

UDC 512.83:  
681.3.06

# 航空宇宙技術研究所資料

TECHNICAL MEMORANDUM OF NATIONAL AEROSPACE LABORATORY

TM-277

## 1 次方程式系の解法 I

—— 係数行列が正定値である場合 ——

福田正大・末松俊二

1975年5月

航空宇宙技術研究所  
NATIONAL AEROSPACE LABORATORY

## 目 次

1. はじめに .....	1
2. 条件数について .....	2
3. 数値実験 .....	3
4. コレスキー法 .....	4
方法の基礎 .....	4
サブルーチンの説明 .....	5
数値実験及び結果 .....	14
5. 逐次法による解の精密化 .....	14
方法の基礎 .....	14
サブルーチンの説明 .....	17
数値実験及び結果 .....	20
6. 帯状行列 .....	20
方法の基礎 .....	20
サブルーチンの説明 .....	20
数値実験及び結果 .....	23
7. ガウス・ジョルダン法 .....	25
方法の基礎 .....	25
サブルーチンの説明 .....	25
数値実験及び結果 .....	27
8. バイマトリックスに関する若干の考察 .....	27
9. 結 語 .....	30

# 1 次方程式系の解法 I \*

— 係数行列が正定値である場合 —

榎 田 正 大\*\*・ 末 松 俊 二\*\*

## 概 要

当所に大型計算機が導入されてから十数年になるが、科学技術計算用の標準的なサブルーチンというものがない。そこで、これらのものの作成にとりかかるとことにした。まず第一に、線型計算関係を取り扱うことにし、今回はその中でも、係数行列が正定値である場合の、方程式系の解法及び逆行列の計算に関するサブルーチンについて報告する。解法としては、本質的には2つの方法—コレスキー分解法とガウス・ジョルダン法—を採用した。

## 1 章 はじめに

計算センターの主要な任務の一つとして、科学技術計算用サブルーチンライブラリーの標準化ならびに充実化というものが考えられる。当所に計算機が導入されてから十数年になるにもかかわらず、過去この様な研究が系統的に行なわれてこなかったのは遺憾とするところである。

以上のようなことを踏まえた上で、まずその第一段階として、線型計算関係のルーチンからとりかかるとことにした。その際、線型計算関係については定評のある、J. H. Wilkinson & C. Reinisch編の“HANDBOOK FOR AUTOMATIC COMPUTATION”を参考にし、これにプログラミング技術上の改変を加え、テスト計算を系統的に行ない、精度、計算速度等に関する資料をととのえ、それを当計算センターにおける線型計算の標準サブルーチンとすることにした。

今回は、その第一回分として、稠密もしくは帯状の正値対称行列を対象とする、1次方程式系或は行列の逆転に関するルーチンについて報告する。

プログラミングするにあたり、次のようなことに注意した。

1) IF文による制御はなるべく避けて、DO文による

制御とする。

- 2) 2次元アレイの場合には、行毎の計算よりは列毎の計算になるようにする。(データの連続性のため)
- 3) 行列式の値を副産物として計算しているサブルーチンでは、その値が単一変数ではオーバーフローもしくはアンダーフローを起こす可能性のあることを考えて、仮数部と指数部とに分けて表示するようにした。

このことをFORTRANですると、例えば底を $a$ とし、 $\det A = D \times a^M$   $a^{-1} < |D| < 1$  とするためには、

```
1 IF (ABS(D). LT. 1.0) GO TO 2
  D = D * a-1
  M = M + 1
  GO TO 1
2 IF (ABS(D). GE. a-1) GO TO 3
  D = D * a
  M = M - 1
  GO TO 2
3 CONTINUE
```

というようなことになる。しかしこうすると、ある場合には、計算時間のかなりの部分がここでくわれることになる。そこで、2進計算機であるので、底を2に取り、この部分をアセンブラで書くことにした。そうすると、毎回 $D$ にでてくる指数部を $M$ に足しこみ、しかる後に $D$ の指数部を0にする操作だけで済む。

このサブルーチンを、単精度のものには“DEC”、倍精度のものには“DECD”と名付けた。

- 4) サブルーチン内でワーキングエリアとしてアレイを使用する場合、2次元であれば引数に加え、1次元であれば引数に加えなかった。従って、1次元アレイのワーキングエリアを使用しているサブルーチンでは、行列の次数 $N$ に上限が生じる。

\* 昭和50年2月26日 受付

\*\* 計算センター

## 2章 条件数について

本章では、ベクトルを小文字で、行列を大文字で表記する。又、それぞれのノルムを同一の記号 $\|\cdot\|$ で表わす。なおベクトルノルムと行列ノルムの関係について、3節に若干記しておく。詳しくは航技研資料No.53<sup>(8)</sup>を参照。

### 2-1 方程式系の perturbation

方程式系  $Ax = b$  において、 $A$  や  $b$  に擾動が加わったときの、解  $x$  の変化を評価してみよう。

$b$  が  $b+k$  になり、 $x$  が  $x+h$  になったとすると、

$$A(x+h) = b+k \quad (2.1.1)$$

従って、

$$\|h\| \leq \|A^{-1}\| \cdot \|k\|$$

通常は、相対誤差が問題になることが多いから、

$$\|h\|/\|x\| \leq \|A\| \cdot \|A^{-1}\| \cdot \|k\|/\|b\| \quad (2.1.2)$$

ここで現われた、 $\|A\| \cdot \|A^{-1}\|$  という数は以下しばしば現われるので、これを  $P(A)$  と書くことにする。そうすると上記不等式は、“解の相対誤差は、右辺の相対誤差の  $P(A)$  倍を越えない”ことを意味している。

次に、 $A$  が  $A+F$  に変化して、 $x$  が  $x+h$  になったときを考える。そうすると、

$$(A+F)(x+h) = b \quad (2.1.3)$$

従って、

$$(A+F)h = -Fx$$

ところで、 $A$  が正則であっても、 $F$  になんらかの制限をつけない限り、一般には、 $A+F$  は非正則になり得る。そこで

$$A+F = A(E+A^{-1}F) \quad E: \text{単位行列}$$

だから、 $\|A^{-1}F\| < 1$  を仮定すれば、 $A+F$  は正則になる。さらに、後の計算上

$$\|A^{-1}F\| \leq \|A^{-1}\| \cdot \|F\| < 1 \quad (2.1.4)$$

を仮定する。そうすれば、

$$h = -(E+A^{-1}F)^{-1} A^{-1} Fx$$

より

$$\|h\| \leq \|A^{-1}F\| \cdot \|x\| / (1 - \|A^{-1}F\|)$$

が得られ、結局相対誤差は

$$\|h\|/\|x\| \leq (P(A) \cdot \|F\|/\|A\|) / (1 - P(A) \cdot \|F\|/\|A\|) \quad (2.1.5)$$

