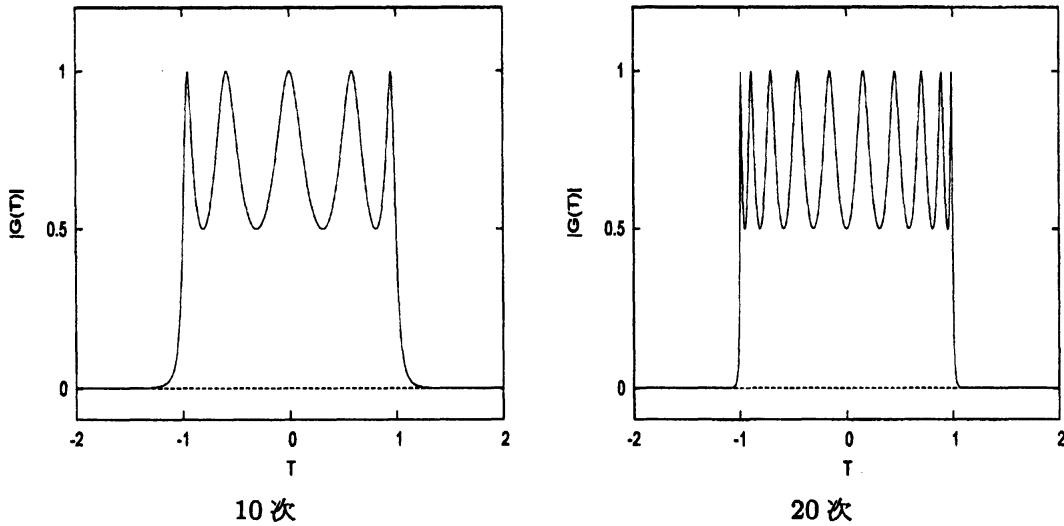
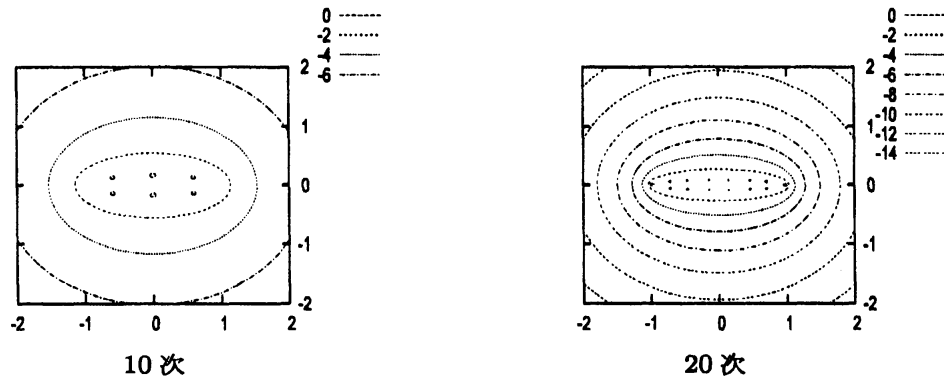


図 5: 例 2 (値シフト Chebyshev 多項式 ( $\gamma=1$ ))  $\log_{10} |f(\lambda)|$  の等高線 (複素平面上, 相対座標).



### 3.5 複素平面上での分点の一般的な配置と共鳴の困難の事後の回避の処方

分点  $\rho_i$  の配置は比較的自由であり, 必ずしも線分や円, 楕円などの曲線上に分布させなくてもよい. 固有値を求めたい領域を  $\mathcal{D}$  とするとき, 関数  $|f(\lambda)|$  の値が,  $\mathcal{D}$  の内では相対的に大きく,  $\mathcal{D}$  の外では極めて小さくなればよい. (演算精度が限られている数値計算では, 分点の配置はフィルタの出力ベクトルの数値誤差の特性にも影響がある. この観点からの望ましい分点の配置の理論については今後の課題である.)

複素平面上でフィルタの分点  $\rho_i$ ,  $i = 1, 2, \dots, k$  全体を並行移動, 拡大縮小, 回転で座標変換すると, 線形結合の係数  $\omega_i$  は高々一斉に定数倍される. 分点の組と重み  $\rho_i, \omega_i$ ,  $i = 1, 2, \dots, k$  を標準的な座標で用意

しておけば、座標の複素線形変換で移動し利用できる。

共鳴の困難を事後に避けるには、共鳴に近いレゾルベントだけを分点をずらし再計算すれば簡単だが、並列処理では不利なので、共鳴に近いレゾルベントだけを除去して、フィルタの係数を作り直す方法を述べる。

