

78

対称帯行列固有値解析

— ページスワップに注目して —

日立 中研 村田 健郎

1. はじめに. 最近, 計算機の内部演算能力の向上に比べて, ページスワップ時間の短縮の方には見るべき進歩がないために, 待ち時間ネックが益々目立ってきた。流れ場や, 電磁場の偏微分方程式を有限要素法で解くとか, 固有値問題を解こうとかのとき特に著しい。ここでは, 代表的な対称帯行列固有値解析アルゴリズムについての, そういう観点からの見直しを行い, 併せてそのうちの二つ, すなわちスツルム・逆反復法系統に属するものと帯QR法についての改良アルゴリズムを提示したい。

初めに, 周知のスツルム・逆反復法を例にとつて, 問題の捉え方を例示したい。対称帯行列固有値問題 $Ax = \lambda x$ について考える。帯半巾を m , 元数を n とする。言葉を節約するために A は正定値とする。与えられた数 α より小さい固有値と固有ベクトル:

$$0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p < \alpha$$

$$v_1, v_2, \dots, v_p$$

を求めたいとする。大筋を示そう。

[1] $A - \alpha I$ を $U^T D U$ 分解して D の負要素の個数 $c(\alpha)$ をスツルム法によって求める。それが p である。 $p = c(\alpha)$ 。

[2] バイセクション法を施行する： 分点 α_k 毎にスツルム法を使用し、すなわち $A - \alpha_k I = U^T D U$ を行って、 D の負要素の個数 $c(\alpha_k)$ を求めることによって、各 $\lambda_1, \dots, \lambda_p$ を分離する。これで各 λ_k 毎にその下界と上界： $\alpha_{uk} \leq \lambda_k < \alpha_{lk}$ が判る。

[3] 各 k 毎に、 $\alpha_k = (\alpha_{uk} + \alpha_{lk})/2$ をシフト量にえらび、 $A - \alpha_k I$ を $U^T D U$ 分解した結果を使用して逆反復を行う：

$$\left[\begin{array}{l} (A - \alpha I) b^{(l)} = w^{(l-1)}; \quad eig = \alpha_k + 1/w^{(l-1)T} b^{(l)} \\ w^{(l)} = b / (b^T b)^{\frac{1}{2}}; \quad \text{収束判定}; \quad l = l+1 \end{array} \right.$$

[4] 上で求めた eig から改めて $A - eig I$ を作り、それを $U^T D U$ 分解した結果を使用しての逆反復によって固有ベクトルを求める。最後に固有値も再びレーリー商によって更新する。（[4] は、粗い精度でよいときには省略される。）

スツルム・逆反復法と呼ばれる計算法の大筋は上述の如くであるが、問題点を整理しよう。

(1) CPU タイムと USE タイムに関し：

上述の方法に関し、従来言われて来たことは、 $U^T D U$ 分解のための $\frac{1}{2} m^2 n$ 回積和（ガウスなら $2 m^2 n$ ）が一体何回必要かが主たる関心事であって、逆反復法に入ってから以後の、

反復1回当りの $2mn$ 回積和 (ガウスなら $3mn$) は大したことない, ということであった。

CPU タイムについては、大筋はその通りである。ところがページスワップに関しては、 $U^T D U$ 分解のためには1回当り $mn/500$ 回 (ガウスなら $3mn/500$) , 逆反復のためには1回当り $2mn/500$ 回 (ガウスなら $3mn/500$ あるいは $6mn/500$ 回)*, 逆反復回数の方も大いに問題になる。

[例] 積和 $1\mu s$, ページスワップ1回当り $20ms$ の機械で、 $m=80$, $n=8000$ のとき、1回当りの CPU タイムとページスワップタイムは、

$U^T D U$ 25 sec : 25 sec 逆反復 0.3 sec : 50 sec

である。スツルムは、 λ_1 のため $10\sim 20$ 回程度、あと $\lambda_2, \dots, \lambda_p$ のためにはひとつ当り平均 $3\sim 5$ 回程度が普通である。いっぽう、逆反復の方は、よく分離された固有値に対しては $3\sim 10$ 回程度であるが、近接固有値に対しては逆反復が 50 回 ~ 100 回にも及ぶことが稀ではない。

[2] 計算の信頼性・安全性 :

特に近接固有値が密集している場合、上述の方法そのままでは収束が甚しく遅いことがある。ときには真値からほど遠い値に '収束' したかの如くふるまったり、あるいは時間制限にか、つて途中で打ちきられることがしばしばである。

この事情のため、汎用の数値計算ライブラリには採用し難いということになっている。

* 本稿では、もつぱら、 mn 語が、そのプログラムに割り当てられた実メモリ容量をはるかに超える場合について議論する。

2. $Ax = \lambda Mx$ に対するスツルム・逆反復法の改良.

初めに、改良の要旨を示そう：

[1] STURM プログラムを、ページスワップ負擔が逆反復のためのガウスとくらべて無視できるほど小さくなるよう、巧妙なプログラムを書く。BISECT と INVRAY で使用。

[2] BISECT プログラム： ϵ_{psb} を与え、その大きさに応じて固有値を分離する。(STURMを使う。目新しいことはない。)

[3] INVRAY プログラム： 次の三つの部分から成る。

(1) (BISECTによって) 単純根と判定された固有値に対し、再度スツルムを2回行って、シフト量 α を決定する。後続の二段の逆反復過程を、安全かつ効果的に行わせるためのものである。

(2) シフト量を、各分離区間毎に固定した通常の逆反復を行う。当該区間のそとの固有値に属する固有ベクトル成分を減衰させるためのものである。単純固有値のあるものについては、この段で固有値、固有ベクトル共に 所望の精度に達することがある。(特に目新しいことはない。)

(3) レーリーシフトつき逆反復： アルゴリズムに目新しいことはないが、使用するガウス消去法プログラムに工夫がある。

[4] 全体のコントロール： CPUタイムとUSEタイムの何れを重視するかに応じたコントロールが可能であるようにする。

2.1 STURM と BTSECT

オーソドックスな Martin·Wilkinson のアルゴリズムによることにする⁽¹⁾。その原理：

α より小さい $Ax = \lambda Mx$ の固有値の個数は、主座小列式：
 $\det(A_k - \alpha M_k)$, $k=1, \dots, n$ の符号変化の回数 $C(\alpha)$ に等しい。