となる。

或は又、計算機等で実際に計算するときには、有限桁数しか表示し得ないことよって、 $A$  や  $b$  に誤差が生じる。今、仮数部が  $t$  桁の浮動小数点表示を採用したとすると、

$$Ax = b \quad (2.1.6)$$

は、

$$Ax' = b' \quad (2.1.7)$$

$$\text{ここで、 } A = A + F \quad |f_{ij}| < 2^{-t} \cdot |a_{ij}|$$

$$b' = b + k \quad |k_i| < 2^{-t} \cdot |b_i|$$

となる。(  $F, k$  は  $A, b$  を 2 進表示するためのまるめ誤差 ) このことから

$$\|x-x'\|_2 \leq x \cdot (\|F\|_2/\|A\|_2 \cdot \|x\|_2 + \|b\|_2/\|A\|_2) / (1 - x \cdot \|F\|_2/\|A\|_2) \quad (2.1.8)$$

が得られる。ここで、 $\|x\|_2$ 、 $\|b\|_2$  はベクトルのユークリッドノルム、 $\|A\|_2$ 、 $\|F\|_2$  は行列のスペクトルノルム、 $x$  は行列ノルムとして、スペクトルノルムを用いたときの  $P(A)$  の値である。(  $x = \|A\|_2 \cdot \|A^{-1}\|_2$  )

### 2-2 条件数

以上見てきたように、解の誤差評価には、いつも  $P(A) = \|A\| \cdot \|A^{-1}\|$  という因子が現われ、これが行列のなんらかの性質を表現しているものと思われる。そこで、これを行列の条件数と呼び、数値実験においても、この値と解の精度との関係を見ることに主眼点を置いた。

行列ノルムには、ベクトルノルムに応じて、それに従属しているものが幾つかある。通常、ベクトルノルムとしては、ユークリッドノルムがよく用いられるので、行列ノルムとしてはスペクトルノルムが用いられる。従って、 $P(A)$  としては、 $x$  を採用することになる。

$x$  の性質を幾つか述べる。まず、行列  $A$  を定数倍しても、或はユニタリ一行列を掛けても、 $x$  の値は変化しない。又定義から明らかなように、 $x$  の値は常に 1 以上である。さらに、 $A$  が対称行列の場合には、 $x$  は次式により計算できる。

$$x = |\lambda_1| / |\lambda_n| \quad (2.2.1)$$

$\lambda_1$ :  $A$  の絶対値最大の固有値

$\lambda_n$ : " 最小の "

しかし、通常その値はアブリアリにはわからないし、行列の条件数が大きい — このとき、行列の条件は悪いと云う — ことと、行列が非正則に近いこととの関係がそれほど密接ではない。即ち、行列  $A$  の次数  $N$  を固定し、 $\|A\| = 1$  となるように  $A$  を正規化したとき、その絶対値最小の固有値を  $\lambda$  とすると、

$$|\lambda| \leq (N^{1/2(N+1)} / \|A^{-1}\|_2)^{1/N} \quad (2.2.2)$$

という結論が得られる。この式から、 $\|A^{-1}\|_2 (=x) \rightarrow \infty$  のとき、 $|\lambda| \rightarrow 0$  となることがわかる。しかし、それは  $\|A^{-1}\|_2$  の  $1/N$  のオーダーでしか 0 に収束しない。

### 2-8 行列のノルム

これまでの評価式の中でもしばしば用いたが、行列とベクトルとが同時に考察されるときには、行列ノルムとベクトルノルムとの間に、

$$\|Ax\| \leq \|A\| \cdot \|x\| \quad (2.3.1)$$

という不等式が成立することが要求される。任意のベクトルノルムに対して、(2.3.1)の不等式が成立するような行列ノルムを、“ベクトルノルムと両立する行列ノルム”という<sup>(9)</sup>さらに、ベクトルノルムと両立する行列ノルムのうちで最小のものを、“ベクトルノルムに従属する行列ノルム”という<sup>(9)</sup>このノルムは

$$\|A\| = \sup_{x \neq 0} \|Ax\| / \|x\| \quad (2.3.2)$$

で与えられる。

通常、ベクトルノルムとしては、ユークリッドノルムが用いられるが、このノルムと両立するような行列ノルムとしては、

$$\begin{aligned} \|A\|_\infty &= n \times \max_{i,j} |a_{ij}| && n: \text{行列の次数} \\ \|A\|_E &= \left( \sum_{i,j} |a_{ij}|^2 \right)^{1/2} \\ \|A\|_2 &= (A^*A \text{の最大固有値})^{1/2} \\ &&& A^*: A \text{の随伴行列} \end{aligned}$$

等がある。このうち、従属ノルムは $\|A\|_2$ である。

誤差等の評価では、上からおさえる数値はできるだけ小さい方が好ましい。そのため、行列ノルムとしては、ベクトルノルムに従属するものが採用される。従って、ベクトルノルムとしてユークリッドノルムを採用したときには、行列ノルムとしてはスペクトルノルム $\|A\|_2$ が用いられ、 $P(A)$ としては $x$ が用いられることになる。

### 3章 数値実験

#### 3-1 テストマトリックスの選択

数値実験としては、一次方程式を解くものではなく、逆行列を求めるものにした。この方が、定数項に対する考慮を必要としないので、行列の性質がみやすい。そこで、どのような行列をテストマトリックスとして採用するか、という問題がまず起きる。そのため、次のような方針でテストマトリックスを選んだ。

- (1) スペクトルノルムに関する条件数がわかっているか、もしくは計算可能なもの。
- (2) 逆行列の要素が、数式もしくは数値で、必要な精度で正確に与えられるもの。
- (3) 次数 $N$ を変えることによって、行列の大きさや条件数が、適当な範囲で変化するもの。
- (4) なるだけ質的に異なるものが望ましい。
- (5) 当然のことながら、正値対称行列でなければならない。