いま除去前のフィルタの特性関数を  $f(x) \equiv \sum_{i=1}^k \omega_i / (x - \rho_i)$  とすると、 $f(x)$  は多項式の商の形で、 $f(x) = \psi(x) / \varphi(x)$  と書いて、分母は  $\varphi(x) \equiv \prod_{i=1}^k (x - \rho_i)$  となる。遠方  $|x| \rightarrow \infty$  で最も強く減衰する条件を課すと分子  $\psi(x)$  は定数で、それを 1 と置くと係数  $\omega_i = 1 / \varphi'(\rho_i)$  である。分点  $\rho_i$  の添字集合を  $\mathcal{G}$  と書くとき、 $\varphi(x) = \prod_{i \in \mathcal{G}} (x - \rho_i)$  で、係数は  $\omega_i = 1 / \varphi'(\rho_i) = 1 / \prod_{j \in \mathcal{G} - \{i\}} (\rho_i - \rho_j)$  と決まるのであった。

共鳴の困難の回避の為に、ある閾値以上の増幅率を持ったレゾルベントをフィルタの構成から除外する方法で、シフト量の分点の添字集合  $\mathcal{G}$  から除外する分点の添字集合を  $\mathcal{E}$  と書き、 $\hat{f}(x) = \sum_{i \in \mathcal{G} - \mathcal{E}} \hat{\omega}_i / (x - \rho_i)$  とすれば、遠方で最も強く減衰するように、 $\hat{f}(x) = 1 / \hat{\varphi}(x)$  とすると、 $\hat{\varphi}(x) = \prod_{i \in \mathcal{G} - \mathcal{E}} (x - \rho_i)$ 。添字  $i \in \mathcal{G} - \mathcal{E}$  での係数  $\hat{\omega}_i = 1 / \hat{\varphi}'(\rho_i)$  が  $\omega_i = 1 / \varphi'(\rho_i)$  を用いて  $\hat{\omega}_i = 1 / \hat{\varphi}'(\rho_i) = 1 / \prod_{j \in \mathcal{G} - \mathcal{E} - \{i\}} (\rho_i - \rho_j) = \omega_i \prod_{j \in \mathcal{E}} (\rho_i - \rho_j)$  と決まる。これは、分点を除去する前のフィルタの係数に補正項をかけた形になっている。

## 4 係数 $A - \rho I$ の線形方程式の解法について

レゾルベントの線形結合を用いたフィルタによる固有値解法では、係数  $A - \rho I$  を持つ線形方程式が高速に能率良く解けることが本質的に重要である。一変数方程式の場合の Frobenius の随伴行列や一般化随伴行列は、帯行列に列を 1 つ付け加えた構造を持ち、帯行列用の LU-分解法 [6] と類似した能率的な LU-分解法がある。

LU-分解で  $N$  次行列  $A - \rho I$  の線形方程式を解く演算量は：

- $A$  の Hessenberg 性だけで疎性を仮定しない場合は、行交換付きの LU-分解は  $O(N^2)$ 、LU-分解後の方程式の解法は  $O(N^2)$  である。
- $A$  が一変数多項式の Frobenius の随伴行列の場合は、行交換付きの LU-分解は  $O(N)$ 、LU-分解後の方程式の解法は  $O(N)$  である。
- $A$  が一変数多項式の直交多項式表現による一般化随伴行列の場合は、行交換付きの LU-分解は  $O(N)$ 、LU-分解後の方程式の解法は  $O(N)$  である。

(注：これら LU-分解による解法では、誤差による数値安定性までが保証されているわけではない。)

多変数方程式系の随伴行列でも疎性を活かした LU-分解はある程度実施可能であろう。しかし、行列  $A - \rho I$  の線形方程式を直接法で解く際に大規模疎行列  $A$  の構造が十分活用出来ない場合には、消去過程での fill-in による非零要素の増加から実施が困難となり、行列の変形を伴わない反復法系の解法がおそらく必要になるであろう。例えば CG 法などの Krylov 部分空間法の系統の方法を用いれば、行列の変形なしで、非零要素が 1 行あたり平均  $f$  個の  $N$  次行列の線形方程式を 1 回当たり実数演算量  $O(fN)$  の処理を  $N$  回繰り返すと、正確な解が得られる。すなわちトータルの演算量  $O(fN^2)$  で解ける。実際は演算には数値誤差が入るため、解法は繰り返し  $N$  回では正解に到達せず、必要な精度に到達するまでには反復がより多く必要、もしくは収束しない。反復法により疎行列の線形方程式を解く技術は現在も発展の途上であり、収束性を高める為の前処理法や、Krylov 部分空間法の系統で対角シフトが異なる複数の問題を能率良く解く方法 [18] などが開発されつつある。