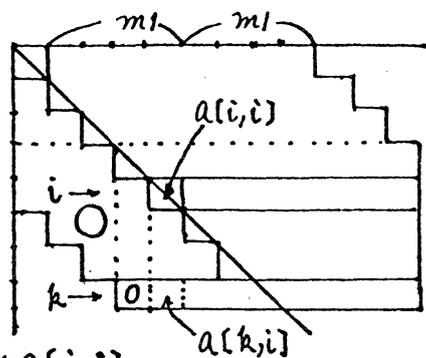
この原理による Martin·Wilkinson のアルゴリズムは、

STURM $a[i, j] = a_{i,j} - \alpha * m_{i,j}$ として、

```

P[1] = sign(a[1,1]) ; ip = 1 ; count = 0
do k = 2, n
  imin = max(1, k-m1) ; jmax = min(k+m1, n)
  do i = imin, k-1
    if abs(a[i,i]) < abs(a[k,i]) then
      ip = -ip
      do j = i, jmax
        a[i,j] = a[k,j]
      if a[i,i] ≠ 0 then
        t = -a[k,i]/a[i,i]
        do j = i, jmax
          a[k,j] = a[k,j] + t * a[i,j]
    p[k] = ip * Πi=1k sign(a[i,i])
    if P[k-1] * P[k] < 0 then count = count + 1

```



上のスツルムを使って、 α より小さい固有値の個数 $p = \text{count}$ を求めたのち、 $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p < \alpha$ なる λ_k を epsbc の粗さで分離するためのプログラム BISECT は次のようになる：

BISECT

$xu[k], xl[k], nc[k], k=1, \dots, p$ を用意する。

```

do k=1, p
  xl[k] = dl ; nc[k] = p ; xu[k] = 0.0
k=1
repeat
  repeat
     $\alpha = (xu[k] + xl[k]) / 2$  ; Call STURM
    do i=k, count
      xl[i] =  $\alpha$  ; nc[i] = count
    do i=count+1, p
      xu[i] = max( $\alpha, xu[i]$ )
    until  $k \geq nc[k]$  or  $xl[k] - xu[k] \leq \epsilon p b_i$ 
  if  $k = nc[k]$  then
    k = k+1
  else
    do i=k, count
      xu[i] = xu[k]
    k = count+1
  until  $k > p$ 

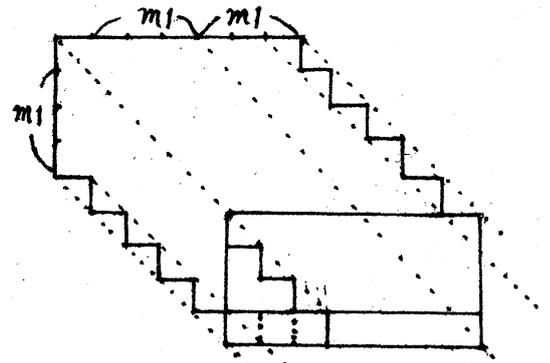
```

STURM のプログラムについて。

スツルムにおいては、欲しいのは Count 値だけであって、三角分解結果は不要である。そこで、ページスワップを減らすために、約 $2m^2$ 語のワークエリアを用意して、それをサイクリックに使うことによって目的を達するようにプログラムを書く。また、 $A - \alpha M$ を先に作ってしまうのではなく、各 k 段のガウス過程の直前に $k+m_1$ 行の $a_{k+m_1, j} + \alpha \cdot m_{k+m_1, j}$ を作る。

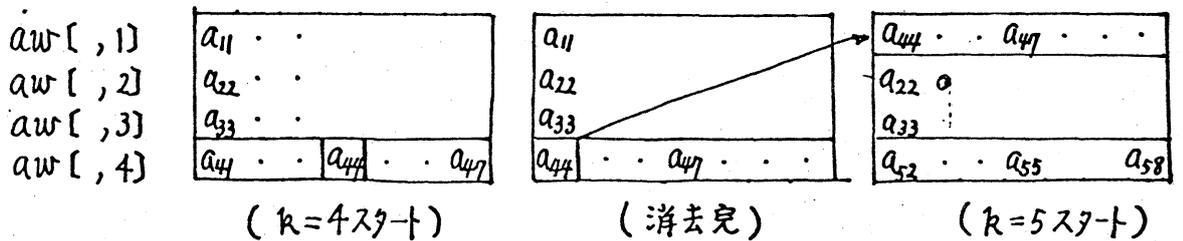
プログラムの解読を助けるために、 $m_1=3$ の場合について例示しよう。

消去の度に左シフトという技巧を使う。 $k=1, 2, 3$ までは特別だが簡単で、下図左のようになる。



$k=4$ の段 : 4番の方程式の係数 $a_{4,j} - \alpha * m_{4,j}$ を $aw[,4]$ に作り、これに対する消去作業を行う。消去の度に左シフトを行う。完了したらそれを $aw[,1]$ に収める。

$k=5$ の段 : $aw[,2], aw[,3], aw[,1]$ の順にピボット方程式が使われる。



$k=m_1$ までのところは簡単だから省略して、 $k=m_1+1$ 以降のプログラムを示そう。(次頁)

[ノート] 差分法や有限要素法によって作られた行列 A や M は、帯のふちと、真中三本がノンゼロで、あと帯の中もゼロというようになっているから、これらは、インデックス付きのパックされたアレイに収めておき、消去作業の直前に $aw[,m]$ に $aw[j-k+m, m] = a_{k,j} - \alpha * m_{k,j}$ とやって作る。({1} の部分)