以上の方針でテストマトリックスを選んだところ、結果的には、行列自身及び逆行列が共に稠密なものとしては、一連のいわゆるヒルベルト行列 $H_N$ 、及びバイ行列 $P_N^d$ を、又帯状行列としては、2階常微分を中心差分して得

られる行列 $B_N$ 、及び $B_N$ を自乗した $B_N^2$ を採用した。さらに、最初に与える行列が、稠密であるか疎であるかによって、精度に影響する可能性を考えて、 $B_N^2$ の逆行列 $T_N^2$ についても実験した。

ここで、 $H_N$ 、 $P_N^d$ 、 $B_N$ 、 $B_N^2$ 及びそれらの逆行列は、次の通りである。

$$\begin{aligned} (1) \quad (H_N)_{ij} &= 1 / (i+j-1) \\ (H_N^{-1})_{ij} &= (-1)^{i+j} (N+i-1)! (N+j-1)! / \\ &\quad \{ (i+j-1) [(i-1)! (j-1)!]^2 (N-i)! (N-j)! \} \end{aligned}$$

$x(H_N)$ は、テーブルより計算

$$\begin{aligned} (2) \quad (P_N^d)_{ij} &= d \quad i=j \\ &= 1 \quad i \neq j \\ (P_N^d)^{-1} &= (d+N-2) / \{ (d-1)N + (d-1)^2 \} \\ &= -1 / \{ (d-1)N + (d-1)^2 \} \end{aligned}$$

$x(P_N^d) = 1 + N / (d-1)$

$$\begin{aligned} (3) \quad (B_N)_{ij} &= 2 \quad i=j \\ &= -1 \quad |i-j|=1 \\ &= 0 \quad |i-j| > 1 \\ (B_N^{-1})_{ij} &= i(N-j+1) / (N+1) \quad i < j \\ &= (B_N)_{ji} \quad i > j \\ x(B_N) &= \{ \sin(N\pi/2(N+1)) / \sin(\pi/2(N+1)) \}^2 \approx 4N^2 / \pi^2 \end{aligned}$$

$$\begin{aligned} (4) \quad (B_N^2)_{ij} &= 5 \quad i=j=1 \text{ or } i=j=N \\ &= 6 \quad 1 < i=j < N \\ &= -4 \quad |i-j|=1 \\ &= 1 \quad |i-j|=2 \\ &= 0 \quad |i-j| > 2 \\ (B_N^2)^{-1} &= \{ ijN(2N+1) - ij(i-1)(3N+2-i) \\ &\quad - j(j^2-1)(N+1-i) \} / 6(N+1) \quad i > j \\ &= (B_N^2)_{ji} \quad i < j \\ x(B_N^2) &= x(B_N)^2 \end{aligned}$$

これらの行列を実際に使うときには、入力データに誤差が入らぬようにするために、適当に定数倍して、要素が整数になるようにした。

#### 3-2 実験結果の整理方法

次に、誤差評価として、どのような数値を採用するかが問題になる。そのため、誤差行列 $R$ として、Newman & Todd<sup>(2)</sup>では、計算された逆行列を $X$ としたとき、 $XA-E$ 、 $AX-E$ 、 $X-A^{-1}$ 、 $X^{-1}-A$ の4つを挙げている。そして結果の整理には、 $R=AX-E=(r_{ij})$ を採用して、 $m = \max |r_{ij}|$ の値を与えている。さらにLietzke & Stoughton<sup>(3)</sup>では、 $R_1=XA-E=(r_{ij}^{(1)})$ 及び $R_2=A^{-1}-X=(r_{ij}^{(2)})$ を採用して、 $a=1/N^2 \sum |r_{ij}^{(1)}|$

$f=1/N \{ \sum (\gamma_{ij}^{(1)})^2 \}^{1/2}$ ,  $Q=1/N^2 \sum |\gamma_{ij}^{(2)}|$  の3つの値を与えている。

ところで、今  $A$  として  $H_3$  をとろう。  $A = \begin{pmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{pmatrix}$

そして、これの逆行列として、  $X_1 = \begin{pmatrix} 1.50042_{-1}, \\ -6.00232_{-1}, \\ 5.00223_{-1}, \end{pmatrix}$

$\begin{pmatrix} -6.00185_{-1}, & 5.00161_{-1} \\ 3.20103 & , & -3.00090 \\ -3.00099 & , & 3.00086 \end{pmatrix}$  及び、  $X_2 = \begin{pmatrix} 1.5_{-1}, \\ -5.99999_{-1}, \\ 5_{-1}, \end{pmatrix}$

$\begin{pmatrix} -6_{-1}, & 5_{-1} \\ 3.19999, & -2.99999 \\ -3 & , & 3 \end{pmatrix}$  の2つを比べてみる。正確な  $A^{-1}$

は、  $A^{-1} = \begin{pmatrix} 1.5_{-1}, & -6_{-1}, & 5_{-1} \\ -6_{-1}, & 3.2, & -3 \\ 5_{-1}, & -3, & 3 \end{pmatrix}$  だから、“ $X_1$  よりも

$X_2$  の方が  $A^{-1}$  の良い近似である”と云える。一方、  $R_i = AX_i - E$   $i=1,2$  として、それぞれの  $m$ ,  $a$ ,  $f$  を求めてみると、それら3つの値全部について、 $X_1$  の方が  $X_2$  よりも小さい。従って  $m$ ,  $a$ ,  $f$  等は誤差の評価として望ましいものではない、と考えられる。

そこで、われわれは  $R$  として、絶対誤差及び相対誤差行列を採用することにした。そして、それぞれの最大値  $m$ , 算術平均  $a$ , 2乗平均  $f$  の3つの値を求めた。なお、最大値と2乗平均は、行列の次数  $N$  を掛けることによって、行列のノルムになり得る。

結果の整理等では、以下のことに注意した。

- (1) 精度と条件数の関係を見るには、絶対誤差よりも、相対誤差の方が好ましい。
- (2) テストマトリックスとして  $T_N^2$  を用いると、その逆行列は  $B_N^2$  という帯状行列になる。このとき、零要素部分の相対誤差として、どのような数値を採用かが難しい。そのため、零要素部分と非零要素部分の2つに分け、零要素部分では、絶対誤差をそのまま相対誤差とした。条件数と精度の関係をみるのには、非零要素部分だけの数値を採用した。
- (3) 正定値行列の直接解法では、コレスキー法が最もポピュラーと思われるので、このアルゴリズムについて数値実験を重点的に行った。
- (4) なお、本章でも無断で用いたが、数の表記法で、「 $1.23_{-4}$ 」は「 $1.23 \times 10^{-4}$ 」を意味する。