**例：一変数方程式の Frobenius 随伴行列の場合** 右 Hessenberg 形の「Frobenius 随伴行列」の形と、行交換を行う LU-分解を用いる場合の  $A - \rho I$  の LU-分解の記憶箇所を ×印で示す。最後の列以外は帯三重対角形である（随伴行列  $A$  に balancing が施されていても同様）。LU-分解の記憶に用いる非零要素数は  $4N - 4$



である.

$$\begin{bmatrix} 0 & & & -a_0 \\ 1 & 0 & & -a_1 \\ & 1 & \ddots & -a_2 \\ & & \ddots & 0 \\ & & & 1 & -a_{N-1} \end{bmatrix}, \begin{bmatrix} \times & \times & & & \times \\ \times & \times & \ddots & & \times \\ & \times & \ddots & \times & \times \\ & & \ddots & \times & \vdots \\ & & & \times & \times \end{bmatrix}$$

例：一変数方程式の一般化随伴行列の場合 右 Hessenberg 形の「一般化随伴行列」の形と，行交換を行う LU-分解を用いた場合の， $A - \rho I$  の LU-分解の記憶箇所を×印で示す（一般化随伴行列  $A$  に balancing が施されていても同様）．最後の列以外は，対角線の下側 1 本，上側 2 本の帯行列である．LU-分解の記憶に用いる非零要素の数は  $5N - 7$  になる．

$$\begin{bmatrix} 0 & 1/2 & & & \gamma_0 \\ 1 & 0 & 1/2 & & \gamma_1 \\ & 1/2 & 0 & \ddots & \gamma_2 \\ & & 1/2 & \ddots & 1/2 \\ & & & \ddots & 0 \\ & & & & 1/2 & \gamma_{N-1} \end{bmatrix}, \begin{bmatrix} \times & \times & \times & & \times \\ \times & \times & \times & \ddots & \times \\ & \times & \times & \ddots & \times \\ & & \times & \ddots & \times \\ & & & \ddots & \times \\ & & & & \times & \times \end{bmatrix}$$

例：LU-分解（行列  $A$  は必要な要素だけを記憶）．分解後の  $LUx = b$  の解法（ $x$  と  $b$  は記憶を共用）．

```

! 右 Hessenberg 形の随伴行列専用 LU-分解：
! A --> LU, (行ピボット交換).
! 右半帯幅 rh=0 は「Frobenius」,
! rh=1 は「一般化」の随伴行列.
do k = 1, n-1
  amx=abs(A(k,k)); ip=k
  if (abs(A(k+1,k)) > amx) then
    amx=abs(A(k+1,k)); ip=k+1
  endif
  ipiv(k) = ip
  if (ip /= k) then
    do j = k, min(k+1+rh, n-1)
      t=A(k,j); A(k,j)=A(ip,j); A(ip,j)=t
    enddo
    t=A(k,n); A(k,n)=A(ip,n); A(ip,n)=t
  endif
  A(k+1,k)=A(k+1,k)/A(k,k)
  do j = k+1, min(k+1+rh, n-1)
    A(k+1,j)=A(k+1,j)-A(k+1,k)*A(k,j)
  enddo
  A(k+1,n)=A(k+1,n)-A(k+1,k)*A(k,n)
enddo

! 前進消去: L y = b.
do k = 1, n-1
  ip=ipiv(k)
  t=x(ip); x(ip)=x(k); x(k)=t
  x(k+1)=x(k+1)-A(k+1,k)*t
enddo

! 後退代入: U x = y.
do k = n, 1, -1
  t=x(k)/A(k,k); x(k)=t
  if (k == n) then
    do i = n-1, 1, -1
      x(i)=x(i)-A(i,n)*t
    enddo
  else
    do i = k-1, max(k-1-rh, 1), -1
      x(i)=x(i)-A(i,k)*t
    enddo
  endif
enddo

```

例：モニックな行列多項式の方程式と随伴行列  $A_i$  は  $p$  次の正方行列で  $A_n = I_p$ .  $P(x) \equiv A_0 + A_1x + \dots + A_{n-1}x^{n-1} + I_p x^n$  とする.  $\det P(x) = 0$  の解は，Frobenius 類似の  $n \times p$  次行列  $C$  の固有値であり， $(C - \lambda I)$  を係数とする線形方程式は，通常の Frobenius 随伴行列の LU-分解法と同様にブロック単位の消去法で効