$m = m1 + 1$; $mm = 2m1 + 1$; $eps = 1.0 * 10^{-16}$

```

do k = m1 + 1, n
  jmax = min(n, k + m1)
  do j = k - m1, jmax
    aw[j - k + m, m] = ak,j - alfa * mk,j
  do j = jmax + 1, k + m1 aw[j - k + m, m] = 0.0
  kq = (k - 1) ÷ m1; ist = k - kq * m1
  do iw = ist, m1
    if abs(aw[1, iw]) < abs(aw[1, m]) then
      ip = -ip
      do jw = 1, mm aw[jw, iw] ⇌ aw[jw, m]
    if aw[1, iw] ≠ 0.0 then
      piv[(kq - 1) * m1 + iw] = aw[1, iw] ; t = -aw[1, m] / aw[1, iw]
      do jw = 2, mm
        aw[jw - 1, m] = aw[jw, m] + t * aw[jw, iw]
      aw[mm, m] = 0.0
    else
      piv[(kq - 1) * m1 + iw] = eps ; ir = ir + 1
      do jw = 2, mm
        aw[jw - 1, m] = aw[jw, m]
      aw[mm, m] = 0.0
  do iw = 1, ist - 1
    if abs(aw[1, iw]) < abs(aw[1, m]) then
      ip = -ip
      do jw = 1, mm aw[jw, iw] ⇌ aw[jw, m]
    if aw[1, iw] ≠ 0.0 then
      piv[kq * m1 + iw] = aw[1, iw] ; t = -aw[1, m] / aw[1, iw]
      do jw = 2, mm
        aw[jw - 1, m] = aw[jw, m] + t * aw[jw, iw]
      aw[mm, m] = 0.0
    else
      piv[kq * m1 + iw] = eps ; ir = ir + 1
      do jw = 2, mm
        aw[jw - 1, m] = aw[jw, m]
      aw[mm, m] = 0.0
  piv[k] = aw[1, m]
  do jw = 1, mm
    aw[jw, ist] = aw[jw, m]
  piv[k] = aw[1, k] ; P[k] = Πi=1k sign(piv[i])
  if P[k - 1] * P[k] < 0 then count = count + 1

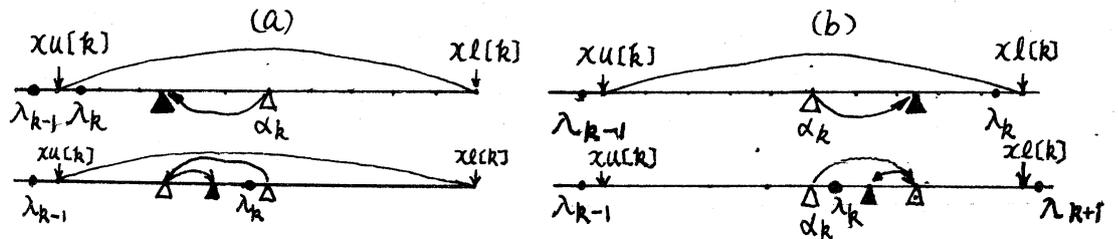
```

2.2 INVRAY

前節の BISECT によって $xu[k]$ と $xl[k]$ の間にはさまれた λ_k を、逆反復法によって求めるわけであるが、原理的には、

$$\alpha_k = (xu[k] + xl[k]) / 2$$

をシフト真にえらんでの逆反復法を行なえば、必ず収束する筈である。ところが、たとえば下図のような状況のとき、収束ははなはだしく遅い。(特にページスワップの負擔が大きい。)そこで、



(1) 逆反復に入る前に、あと二回余計に二分法を行うことによってシフト真を定める。(上例の場合 \blacktriangle に移す。)

ところがこうしても近接根のときは収束がおそいことは相変わらずであるから、

(2) 逆反復をある程度やって、まだ収束しないときは、レーリ商シフトの逆反復にひきつぐ。

じつは(1)において、一回でなく二回スツルムを行なわねばならぬのは、この(2)のことを安全に行なわせたいがためである。もし、シフト真の両側に殆んど等しい距離のところに λ_k と λ_{k+1} がある、ということだと、レーリ商シフトをやってもうまく λ_k に近づいて呉れないことがある。

INVRAY { 固有ベクトルも同時に求めるもの }

$Ax = \lambda x$ 型用

$lmin, lmax, llmax, eps1, epse$ を与える。

v_k ($k=1$ から p) に初期ベクトルを与える。

$k = 1$

while $k \leq p$ do

if $xu[k] \neq xu[k+1]$ or $k = p$ then {1}

$alfa = (xu[k] + xl[k]) / 2$; Call Sturm

 if count = k then {2}

$xl[k] = alfa$; $alfa = (xu[k] + alfa) / 2$; Call Sturm

 if count $\neq k$ then {3}

$xu[k] = alfa$; $alfa = (alfa + xl[k]) / 2$

 else $xl[k] = alfa$ {3'}

 else {2'}

$xu[k] = alfa$; $alfa = (alfa + xl[k]) / 2$; Call Sturm

 if count = k then {4}

$xl[k] = alfa$; $alfa = (xu[k] + alfa) / 2$

 else $xu[k] = alfa$ {4'}

else

$alfa = xu[k]$; $lmin = 1$

100 continue

$epsir = 10^{-10}$; Call GLUBR3 ($ar, ac, ip, n, m1, alfa, epsir, ir$)

$w = v_k$

 do $j = \max(k-8, 1), k-1$

$w = w - (v_j^T v_k) v_j$ {5}

$w = w / (w^T w)^{1/2}$; $alfavo = 0.0$; $iconv = 0$

 do $l = 1, lmax$ {6}

$b = w$; Call GSBK ($ar, ac, b, ip, n, m1, \dots$)