## 4章 コレスキー法

### 4-1 方法の基礎

コレスキーの定理

正値対称行列  $A$  は、対角成分が正なる下三角行列  $L$  を用いて

$$A = LL^T \quad (4.1.1)$$

$L^T$ :  $L$  の転置行列

と分解される。このとき、 $L$  は一意的である。

さらに、対角成分が1なる下三角行列  $\tilde{L}$  と、対角成分が正なる対角行列  $D$  を用いて

$$A = \tilde{L} D \tilde{L}^T \quad (4.1.2)$$

とも分解される。

この定理そのものは、帰納法によって簡単に示される。行列  $L = (l_{ij})$  の要素は、行ごとに或は列ごとに順次計算できる。行ごとに求めると、

$$l_{ij} = (a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk}) / l_{jj} \quad 1 < j < i-1 \quad (4.1.3)$$

$$l_{ii} = (a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2)^{1/2} \quad 1 < i < N \quad (4.1.4)$$

同様に、 $\tilde{L} = (\tilde{l}_{ij})$   $D = (d_i)$  は、

$$\tilde{l}_{ij} = (a_{ij} - \sum_{k=1}^{j-1} \tilde{l}_{ik} d_k \tilde{l}_{jk}) / d_j \quad 1 < j < i-1 \quad (4.1.5)$$

$$d_i = a_{ii} - \sum_{k=1}^{i-1} \tilde{l}_{ik} d_k \tilde{l}_{ik} \quad 1 < i < N \quad (4.1.6)$$

となる。

(4.1.4) 式からわかるように

$$|l_{ik}| < \sqrt{a_{ii}} \quad 1 < k < i$$

と上からおさえられるので、計算の数値的安定性がすぐれている。一方、 $\tilde{l}_{ik}$  の方は、いくらでも大きくなり得る。

さらに、“分解ができる”といっても、理論上そうなるだけで、実際計算上ではまるめ誤差等の影響で、本来正であるべき  $l_{ii}^2$  や  $d_i$  が負になることがある。このとき、それ以上の分解は意味をもたなくなる。

### 4-2 数値的議論

仮数部が  $t$  桁の浮動小数点表示で計算した  $L$  は、

$$LL^T = A + F \quad \|F\|_2 < kn^{3/2} 2^{-t} \|A\|_2 \\ \|F\|_2 < kn^2 2^{-t} \max |a_{ij}|$$

を満足する。ここで、 $k$  はオーダー-1の定数であって、各々異なり得るが、一括して  $k$  を用いた。以後現われる  $k$  も同様である。

又、もしいわゆる内積計算  $\sum a_i b_i$  を倍精度で計算して、最後に単精度にまるめると、 $F$  の評価は、

$$\|F\|_2 < kn^{1/2} 2^{-t} \|A\|_2, \\ \|F\|_2 < kn 2^{-t} \max |a_{ij}|$$

と改善される。

$Ax = b$  の計算された解に誤差をもたらす原因は3つある。まず第一に、 $A$  を  $LL^T$  に分解するとき、ついで  $Ly = b$  及び  $L^T x = y$  を解くときである。後者の2つによる誤差の評価には  $x^{1/2}$  の因子があり、前者の評価には

$\epsilon$ の因子がある。従って $\epsilon$ が大きいときには、 $LL^T$ 分解による誤差が支配的になる。その結果、計算された解 $x$ 、及び計算された逆行列 $X$ の誤差評価は、 $\epsilon^{1/2}$ の項を $\epsilon$ の項に対して無視できるものとして、それぞれ

$$\|x - A^{-1}b\|_2 / \|A^{-1}b\|_2 < k n^{3/2} \epsilon^{1/2} \epsilon(A)$$

$$\|X - A^{-1}\|_2 / \|A^{-1}\|_2 < k n^{3/2} \epsilon^{1/2} \epsilon(A)$$

となる。

#### 4-3 演算回数

##### 1) $LL^T$ 分解

分解そのものに $N^3/6$ 回の乗除と加減、さらに $N$ 回の開平が、

$LL^T x = b$ を解くには、 $N^2$ 回の乗除と加減が $(LL^T)^{-1}$ の計算には、 $N^3/3$ の乗除と加減が、それぞれ必要である。

##### 2) $\tilde{L}D\tilde{L}^T$ 分解

分解そのものに $N^3/6$ 回の乗除と加減が、

$\tilde{L}D\tilde{L}^T x = b$ を解くには、 $N^2$ 回の乗除と加減が、

$(\tilde{L}D\tilde{L}^T)^{-1}$ の計算には、 $N^3/2$ 回の乗除と $N^3/3$ 回の加減が

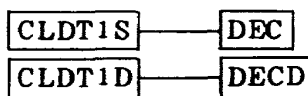
それぞれ必要である。

#### 4-4 サブルーチンの説明

サブルーチン I, IIは $LL^T$ 分解を、III, Nは $\tilde{L}D\tilde{L}^T$ 分解を用いたものである。I, IIIは $A$ を一次元アレイとして、II, Nは $A$ を2次元アレイとして取り扱っている。ここで扱う係数行列は対称行列なので、1次元アレイに格納するときには、その半分の要素だけを格納すればよく、従って $A$ のためのコアが半減される。その代わり、消費時間は多くなる。

各サブルーチン名の、第5文字目の数「1」は $A$ が1次元アレイであることを示し、「2」は $A$ が2次元アレイであることを示す。又、第6文字目の「S」はサブルーチンが単精度のものであることを示し、「D」は倍精度のものであることを示す。

なお、第1章でも述べたように、行列式の値をも求めているようなサブルーチンでは、アッセンブラプログラム“DEC”もしくは“DECD”をコールしている。サブルーチン I について、プログラムの構成を図示すると、