率良く行える。

$$C = \begin{bmatrix} 0 & & & -A_0 \\ I_p & \ddots & & \vdots \\ & & \ddots & 0 \\ & & & I_p \end{bmatrix} \begin{bmatrix} -A_0 \\ \vdots \\ -A_{n-2} \\ -A_{n-1} \end{bmatrix}$$

## 5 数値計算例

200次方程式を  $P(z) = z^n + c_2 z^2 + c_1 z + c_0 = 0$  とする。但し,  $n = 200$ ,  $c_2 = -0.81078$ ,  $c_1 = -9.0617301$ ,  $c_0 = 10.53771414908$ .  $P(z)$  の Frobenius 随伴行列を balancing せずに固有値を求めた。フィルタを「値シフト Chebyshev」型で, 次数  $k = 30$ , パラメタ  $\gamma = 1$  とし, 実軸に近い  $z = 1$  付近の零点を求めるべく区間を  $[0.8, 1.2]$  に設定した。複素乱数から作り Hermite 直交化した初期ベクトルの個数を 20 とし, 特異値切断の相対閾値は  $10^{-5}$  とした。フィルタ出力の組  $Y = \mathcal{F}X$  の特異値分解で得た 20 個の特異値の例を表 1 に掲げる。相対閾値  $10^{-5}$  での  $Y$  の有効階数  $r'$  は 6 で, 切断後に残った 6 個の特異ベクトルの組が張る部分空間に対して Rayleigh-Ritz の方法で近似固有対を求めた。その Rayleigh 商を表 2 に掲げる。

フィルタ対角化法の近似固有対を初期値とする (各対で独立の) 逆反復法による改良過程を表 3 に示す。表中で  $Q$  は近似固有ベクトルの Rayleigh 商を, RES は 2-ノルムで規格化した近似固有ベクトルの残差を表す。フィルタ対角化法の与えた初期近似対は固有値が実軸から遠いほど残差 (逆反復法の初期値は表中で反復回数 0 の行, その RES の値) が大きい傾向がある。これは固有値が実軸から離れると, 値シフト Chebyshev 多項式フィルタの透過率が減少し, 出力中の固有ベクトル成分の有効桁が丸め誤差の影響で減るからである。

本計算では行列が実であることを全く利用していない。また, レゾルベントの作用や逆反復法に必要な線形方程式は, 上述の対角シフト付き随伴行列用の高速な解法を用いたが, 行列や  $LU$ -分解に必要な非零要素だけを記憶する最適化は (行列の次数が大きいと極めて重要になるが) 施していない。この例の経過時間は, intel Core 2 Duo E6600 2.4GHz 4MB L2 のシステムの CPU コア 1 個を用いて IEEE 64bit 浮動小数点数により, フィルタ対角化で 6 個の近似固有対を求めるまでの時間が 10.6 ミリ秒, その後 6 個の近似固有ベクトル全てに逆反復を 2 回ずつ適用して改良を施す時間が 2.2 ミリ秒であった。

表 1:  $Y$  の特異値分布の例.

$\sigma_1 = 4.9E-01$	$\sigma_6 = 1.7E-05$	$\sigma_{11} = 7.2E-09$	$\sigma_{16} = 4.0E-09$
$\sigma_2 = 2.8E-01$	$\sigma_7 = 3.1E-07$	$\sigma_{12} = 6.4E-09$	$\sigma_{17} = 3.2E-09$
$\sigma_3 = 2.2E-03$	$\sigma_8 = 2.0E-07$	$\sigma_{13} = 5.8E-09$	$\sigma_{18} = 2.8E-09$
$\sigma_4 = 1.3E-03$	$\sigma_9 = 1.5E-08$	$\sigma_{14} = 5.3E-09$	$\sigma_{19} = 2.2E-09$
$\sigma_5 = 2.4E-05$	$\sigma_{10} = 1.0E-08$	$\sigma_{15} = 4.1E-09$	$\sigma_{20} = 1.5E-09$

表 2: フィルタ対角化の与えた近似固有対の Ritz 値.

対の番号	Ritz 値
1	(0.997372281698081, -0.07427221462633841)
2	(0.997586635724274, 0.07422785353169536)
3	(0.998112662560635, -0.04406704931348985)
4	(0.99811446614153, 0.04406693534181548)
5	(0.998131634646873, -0.01456568588515524)
6	(0.998131629996097, 0.01456568769251575)

表 3: 逆反復法 (Rayleigh 商シフト) による近似固有対の改良.