$alfav = 1.0 / w^T b$; $eig = alfa + alfav$

 if $|alfav - alfavo| < eps1$ then $iconv = 1$

$alfavo = alfav$; $w = b$

 do $j = \max(k-8, 1), k-1$

$w = w - (v_j^T b) v_j$ {7}

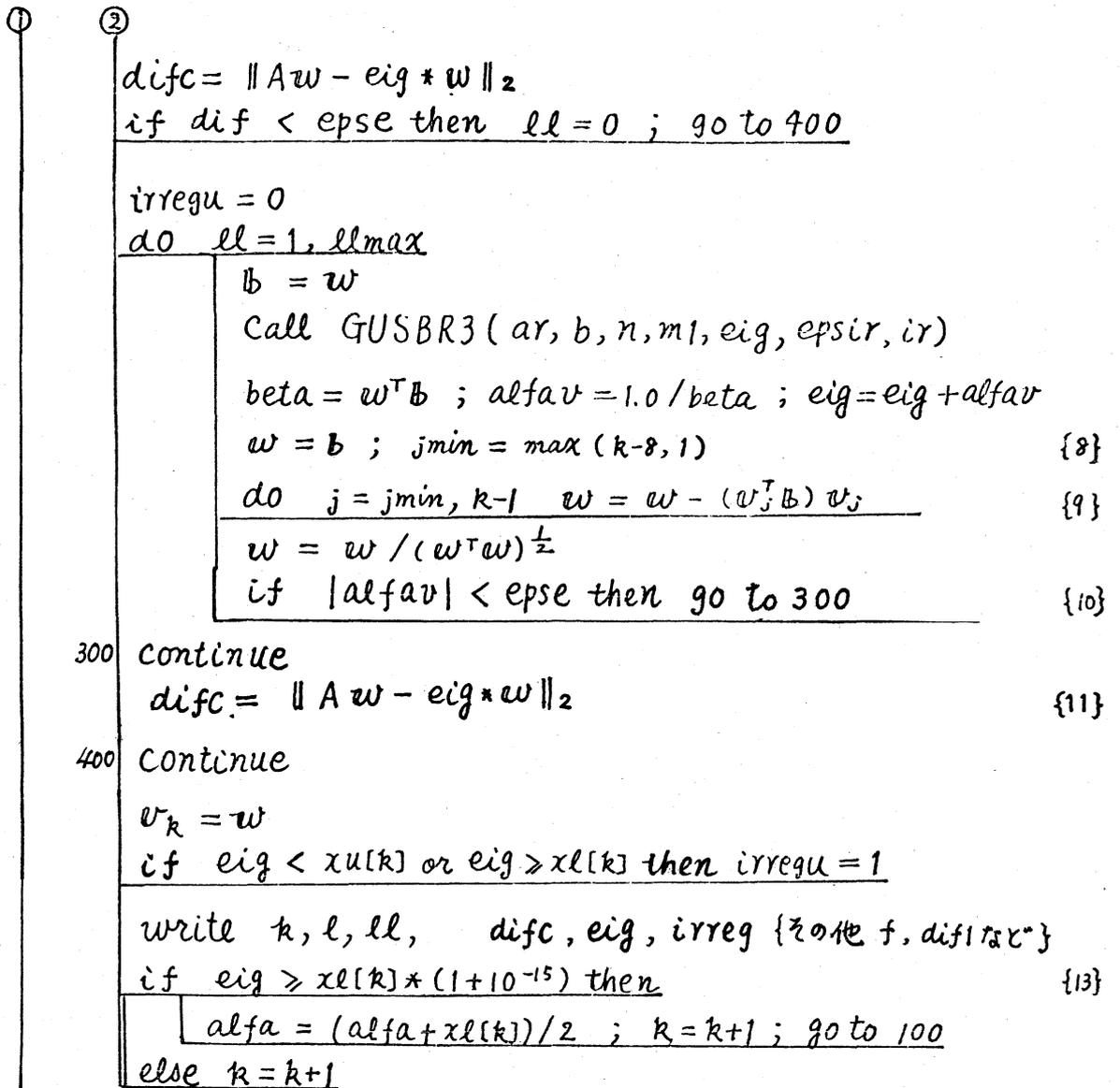
$w = w / (w^T w)^{1/2}$

 if $l > lmin$ and $iconv = 1$ then go to 200

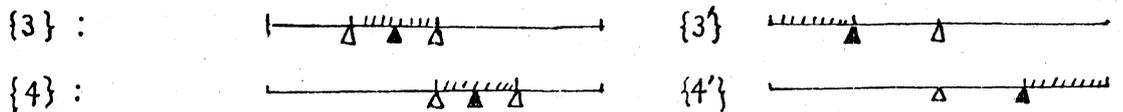
200 continue

①

②



[解説]



{5} : 最近求まった 8 個の固有ベクトルは、 w からぬきとられる。

{7} と {9} : 両直交化である。直前から 8 個前までの固有ベクトルの成分をぬきとった。

次頁に GUSBR3 を示す。これを LU 分解部と右辺に対する前進・後退部とに分けたものが GLUBR3 と GSB である。

GUSBR3 (ar, b, n, m1, eig, epsir, ir)

```

m = m1 + 1 ; ir = 0
do i = 1, m1
  do j = 1, i + m1
    ar[j, i] = a[i, j] - eig * delta[i, j]
  do j = i + m1 + 1, m1 + m1 + 1
    ar[j, i] = 0.0

do k = 1, n
  imax = min(k + m1, n) ; jmax = min(k + m1 + m1, n)
  if k + m1 <= n then
    do j = k, jmax
      ar[1 + j - k, k + m1] = a[k + m1, j] - eig * delta[k + m1, j]
    amax = abs(ar[1, k]) ; ipk = k
    do i = k + 1, imax
      aik = abs(ar[1, i])
      if aik > amax then
        amax = aik ; ipk = i
    jkmax = min(1 + m1 + m1, n - k + 1)
    if k < ipk then
      b[k] <=> b[ipk]
      do jk = 1, jkmax
        ar[jk, k] <=> ar[jk, ipk]
    if amax < epsir then
      ar[1, k] = epsir * sign(ar[1, k]) ; ir = ir + 1
    do i = k + 1, imax
      t = -ar[1, i] / ar[1, k]
      b[i] = b[i] + t * b[k]
      do jk = 2, jkmax
        ar[jk - 1, i] = ar[jk, i] + t * ar[jk, k]
      ar[jkmax, i] = 0.0

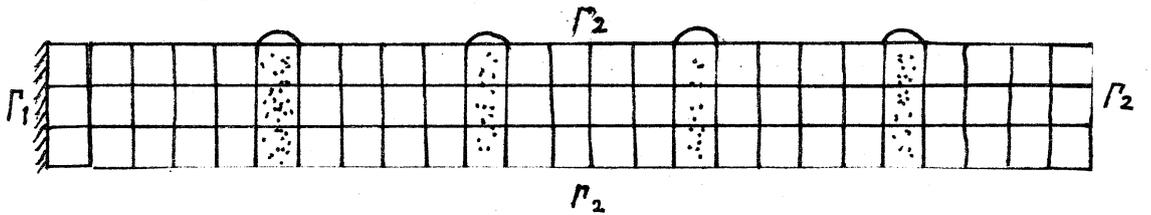
do k = n downto 1
  jmax = min(k + m1 + m1, n)
  sum = 0.0
  do j = k + 1, jmax
    sum = sum + ar[1 + j - k, k] * b[j]
  b[k] = (b[k] - sum) / ar[1, k]

```

2.3 結果に関する中間報告

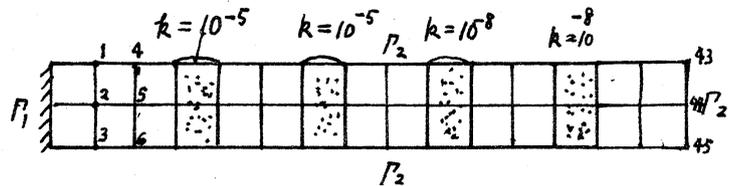
固有値計算プログラムのデバッグと、評価を行うに際して決定的に重要なのは、テスト問題の選択である。いろいろのレベルの近接度をもった固有値系を、システマティックに作るために、有限要素法を利用した。すなわち、下図のような場を作り、部のk値を任意に変えれば、思いのままの近接度をもった行列が生成される。(□部のk=1.0にとる.)

支配方程式 : $\nabla \cdot (-k \nabla u) = \lambda u \rightarrow A u = \lambda M u$ (但し $M=I$ とした.)