大きさ  $KB$  の 2次元アレイ

$KB$  : アレイ  $B$  の大きさを指定,  $KB > N$

整数型変数名または整数定数

$KX$  : アレイ  $X$  の大きさを指定,  $KX > N$

整数型変数名または整数定数

アウトプット

$X$  : 方程式  $AX = B$  の解  $X$  がストアされている。

$N \times NR$  行列

大きさ  $KX$  の 2次元アレイ

エラー処理

$N < 1$  であれば, 何も計算せず, そのままりターンする。 $ILL$  値のセットやメッセージの印字は行わない。

詳細説明

$B$  及び  $X$  は 2次元アレイだから, たとえ  $NR$  が 1 であっても, 2次元で指定する。

$B$  の値は保存されて, 解は  $X$  にストアされる。しかし, 仮引数  $B$  及び  $X$  の実引数を同じものにする, 解は  $B$  にストアされ,  $B$  の元の値は失われる。

$CLSL2D$  を用いるときには,  $A, DIL, B, X$  を倍精度実数型として指定する。

### (3) CLIV2S, CLIV2D

このサブルーチンは, 2次元アレイで指定された正値対称行列  $A$  を,  $A = LL^T$  とコレスキー分解し, ついで  $L^{-1}$  を求め,  $A^{-1} = L^{-1T} L^{-1}$  に従って  $A^{-1}$  を求めるものである。副産物として,  $A$  の行列式の値をも与える。

計算には,  $A$  の対角要素をも含めて下三角部分しか用いないので, その部分の値が正確に与えられていれば十分である。

元の  $A$  の値は保存される。

呼び出し形式

$CLIV2S(N, A, KA, DET, MDET, ILL)$

インプット

$N$  : 行列  $A$  の次数,  $N > 2$

整数型変数名または整数定数

$A$  :  $N \times N$  行列  $A$  の要素

大きさ  $KA$  の 2次元アレイ

配列の様子は, 図 4-5 を参照

$KA$  : アレイ  $A$  の大きさを指定,  $KA > N$

整数型変数名または整数定数

アウトプット

$A$  :  $A^{-1}$  の要素が, アレイ  $A$  の上三角部分にストアされている。

図 4-5 を参照

$DET, MDET$

$A$  の行列式の値を,  $DET \times 2^{MDET}$  の形で与える。

$DET$  は実数型で,  $1/2 < DET < 1$

$MDET$  は整数型

$ILL$  : 計算が正常に終了したときには,  $ILL = 0$  がセットされている。

エラーが生じたときには, 0 以外の値がセットされている。

エラー処理

$N < 1$  であれば,  $ILL = 10$  をセットして, なんらの計算も行わない。改頁して, その旨の印字をする。

分解途中で  $l_{ii}^2 < 0$  となって, コレスキー分解ができなくなったときには,  $ILL = 1$  がセットされ, そこから先の計算は行わない。改頁して, "SUB CLIV2S IS FAIL" と印字し,  $l_{ii}^2 < 0$  となったときの  $I$  の値も併せ印字する。このときの出力は, 意味のあるものではない。

詳細説明

インプット及びアウトプットのアレイ  $A$  のようす。

$A = (a_{ij}), A^{-1} = (a_{ij})$  とする。

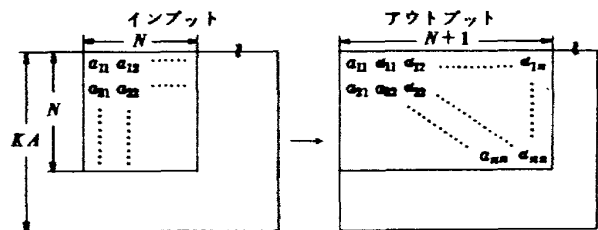


図 4-5

一般に  $a_{ij} = A(I, J+1)$

となっている。

インプットは  $N \times N$  行列であるが, アウトプットは  $N \times (N+1)$  行列であることに注意する。

$CLIV2D$  を用いるときには,  $A$  及び  $DET$  を倍精度実数型として指定する。

サブルーチン III

(1) SMDET1S, SMDET1D

このサブルーチンは、1次元アレイで指定された正値対称行列Aを、 $A = \tilde{L} D \tilde{L}^T$ と分解し、副産物として、Aの行列式の値をも与える。

元のAの値は失われる。

呼び出し形式

SMDET1S ( N, A, DET, MDET, ILL )

インプット

N : 行列Aの次数,  $N > 2$

整数型変数名または整数

A :  $N \times N$ 行列Aの要素

大きさ  $N(N+1)/2$  の1次元アレイ

配列の仕方については、図4-6を参照

アウトプット

A : Aを $\tilde{L} D \tilde{L}^T$ 分解して得られる単位下三角行列 $\tilde{L}$ の要素と、対角行列Dの要素の逆数がストアされている。

図4-7を参照

DET, MDET

Aの行列式の値を、 $DET \times 2^{MDET}$ の形で与える。

DETは実数型で、 $1/2 < DET < 1$

MDETは整数型

ILL :  $\tilde{L} D \tilde{L}^T$ 分解が正常に終了したときにはILL = 0がセットされている。

エラーが生じたときには、0以外の値がセットされている。

エラー処理

$N < 1$ であれば、ILL = 10をセットして、なんらの計算も行わない。改頁して、その旨の印字をする。

分解途中で、 $d_i < 0$ となって分解が意味をもたなくなったときには、ILL = 1がセットされ、そこから先の計算は行わない。改頁して、“SUB SMDET1S IS FAIL”と印字し、 $d_i < 0$ となったときのIの値も併せ印字する。このときの出力は、意味のあるものではない。

詳細説明

インプットAの配列の仕方

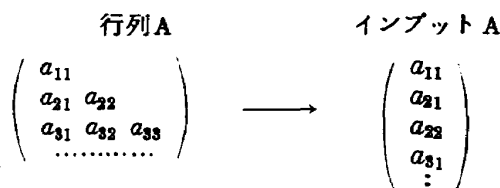


図4-6

一般に、行列Aの第(I, J)-要素は、1次元アレイAの第  $I * (I-1) / 2 + J$  番目の要素となっている。

アウトプットのAは、 $\tilde{L} = (\tilde{l}_{ij})$ ,  $D = (d_i)$ とする

$$(1/d_1, \tilde{l}_{21}, 1/d_2, \tilde{l}_{31}, \tilde{l}_{32}, 1/d_3, \tilde{l}_{41}, \dots)$$