対の番号	反復回数	Q	RES
1	0	(0.99737228169808, -0.07427221462634)	1.0E-03
	1	(0.99750990459796, -0.07424092833867)	1.2E-06
	2	(0.99750964074953, -0.07424067033468)	8.3E-12
2	0	(0.99758663572427, 0.07422785353170)	4.3E-04
	1	(0.99750970896572, 0.07424078541512)	3.5E-07
	2	(0.99750964074733, 0.07424067033387)	9.8E-13
3	0	(0.99811266256063, -0.04406704931349)	9.5E-06
	1	(0.99811271253546, -0.04406587763153)	1.2E-10
	2	(0.99811271257415, -0.04406587763941)	3.0E-16
4	0	(0.99811144661415, 0.04406693534182)	1.2E-05
	1	(0.99811271258096, 0.04406587763006)	1.7E-10
	2	(0.99811271257415, 0.04406587763941)	2.3E-16
5	0	(0.99813163464687, -0.01456568588516)	3.9E-08
	1	(0.99813163796875, -0.01456568522496)	1.5E-15
	2	(0.99813163796875, -0.01456568522496)	1.7E-16
6	0	(0.99813162999610, 0.01456568769252)	4.8E-08
	1	(0.99813163796875, 0.01456568522496)	4.3E-15
	2	(0.99813163796875, 0.01456568522496)	3.5E-16

## 参 考 文 献

- [1] Eberlein, P. J.: Solution to the Complex Eigenproblem by a Norm Reducing Jacobi Type Method, *Numer. Math.*, Vol.14 (1970), pp.232-245.
- [2] Edelman, A. and Murakami, H.: Polynomial Roots from Companion Matrix Eigenvalues, *Math. Comp.*, Vol.64, No.210 (1995), pp.763-776.
- [3] Ericsson, T. and Ruhe, A.: The spectral transformation Lanczos method for the numerical solution of large sparse generalized symmetric eigenvalue problems, *Math. Comp.*, Vol.35 (1980), pp.1251-1268.
- [4] Golub, G. H. and van Loan, C. F.: *Matrix Computations, 3rd Ed.*, The Johns Hopkins Univ. Press (1983).
- [5] Stetter, H. J.: *Numerical Polynomial Algebra*, SIAM, Philadelphia (2004).
- [6] Martin, R. S. and Wilkinson, J. H.: Solution of symmetric and unsymmetric band equations and the calculations of eigenvectors of band matrices, *Numer. Math.*, Vol.9 (1967), pp.279-301.
- [7] 松本 純一: Krylov 部分空間反復法を用いた Arnoldi 法による有限要素並列固有値解析, 日本応用数理学会論文誌, Vol.15, No.2 (2005), pp.145-158.
- [8] 村上 弘: 一変数代数方程式の行列固有値解法について, 京都大学数理解析研究所講究録, No.1395 (2004), pp.198-204.
- [9] 村上 弘: 一変数代数方程式の行列解法とその周辺, 情報処理学会研究報告 2005-HPC-102 (2005), pp.1-6.
- [10] 村上 弘: 代数方程式の多項式基底展開と行列固有値解法, 情報処理学会研究報告 2006-HPC-106 (2006), pp.13-18.
- [11] 村上 弘: 随伴行列法による非線型方程式の解法, 第 35 回数値解析シンポジウム NAS2006, 講演予稿集 (2006), pp.21-24.

- [12] 村上 弘: 区間内での直交多項式展開による非線形方程式の求解, 日本数式処理学会誌, *J.JSSAC*, Vol.14, No.2 (2007), pp.23–26.
- [13] 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, Vol.49, SIG 2 (ACS 21) (2008), pp.66–87, (to appear).
- [14] 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).
- [15] 村田 健郎: 「線形代数と線形計算法序説」, サイエンス社 (1986).
- [16] 小国力 編, 村田 健郎・三好 俊郎・ドンガラ J.J.・長谷川 秀彦 共著: 「行列計算ソフトウェア」, 丸善 (1991).
- [17] Sakurai, T. and Sugiura, H.: A Projection Method for Generalized Eigenvalue Problems, *J. Comput. Appl. Math.*, Vol.159 (2003), pp.119–128.
- [18] 多田野 寛人・櫻井 鉄也: 一般化固有値問題で現れる複素対称連立一次方程式に対する反復解法の性能評価, 第 35 回数値解析シンポジウム講演予稿集 (2006), pp.61–64.
- [19] Toledo, S. and Rabani, E.: Very Large Electronic Structure Calculations Using an Out-of-Core Filter-Diagonalization Method, *J. Comput. Physics*, Vol.180 (2002), pp.256–269.
- [20] Jia, Zhongxiao and Stewart, G. W.: An Analysis of the Rayleigh-Ritz Method for Approximating Eigenspaces, *Math. Comp.*, Vol.70, No.234 (2000), pp.637–647.