$\Gamma_1 : u=0$ (固定) , $\Gamma_2 : \nabla u=0$ (自由)

[簡単な問題例]



節番号を上のようにつけたから、これは、 $m=4$ ($m_1=3$)の帯行列となる、 $n=45$ 。熱伝導の問題だと思えば、のところが、熱伝導率の非常に小さい物質となっている。

次頁上段に行列 A を示す。左端の列にならんでいるのが対角項である。対称だから上半分だけ示している。次頁下段は、 $\text{epsbi} = 10^{-4}$ によって BISECT を行なって、 $\text{epse} = 10^{-10}$ 、 $\text{epse} = 10^{-8}$ で LL 反復を行なったときのもの、 $\text{epsi} = 10^{-4}$ とした。

STRUM KAISU= 88

K	L	LL	F	DIFC	EIG *
1	1	1	2	0.235D-12	0.00000000847585
2	1	2	2	0.245D-12	0.000000005814623
3	1	2	2	0.674D-14	0.000000850397529
4	2	0		0.525D-11	0.000005818384915
5	3	2		0.517D-11	0.130988706559026
6	1	3	2	0.655D-12	0.633974604384458
7	1	3	2	0.435D-12	0.633974599342072
8	1	5	2	0.337D-12	0.633974597550948
9	1	4	2	0.176D-12	0.633975935359265
10	1	2	2	0.465D-12	0.633974604918049
11	2	1		0.235D-12	0.633977726206820
12	2	0		0.307D-10	0.633982769889014
13	2	0		0.307D-10	0.633983300138170
14	4	2		0.581D-15	0.759410727188718
15	3	2		0.427D-14	0.981256155790828
16	1	4	2	0.752D-13	1.000000003964257
17	1	1	2	0.792D-13	1.000000011034313
18	2	2		0.115D-13	1.000003968777679
19	2	0		0.525D-11	1.000011036336149
20	3	2		0.442D-12	1.65555485733648
21	1	4	2	0.655D-12	1.697224367542230
22	1	3	2	0.133D-14	1.697224374668973
23	2	1		0.125D-11	1.697229641804310
24	2	1		0.135D-11	1.697236764061253
25	3	2		0.117D-11	1.933646468546035

LL	F	DIFC	EIG
0	2	0.108D-08	0.000000000854363
1	2	0.179D-09	0.000000005814563
2	2	0.845D-09	0.000000850397529
0		0.200D-09	0.000005818384915
2		0.588D-09	0.130988706559026
2	2	0.224D-09	0.633974604378081
2	2	0.526D-09	0.633974599231349
7	2	0.356D-08	0.633974602328289
4	2	0.642D-09	0.633975935359265
2	2	0.354D-08	0.633974599670639
1		0.142D-10	0.633977726206820
0		0.314D-10	0.633982769889014
0		0.307D-10	0.633983300138170
2		0.580D-15	0.759410727188718
2		0.427D-14	0.981256155790828
1	2	0.324D-08	1.000000006083633
1	2	0.492D-08	1.000000005209902
1		0.494D-11	1.000003968777679
0		0.681D-11	1.000011036336149
2		0.442D-12	1.65555485733648
2	2	0.123D-08	1.697224367760609
2	2	0.125D-08	1.697224374668917
0		0.374D-08	1.697229641806272
0		0.440D-08	1.697236764060845
2		0.312D-09	1.933646468546035

$\text{epsbi} = 10^{-6}$

$\text{epse} = 10^{-10}$

$\text{epse} = 10^{-8}$

以上は、 $\text{epse} = 10^{-10}$ と 10^{-8} のときのものであつた。次に、 $\text{epse} = 10^{-12}$ としたものを示そう。これで見ると、DIFC, すなわち $\|Aw - \text{eig} \cdot w\|_2$ が ^{殆んど} すべて 10^{-15} 以下に収まっている。

K	L	LL	F	DIFC	EIG *	DIF1
1	1	2	2	0.310D-15	0.00000000847585	0.235D-12
2	1	2	2	0.445D-15	0.000000005814623	0.202D-10
3	1	2	2	0.335D-15	0.000000850397528	0.205D-09
4	2	1		0.342D-15	0.000005818384915	0.525D-11
5	3	2		0.233D-15	0.130988706559026	0.162D-10
6	1	4	2	0.204D-15	0.633974604384458	0.495D-12
7	1	4	2	0.170D-15	0.633974599342072	0.916D-12
8	1	6	2	0.235D-15	0.633974597550948	0.129D-12
9	1	4	2	0.204D-15	0.633975935359265	0.170D-10
10	1	2	2	0.325D-15	0.633974604918049	0.172D-08
11	2	2		0.350D-15	0.633977726206820	0.193D-12
12	2	1		0.248D-15	0.633982769889014	0.307D-10
13	2	0		0.728D-14	0.633983300138170	0.723D-14
14	4	2		0.573D-15	0.759410727188718	0.385D-09
15	3	2		0.425D-14	0.981256155790828	0.217D-08
16	1	5	2	0.316D-15	1.000000003964257	0.752D-13
17	1	2	2	0.615D-15	1.000000011034313	0.243D-13
18	2	2		0.303D-15	1.000003968777679	0.123D-11
19	2	1		0.634D-15	1.000011036336149	0.525D-11
20	3	2		0.236D-15	1.65555485733648	0.103D-08
21	1	4	2	0.162D-14	1.697224367542230	0.491D-10
22	1	3	2	0.133D-14	1.697224374668973	0.512D-11
23	2	2		0.255D-15	1.697229641804310	0.106D-14
24	2	1		0.425D-15	1.697236764061253	0.232D-08
25	3	2		0.253D-15	1.933646468546035	0.117D-08

【参考】

右端の DIF1 は、LL の最後のひとつ前のときにおける DIFC の値を示している。

DIF1 と DIFC を比べて見ると、最後の反復でどれぐらい改善されたかが判る。

$\text{epsbi} = 10^{-6}$

$\text{epse} = 10^{-12}$

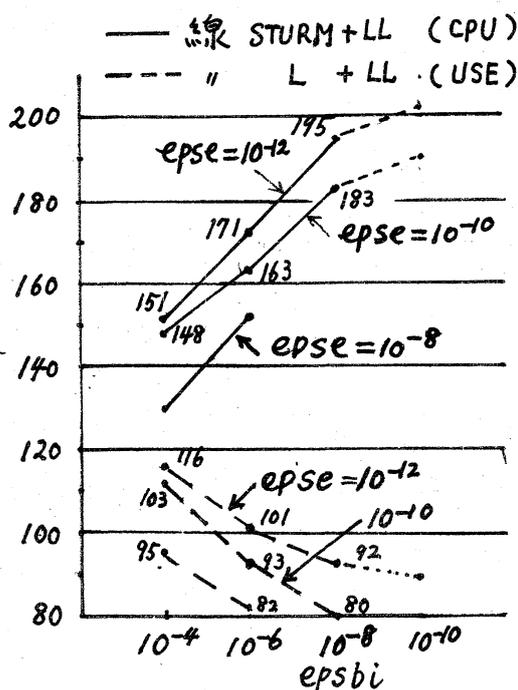
* EIG は、 $\text{epse} = 10^{-10}$ で"のもの"すべてに殆んどすべて 10^{-15} の位まで収まっている。