図4-7

となっている。

SMDET1Dを用いるときには、A, DETを倍精度実数型として指定する。

(2) SMSL1S, SMSL1D

このサブルーチンは、正値対称行列Aを係数行列とする方程式  $AX = B$  を解くもので、 $A = \tilde{L} D \tilde{L}^T$ と分解し(SMDET1S)、ついで $\tilde{L}Y = B$ ,  $D\tilde{L}^T X = Y$ を解いて解Xを求める(SMSL1S)。

Aは1次元アレイであり、また解Xは、アレイBにストアされる。従ってBの元の値は失われる。

前項のサブルーチン(1)からのアウトプットAが、本サブルーチンのインプットAになるので、このサブルーチンを使用する前に、必ず(1)のサブルーチンを使用する。もしくは、それと同等な手続に従って、Aに図4-7のような情報をセットする。一度そのようにセットされれば、このサブルーチンを何度使用してもよい。

呼び出し形式

SMSL1S ( N, NR, A, B, KB )

インプット

N : 行列Aの次数,  $N > 2$

整数型変数名または整数

NR : 方程式  $AX = B$  の右辺の数,  $NR > 1$

整数型変数名または整数

A : 係数行列Aを $\tilde{L} D \tilde{L}^T$ 分解して得られた $\tilde{L}$ , Dの要素が図4-7のようにストアされている。

1次元アレイ

B : 方程式  $AX = B$  の右辺で、 $N \times NR$ 行列 大きさKBの2次元アレイ

KB : アレイBの大きさを指定,  $KB > N$

整数型変数名または整数

アウトプット

B : 方程式  $AX = B$  の解Xがストアされている。

エラー処理

$N < 1$ ならば、何も計算せずそのままリターンする。ILL値のセットやメッセージの印字は行わない。

## 詳細説明

$B$ は2次元アレイだから、たとえ $NR$ が1であっても2次元で指定しなければならない。

$B$ の元の値は失われ、新しく解がストアされる。

SMIS1Dを用いるときには、 $A$ 及び $B$ を倍精度実数型として指定する。

## (3) SMIV1S, SMIV1D

このサブルーチンは、1次元アレイで指定された正値対称行列 $A$ を、 $A = \tilde{L}D\tilde{L}^T$ と分解し、ついで $\tilde{L}^{-1}$ を求め、 $A^{-1} = \tilde{L}^{-1T}D^{-1}\tilde{L}^{-1}$ に従って $A^{-1}$ を求めるものである。副産物として、 $A$ の行列式の値をも与える。

元の $A$ の値は失われる。

呼び出し形式

SMIV1S ( $N$ ,  $A$ ,  $DET$ ,  $MDET$ ,  $ILL$ )

インプット

$N$ : 行列 $A$ の次数,  $N > 2$

整数型変数名または整数

$A$ :  $N \times N$ 行列 $A$ の要素

大きさ $N(N+1)/2$ の1次元アレイ

配列の仕方は、図4-8を参照

アウトプット

$A$ :  $A^{-1}$ の要素がストアされている。

図4-8を参照

$DET$ ,  $MDET$

$A$ の行列式の値を、 $DET \times 2^{MDET}$ の形で与える。

$DET$ は実数型で、 $1/2 < DET < 1$

$MDET$ は整数型

$ILL$ : 計算が正常に終了したときには、 $ILL = 0$ がセットされている。

エラーが生じたときには、0以外の値がセットされている。

エラー処理

$N < 1$ であれば、 $ILL = 10$ をセットして、なんらの計算も行わない。改頁して、その旨の印字をする。

$d_i < 0$ となって、分解が意味をもたなくなったときには、 $ILL = 1$ がセットされ、そこから先の計算は行わない。改頁して、“SUB SMIV1S IS FAIL”と印字する。 $d_i < 0$ となったときの $i$ の値も併せ印字する。このときの出力は、意味のあるものではない。

## 詳細説明

インプット $A$ の配列

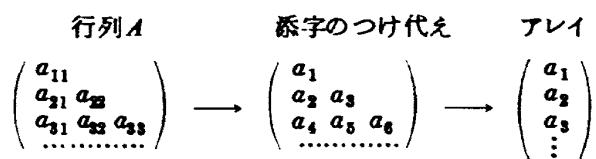


図4-8

アウトプットの逆行列も同じ配列の仕方になっている。

SMIV1Dを用いるときには、 $A$ 及び $DET$ を倍精度実数型として指定する。

サブルーチン N

(1) SMDT2S, SMDT2D

このサブルーチンは、2次元アレイで指定された正値対称行列  $A$  を、 $A = \tilde{L} D \tilde{L}^T$  と分解し、副産物として、 $A$  の行列式の値をも与える。

分解に際しては、 $A$  の対角要素をも含めて下三角部分しか用いないので、その部分の値が正確に与えられていれば十分である。

元の  $A$  の値は保存される。

呼び出し形式

SMDT2S ( N, A, KA, DIL, DET, MDET, ILL )

インプット

N : 行列  $A$  の次数,  $N > 2$

整数型変数名または整定数

A :  $N \times N$  行列  $A$  の要素

配列の仕方は、通常の行列表示と同じ。

図 4-9 を参照

大きさ  $KA$  の 2次元アレイ

KA : アレイ  $A$  の大きさを指定,  $KA > N$

整数型変数名または整定数

アウトプット

A :  $A$  を  $\tilde{L} D \tilde{L}^T$  分解して得られる単位下三角行列  $\tilde{L}$  の要素が、上三角部分にストアされている。

図 4-9 を参照

DIL : 対角行列  $D$  の要素が、逆数の形でストアされている。

大きさ  $N$  の 1次元アレイ

DET, MDET

$A$  の行列式の値を、 $DET \times 2^{MDET}$  の形で与える。

DET は実数型で、 $1/2 < DET < 1$

MDET は整数型

ILL :  $\tilde{L} D \tilde{L}^T$  分解が正常に終了したときには  $ILL = 0$  がセットされている。

エラーが生じたときには、0 以外の値がセットされている。

エラー処理

$N < 1$  であれば、 $ILL = 10$  をセットして、なんらの計算も行わない。改頁して、その旨の印字をする。

分解途中で、 $d_i < 0$  となって、分解が意味をもたなくなったときには、 $ILL = 1$  がセットされ、そこから先の計算は行わない。改頁して、“SUB SMDT2S IS FAIL” と印字し、 $d_i < 0$  となったときの  $I$  の値も併せ

印字する。このときの出力は、意味のあるものではない。

詳細説明

インプット、アウトプットの  $A$  及び  $DIL$

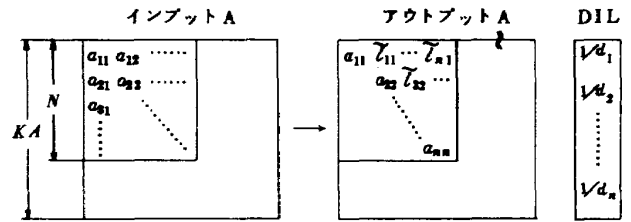


図 4-9

$\tilde{L} = (\tilde{l}_{ij})$ ,  $D = (d_i)$  とすると、

$\tilde{l}_{ij} = A(j, i)$ ,  $1/d_i = DIL(i)$

$1 < j < i < N$

となっている。

SMDT2D を用いるときには、 $A$ ,  $DIL$ ,  $DET$  を倍精度実数型として指定する。

(2) SMSL2S, SMSL2D

このサブルーチンは、正値対称行列  $A$  を係数行列とする方程式  $AX = B$  を解くもので、 $A = \tilde{L} D \tilde{L}^T$  と分解し (SMDT2S), ついて  $\tilde{L} Y = B$ ,  $D \tilde{L}^T X = Y$  を解いて解  $X$  を求める (SMSL2S)。

前項のサブルーチン(1)からのアウトプット  $A$  が、本サブルーチンのインプット  $A$  になる。従って、このサブルーチンを使用する前に、必ず(1)のサブルーチンを使用する。もしくは、それと同等な手続に従って、 $A$  に図 4-9 のような情報をセットする。一度そのようにセットされれば、以後このサブルーチンを何度使用してもよい。

解は、アレイ  $X$  にストアされるので、 $B$  の元の値は保存される。(後註参照)

呼び出し形式

SMSL2S ( N, NR, A, KA, DIL, B, KB, X, KX )

インプット

N : 行列  $A$  の次数,  $N > 2$

整数型変数名または整定数

A : 係数行列  $A$  を  $\tilde{L} D \tilde{L}^T$  分解して得られた  $\tilde{L}$  の要素が図 4-9 のようにストアされている。

大きさ  $KA$  の 2次元アレイ

KA : アレイ  $A$  の大きさを指定,  $KA > N$

整数型変数名または整定数

DIL : 対角行列  $D$  の要素が、逆数の形でストアされている。

大きさ  $N$  の 1次元アレイ

$B$  : 方程式  $AX=B$  の右辺,  $N \times NR$  行列  
 大きさ  $KB$  の 2次元アレイ  
 $KB$  : アレイ  $B$  の大きさを指定,  $KB > N$   
 整数型変数名または整数  
 $KX$  : アレイ  $X$  の大きさを指定,  $KX > N$   
 整数型変数名または整数

アウトプット

$X$  : 方程式  $AX=B$  の解  $X$  がストアされている。  
 $N \times NR$  行列  
 大きさ  $KX$  の 2次元アレイ

エラー処理

$N < 1$  であれば, 何も計算せずにそのままリターンする。 $ILL$  値のセットやメッセージの印字は行わない。

詳細説明

$B$  及び  $X$  は 2次元アレイだから, たとえ  $NR$  が 1 であっても, 2次元で指定する。

$B$  の値は保存されて, 解は  $X$  にストアされる。

仮引数  $B$  及び  $X$  の実引数を同じものにしてもよい。但し, このときには,  $B$  の元の値は失われる。

$SMISL2D$  を用いるときには,  $A, D, L, B, X$  を倍精度実数型として指定する。

(3) SMIV2S, SMIV2D

このサブルーチンは, 2次元アレイで指定された正値対称行列  $A$  を,  $A = \tilde{L}D\tilde{L}^T$  と分解し, ついで  $\tilde{L}^{-1}$  を求め,  $A^{-1} = \tilde{L}^{-1T}D^{-1}\tilde{L}^{-1}$  に従って  $A^{-1}$  を求めるものである。

副産物として,  $A$  の行列式の値をも与える。

計算には,  $A$  の対角要素をも含めて下三角部分しか用いないので, その部分の値が正確に与えられていれば十分である。

元の  $A$  の値は保存される。

呼び出し形式

$SMIV2S(N, A, KA, DET, MDET, ILL)$

インプット

$N$  : 行列  $A$  の次数,  $N > 2$   
 整数型変数名または整数  
 $A$  :  $N \times N$  行列  $A$  の要素  
 大きさ  $KA$  の 2次元アレイ

配列のようすは, 図 4-10 を参照

$KA$  : アレイ  $A$  の大きさを指定,  $KA > N$   
 整数型変数名または整数

アウトプット

$A$  :  $A^{-1}$  の要素が, アレイ  $A$  の上三角部分にストアされている。

図 4-10 を参照

$DET, MDET$

$A$  の行列式の値を,  $DET \times 2^{MDET}$  の形で与える。

$DET$  は実数型で,  $1/2 < DET < 1$

$MDET$  は整数型

$ILL$  : 計算が正常に終了したときには,  $ILL=0$  がセットされている。

エラーが生じたときには, 0 以外の値がセットされている。

エラー処理

$N < 1$  であれば,  $ILL=10$  をセットして, なんらの計算も行わない。改頁して, その旨の印字をする。

$\tilde{L}D\tilde{L}^T$  分解の途中で,  $d_i < 0$  となって, 分解が意味をもたなくなったときには,  $ILL=1$  をセットして, そこから先の計算は行われぬ。改頁して, "SUB SMIV2S IS FAIL" と印字し,  $d_i < 0$  となったときの  $I$  の値も併せ印字する。このときの出力は, 意味のあるものではない。