epsbiとepseを動かすとき、CPUタイムとUSEタイムが傾向としてどのように変化するかをしらべてみる。CPUタイムにとっては、STURM回数とLU分解数が利く。USEタイムにとっては、LU分解数のほかにL反復の回数か利く。下表にそのとりまとめた結果を示す。(L反復の前のLU分解はいつでも、必ずやるから表にはあげない。)

epsbi	epse	(BISEC) STURM	(INVRA) STURM	L	LL	STURM+LL	L+LL
10 ⁻⁴	10 ⁻¹²	60	12	37	79	151	116
	10 ⁻¹⁰				76	148	113
	10 ⁻⁸					130	95
10 ⁻⁶	10 ⁻¹²	88	26	44	57	171	101
	10 ⁻¹⁰				49	163	93
	10 ⁻⁸				38	152	82
10 ⁻⁸	10 ⁻¹²	115	38	50	42	195	92
	10 ⁻¹⁰				30	183	80
	10 ⁻⁸						

上の表では印象が不鮮明ゆえ、ぐらぶにしてみたのが右のものである。大体、思惑通りの結果ではあるが、epseを甘くしても、そう大して計算時間がみぢかくならぬのは意外に近い印象を受けた。

改善の余地はある。



〔ひとつの改善案〕

初め粗い eps_{bi} (たとえば 10^{-4}) で BISECT を行なうと、その段階で各固有値が密集したあるグループに属しているか、それとも孤立しているかが判る。(たとえば先程の例で $\text{eps}_{bi} = 10^{-4}$ を使用すれば、 $\lambda_1 \sim \lambda_4$ が第1密集グループ、 λ_5 が孤立根、 $\lambda_6 \sim \lambda_{13}$ が第2密集グループ... という具合である。)

そこで、孤立根に対してはほぼ先程の通りの処理を行なうが、密集グループに属するものに対しては、各グループ毎にシフト英をえらんで、シュミット直交化つきの右典的な同時逆反復法をほどこす。

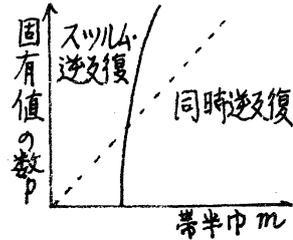
この同時逆反復法で、各固有値への収束は比較的速い。固有ベクトルの方は、各固有値に対応させて計算機の限度に近いほどの精度で求めようとするともやはり収束に手間どる。従って、適当なところで各固有値毎にシフト英を移しての逆反復にひきつぐがよい。

しかし、固有ベクトルについては、それらが張るところの部分空間が、問題の密集固有値に対応する固有空間に収束すればそれで十分だという考えならば、その必要はない。

同時逆反復法は、ベクトルの収容法をうまく考え、プログラムに若干の工夫をこらすことによって、ページスワップを大中に減らすことができる。

3 同時逆反復法

CPUタイムについては、スツルム・逆反復法と、同時逆反復法の適用範囲の傾向は、およそ右図のようになる。ページスワップについてはどうであろうか、ここでは Bathe の Subspace法をとり上げる。常識的なアルゴリズムはおよそ下記の如くなるよう:



初期 $X = \{x_1, \dots, x_q\}$ を用意, $q = \min(p+b, 2p)$

A を LDU 分解 ; $k=0$

while $k \leq k_{max}$ and $\|Ax_p - \tilde{x}_p M x_p\| / \|\tilde{x}_p M x_p\| > \epsilon_{ps}$ do

do $i=1, q$ { $Y = MX$ を作る }

$x_i = x_i / \sqrt{x_i^T x_i}$; $y_i = M x_i$

do $i=1, q$ { $A\bar{X} = Y$ を解く } (*)

$z = L^{-1} y_i$; $\bar{x}_i = (DL^T)^T z$

do $i=1, q$ { $q \times q$ 行列 $\tilde{A} = \bar{X}^T A \bar{X}$, $\tilde{M} = \bar{X}^T M \bar{X}$ を作る }

$z = M \bar{x}_i$; $\tilde{m}_i = \bar{X}^T z$; $\tilde{x}_i = \bar{X}^T y_i$

solve eigenprob: $\tilde{A} \tilde{q} = \tilde{M} \tilde{q} \tilde{\lambda}$

do $i=1, q$ { m の空間のベクトルにひきのぼす }

$x_i = \bar{X} \tilde{q}_i$

$k = k+1$

研究集会当時にはこれでもって評価し、スツルム逆反復法とくらべた。そして、'ページスワップに関しては $q > m$ になると全くサブスペースは不利である。' と結んだ。実は、上の通りのアルゴリズムだとそうなるが、サブスペースにおいては、右辺のベクトル Y をブロック化して収め、ブロック毎に前進・後退代入をやればページスワップを大中に減らす

ことができることが判った。SAP IV の Subspace プログラムを調べてみると、これは仮想メモリ方式向けでないが、A と Y と共によこ方向にブロックわけして処理することによって、ディスクと主メモリの転送回数を減らす工夫がすでにしてあった。我々の場合、仮想メモリ方式の利点を利用できるから、A の方は普通の通りとして、(すなわち DL は普通の通りとして、Y (したがって X も) だけをよこ方向にブロック化して收容してのアルゴリズムを採用すればよい。

— 以上、前言を訂正し、おわびをしたい。 —

さて、問題のサブスペースに関し、極く最近原著者 Bathe 自身による注目すべき改良が発表された⁽³⁾。

4. 対称帯 $Ax = \lambda x$ に対する直接 QR 法

問題をこの型に限るならば、A を特殊なハウスホルダ変換の列によつて帯巾をもとの二倍におさえて三重対角化する方法が、筆者らによつて考案、開発された。⁽⁴⁾ この方法は汎用的ではあるが、帯巾が大きくなると、やはりページスワップネックとなる。今までやった問題の上限は、帯半巾 80, 元数 8000 である。それにしても、所望の固有値の数 p が大きくなるとこの方法によらざるを得ないが、 p が比較的小さい問題には、帯 QR 法の方が有望となる。そういう観点から、帯 QR 法をとり上げることにした。

[原理の復習] $A = A_1$ とする. 単位直交変換 Q_l^T により, A_l を三角化: $Q_l^T A_l = R_l$ 即ち $A_l = Q_l R_l$, $Q_l^T Q_l = I$ する. ここに $Q_l^T = (I - \alpha_l^{(n-1)} w_l^{(n-1)} w_l^{(n-1)T}) \cdots (I - \alpha_l^{(1)} w_l^{(1)} w_l^{(1)T})$.

$Q_l R_l$ を逆順に掛けて $A_{l+1} := R_l Q_l (= Q_l^T A_l Q_l)$ とすると $l \rightarrow \infty$ のとき A_{l+1} は $\lambda_n \geq \cdots \geq \lambda_1 > 0$ を対角要素とする対角行列に収束する. 内容的には, $P_l := Q_1 \cdots Q_l$ の各列ベクトルは, 単位行列 I の各列ベクトルを初期ベクトルとしたシムリット直交化つき同時逆反復法によって得られたベクトル \bar{x}_l に等しい. すなわち 同時逆反復法 の一種と見なされる. (2)

我々のアルゴリズムは, 帯QR法がシムリット直交化つきの同時逆反復法だという見方に強く影さようされているので, 以下 (本質的には (2) に従って) その証明の大筋を見ておく.

$$A_{l+1} := P_l^T A P_l, \text{ 但し } P_l := Q_1 \cdots Q_l, S_l := R_l \cdots R_1 \quad \cdots (1)$$

$$\text{と置く. } P_l S_l = Q_1 \cdots Q_{l-1} Q_l R_l R_{l-1} \cdots R_1 = P_{l-1} A_l S_{l-1} \quad \cdots (2)$$

(1) より, $A_l = P_{l-1}^T A P_{l-1}$, 従って $P_{l-1} A_l = A P_{l-1}$, これを (2) に入れて,

$$\underline{P_l S_l} = A P_{l-1} S_{l-1} = A^2 P_{l-2} S_{l-2} = \cdots = \underline{A^l} \quad \cdots (3)$$

同時逆反復法の方を n 列ベクトルの方から順に直交化して

$$A X_l = \bar{X}_{l-1}, \bar{X}_l = L_l X_l \text{ 即ち } A \bar{X}_l L_l^T = \bar{X}_{l-1} \quad \cdots (4)$$

と書くと, $\bar{X}_l = A^{-1} \bar{X}_{l-1} L_l = A^{-2} \bar{X}_{l-2} L_{l-1} L_l = \cdots = A^{-l} \bar{X}_0 \underline{L_1 \cdots L_l}$

より, $\bar{X}_l = A^{-l} \tilde{L}_l$, $A^l = \tilde{L}_l \bar{X}_l^T$. これを転置して, \tilde{L}_l とおく.

$$A^l = \bar{X}_l \tilde{L}_l^T \quad \cdots (5)$$

(3), (5) より $A^{2l} = S_l^T S_l$, $A^{2l} = \tilde{L}_l \tilde{L}_l^T$ を得るゆえコレスキー分解の一意性によって, $S_l = \tilde{L}_l^T$ が従がう. 一方, (3), (5) から $P_l = A^l S_l^{-1}$, $\bar{X}_l = A^l (\tilde{L}_l^T)^{-1}$ ゆえ, $P_l = \bar{X}_l$ が従がう. (終)

シュミット直交化つき同時逆反復法だということから、重複あるいは近接固有値があってもうまく行く筈である。また原理通りやったのでは近接固有値があるとき収束ははなはだ遅い。そこで、原素シフトを必ずやることに昔からなっている。

[プログラム作成上の要訣をおさえるための考察]

$$\bar{x}_1^{(0)} = c_1 u_1 + c_2 u_2 + c_3 u_3 + \dots + c_n u_n, \quad c_1 \neq 0 \quad (\text{初期ベクトル})$$

と書く。(QR法ではこれは implicit である。同時逆反復とみての話)

$\lambda_1 < \lambda_2$ のときは、逆反復法の原理からして、

$$\bar{x}_1^{(l)} = \left(\frac{1}{\lambda_1}\right)^l \left\{ c_1 u_1 + \left(\frac{\lambda_1}{\lambda_2}\right)^l c_2 u_2 + \left(\frac{\lambda_1}{\lambda_3}\right)^l c_3 u_3 + \dots \right\} / k \quad \dots \dots$$

$$= \alpha_1 u_1 + \varepsilon_{12} u_2 + \dots \quad \text{と書く。}$$

$\lambda_1 = \lambda_2 < \lambda_3$ なら、{ k は正規化のための因子 }

$$\bar{x}_1^{(l)} = \left(\frac{1}{\lambda_1}\right)^l \left\{ c_1 u_1 + c_2 u_2 + \left(\frac{\lambda_1}{\lambda_3}\right)^l c_3 u_3 + \dots \right\} / k$$

$$\underline{A}_{l+1} = P_l^T A P_l = \underline{X}_l^T A \underline{X}_l, \quad \underline{X}_l = [\bar{x}_n^{(l)}, \dots, \bar{x}_1^{(l)}], \quad \bar{x}_1^{(0)} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad (\text{例えは})$$

となつているから、

A_{l+1} の右下隅に近いところの要素：(3,3) . . .
. (2,2) (2,1)
をしらべてみる。それらは $\bar{x}_i^T A \bar{x}_j$ で、. (1,2) (1,1)

$$\bar{x}_i^T A \bar{x}_i = (\alpha_i)^2 \lambda_i + \sum_{k \neq i} (\varepsilon_{ik})^2 \lambda_k \quad (\text{対角項})$$

$$\bar{x}_i^T A \bar{x}_j = \alpha_i \varepsilon_{ji} \lambda_i + \varepsilon_{ij} \alpha_j \lambda_j + \sum_{k \neq i, j} \varepsilon_{ik} \varepsilon_{jk} \lambda_k \quad (\text{非対角項})$$

いま、 $\lambda_1 < \lambda_2$ だが $\lambda_1 \doteq \lambda_2$ (近接) だとしてみると、

右下隅の (1,1) 要素 $\bar{x}_1^T A \bar{x}_1 = (\alpha_1)^2 \lambda_1 + (\varepsilon_{12})^2 \lambda_2 + \dots$ の収束は

順調であるが、(1,2) = (2,1) 要素 $= \bar{x}_1^T A \bar{x}_2 = \alpha_1 \varepsilon_{21} \lambda_1 + \varepsilon_{12} \alpha_2 \lambda_2 + \dots$

は、 ε_{12} ($\bar{x}_1^{(0)}$ の u_2 成分) のため収束がおそい。

プログラム作成上の要奥としては、重複、あるいは極度に近接した固有値が密集しているとき、如何にうまく処理するか、ということがその第一である。

〔アイデアの解説〕

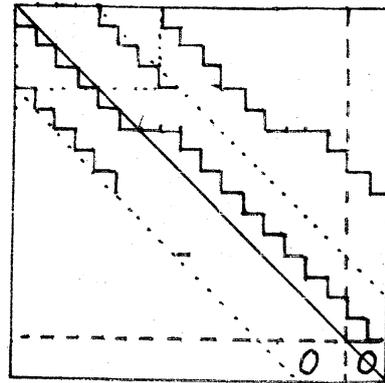
$0 < \lambda_1 \leq \lambda_2 \leq \dots$ とし、 $\bar{A} = A - \lambda_1 I$ とおく。行列 \bar{A} の階数不足値 ν を tol の粗さ(精しさ)に応じて決定できる特別のハウスホルダ三角分解アルゴリズム⁽⁵⁾ を応用した新しい $Q^T A Q$ 変換のプログラム: $QTAQBR$ を使用する。

λ_1 が s 重の固有値なら、 \bar{A} は s 重のゼロ固有値を持ち、 \bar{A} の階数不足値 ν がちょうど s となる。($s=2$ の場合) →

例えば $s=2$ の場合、三角分解を

$$Q^T \bar{A} = R, \quad \bar{A}_2 = Q^T \bar{A} Q$$

を作ると、 \bar{A}_2 は元の帯のワク内に収



まり、かつ下二行はゼロとなる。(今は対称ゆえ右二列もゼロ)

これは2重のゼロ固有値をもっていることは明らかである。

このときは $\bar{A}_2 = Q^T \bar{A} Q$ の下二行、右二列を一挙に落して先に進めばよい。従って、重複固有値があれば、それだけ仕事はかどる、という好ましい結果になる。 λ_1, λ_2 を重複とみるかどうかは、 tol をどう与えるかによるゆえ、これを特に粗くとれば、仕事を早く済ますことができる。

EIGQR { tol (精), tolc (粗) を与える. (例 $tol = 10^{-10}$, $tolc = 10^{-4}$) }
 alfa : 所望の固有値 $\lambda_1 \dots \lambda_p$ の上界とする. (あるいは $anorm = \|A\|$)

$k = 1$; $nv = n$; $eig[0] = 0.0$; $eigv = 0.0$

$eps = 10^{-16} * \underline{alfa}$

行列 $A(n, n)$ に対し QTAQBR を行う

if $iy > 0$ then {ゼロ固有値を始末する}

do $i = 1, iy$

$eig[i] = eig[0]$

$nv = n - iy$; $k = k + iy$

while $k \leq p$ do

repeat {シフト無固定のQTAQ粗反復}

$diag = a[nv, nv]$; $eps = 10^{-16} * alfa$

行列 $A(nv, nv)$ に対し QTAQBR を行う. {*

until $abs(a[nv, nv] - diag) < tolc * alfa$

$eigv = a[nv, nv]$; $eig[k] = eig[k-1] + eigv$

$eps = \max(10^{-16} * alfa, tol * abs(eigv))$

repeat {毎回シフトを行ないながらのQTAQ精反復}

do $i = 1, nv$ {シフト}

$a[i, i] = a[i, i] - eigv$

行列 $A(nv, nv)$ に対し QTAQBR を行う

$eigv = a[nv, nv]$; $eig[k] = eig[k] + eigv$

until $iy > 0$ {昇根なら $iy = 1$ }

do $i = 1, iy - 1$ { $iy = 1$ ならやめ}

$eig[k+i] = eig[k]$

$nv = nv - iy$; $k = k + iy$

{*}のQTAQBRは、これから求めようとしている λ_k (およびそれと近接して λ_{k+1} がもしあれば λ_{k+1} も) に対応する固有ベクトル成分を、他とくらべて卓越させるためのものである。

QTAQBR (a[n,nv], eps, ir, irow, l, m1

ir = 0 ; m = m1 + 1 ; irmax = l - m ; irow = 0

do kk = 1, nv

 k = kk - ir ; imax = min0(kk + m1, nv)

 sum = $\sum_{i=k}^{imax} a[i, kk]**2$

 sk = sqrt(sum) ; jmax = min0(kk + 2*m1, nv)

 if sk > eps then

 hk = sk * (sk + abs(a[k, kk])) ; alf[k] = 1.0 / hk

 if a[k, kk] < 0.0 then sk = -sk

 wk[1, k] = a[k, kk] + sk ; a[k, kk] = -sk

 do i = k + 1, imax

 wk[i - k + 1, k] = a[i, kk] ; a[i, kk] = 0.0

 p^T = $\alpha w^T A$ を行う

 A = A - w p^T を行う

 else

 ir = ir + 1

 do i = k, imax a[i, kk] = 0.0

 if (ir > irmax) \wedge (k + l - 1 \leq n) then

 irow = 1 ; return

 irk[k] = ir

 ⊙ if kr + m1 + irk[kr] < kk then

 R(I - $\alpha w w^T$) を行う { P = $\alpha R w$, A = R - p w^T }

 kr = kr + 1

○ while kr \leq nv - ir do

 R(I - $\alpha w w^T$) を行う { P = $\alpha R w$, A = R - p w^T }

 kr = kr + 1

return

RQ 変換を⊙と○に分けることによってページスワップ°

13頁に示した例題を従来の帯QR法でやると、固有値の求まる順が、はなはだしく乱れる。(左端の場合からの単根 $0.1309\dots$, $0.7594\dots$, のたぐいがづい分あとになってやっと求まる。) 新しいEIGQRで相当改善はされるが、不十分である。こういう問題に対しては、

- (1) 粗い Tol (例えば 10^{-4}) を使って EIGQR を走らせる。その結果 Tol 値以下に近接したものは '重複' と判定される。重複度を含め所望の箇数を可なり上まゆって求める。
- (2) スツルムチェフクにより所望の固有値は上で求めたものに全部含まれているかどうか調べる。
- (3) '重複' 扱いを受けた固有値グループ毎にシフト臭をえらび、今度は精しい Tol (例えば 10^{-12}) によって EIGQR を走らせる。

[文献]

- (1) Martin·Wilkinson : 'Solution of Symmetric and Unsymmetric Band Equations and the Calculations of Eigenvectors' Numer.Math.9, 279-301(1967)
- (2) Bathe/Wilson, 菊地記 : 有限要素法の数値計算 530頁~531頁
- (3) Bathe/Ramaswamy : 'An accelerated subspace iteration method' Comp.Meths. Appl. Mech. Eng. 23 (1980) 313-331
- (4) 村田/堀越 : '対称帯行列を三重対角化するための新アルゴリズム' 情報処理 Vol16, No2, (1975年2月)
- (5) 村田/二村/門間 '仮想記憶方式の下での大形行列計算技法' 情報処理 Vol21, No2, (1980年4月)