詳細説明

インプット, アウトプットのアレイ  $A$  のようす  $A = (a_{ij}), A^{-1} = (a_{ij})$  とする。

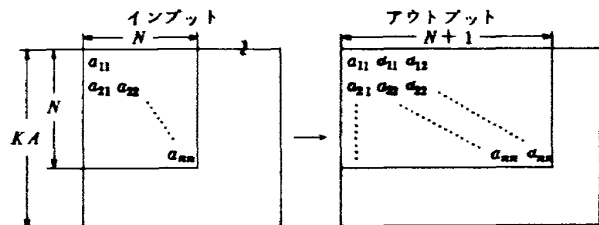


図 4-10

一般に,  $a_{ij} = A(I, J+1)$  となっている。

インプットは  $N \times N$  行列であるが, アウトプットは  $N \times (N+1)$  行列になっている。

$SMIV2D$  を用いるときには,  $A$  及び  $DET$  を倍精度実数型として指定する。

4-5 用途

ここに掲げたサブルーチンの直接的用途は自明である。演算回数という点からは、 $\tilde{L}D\tilde{L}^T$  分解の方が若干すくれている。

$LL^T$  分解は、固有値を求める RUTISHAUSER の LR-アルゴリズムで応用される。

又、 $LL^T$  分解は、固有値問題

$$\det(A - \lambda B) = 0$$

$$\det(AB - \lambda E) = 0$$

ここで、 $A$  : 対称行列

$B$  : 正値対称行列

の解を求める際にも用いられる。実際、 $B = LL^T$  と分解すると、これらの問題は、

$$\det\{L^{-1}A(L^{-1})^T - \lambda E\} = 0$$

$$\det\{L^TAL - \lambda E\} = 0$$

ここで、 $L^{-1}A(L^{-1})^T$  : 対称行列

$L^TAL$  : 対称行列

という通常の固有値問題に帰着できる。

時間は少し多くかかっていいが、メモリーの節約をしたい、というときにはサブルーチン I 或は III を用いて、 $A$  を 1 次元アレイとして扱う。又、インプット行列に関する情報を保存しておきたい場合には、サブルーチン II 或は IV を用いて、 $A$  を 2 次元アレイとして扱う。

4-6 数値実験及び結果

逆行列を計算して種々の数値を比較する方針を採ったので、サブルーチンとしては、CLIV2S 及び CLIV2D を用いた。消費時間の差を比較するために、若干の行列について、CLIV1S、CLIV1D を用いた場合の消費時間を並記してある。

表中には、絶対誤差・相対誤差のそれぞれ最大値・算術平均値・2乗平均値が載せてある。ただし、このサブルーチンでは、逆行列要素のうち右上部分だけしか計算していないので、上記3つの値も  $N(N+1)/2$  の値について計算してある。 $T_N^2$  については、3-2の(2)を参照。

「時間」の欄の数値は、逆行列を計算するのに要した時間である。2つ書いてある場合は、左側が CLIV2S による消費時間、右側が CLIV1S による消費時間である。記号“-”は、消費時間が 100 msec 以下であることを示す。又、記号“×”は、時間を測ってないことを示す。

次に、第2章の(2.1.5)-式を背景にして、計算結果の精度  $n$  桁、マシン精度  $K$  (仮数部  $t$  桁の浮動小数点表示の場合、 $K=2^t$ )、条件数  $P$  との間

$$n = \log K - \log P \tag{4.6.1}$$

なる関係式が、大略成立することが以前から指摘されている。そのためここでは、 $P$  として  $\chi$  (第2章参照) を採り、精度  $n$  桁の  $n$  は通常整数であるが、計算機の扱いなどで 7.8 桁という表現も行われており、又ある程度の量的関係をみるために、 $n$  として相対誤差の最大値  $m$ 、自乗平均  $f$  の対数の負数を採用した。そして各表に、 $\log K - \log \chi$ 、 $-\log m$ 、 $-\log f$  の値を載せておいた。

なお若干の比較の為に、同じサブルーチンで HITAC-5020 を使用した際の、相対誤差の最大値と消費時間も載せておく。このときのサブルーチンでは、行列式の値は計算していない。又、プログラム自体が必ずしも効率化されていないので、消費時間は多い目になっている。

4-4 の演算回数でみるとその主要項は  $N^3$  である。このことから、CLIV2S に要する「時間」も  $N^3$  に依存するだろうことが推察される。そしてこのことは、上に掲げた表によって、まず成立しているといつてよい。

式(4.6.1)については、蓋然性のある結論として、概ね成立しているといえる。 $N$  が小さいのに条件数が非常に大きい行列  $H_N$  では、(4.6.1)-式から得られる数値よりも 1 桁以上良い。又、 $N$  によって条件数の変化しない行列  $P_N^{(N+1)}$  では、 $N$  が大きくなるほど精度が悪くなる。これは演算回数が多くなることによる。まるめ誤差の累積に依ると思われる。

精度を、HITAC-5020 のときと比較すると、1~2 桁良くなっている。これは、FACOM 230/75 の方が仮数部が 3 ビット長く、10進精度で 1 桁良くなるのは期待される場所である。

なお、FACOM 230/75 の FORTRAN には D、H-OPT0、H-OPT2 の 3 種があるが、これらによる結果の表については、7-4 を参照。OPT0 と OPT2 の比較では、CLIV2S の消費時間は  $1/3$  以下になっている。又、 $H_N$  を使ったテストでは、OPT2 の方が精度も若干上回っている。

5章 逐次法による解の精密化

5-1 方法の基礎

正値対称行列  $A$  を係数行列とする方程式  $Ax = b$  の精密解を、次のような逐次法によって得る。

$$\begin{aligned} A &= LL^T, \quad x^{(0)} = 0 \\ r^{(s)} &= b - Ax^{(s)}, \quad (LL^T)d^{(s)} = r^{(s)} \\ x^{(s+1)} &= x^{(s)} + d^{(s)} \end{aligned} \tag{5.1.1}$$

この過程は、 $A$  が非常に大きい条件数をもたなければ、 $r^{(s)}$  を倍精度で計算することによって、必要な程度の精度でもって収束することが示されている。実際、4章



表4-1  $H_N$ : CLIV2S

FORTRAN-H OPT-0

		$N$	3	4	5	6	7
絶対誤差	$m$		1.192 <sub>-7</sub>	2.337 <sub>-8</sub>	2.185 <sub>-2</sub>	5.497 <sub>-1</sub>	6.901
	$a$		7.140 <sub>-8</sub>	7.175 <sub>-8</sub>	5.108 <sub>-3</sub>	1.216 <sub>-1</sub>	1.265
	$f$		8.615 <sub>-8</sub>	1.006 <sub>-3</sub>	8.095 <sub>-3</sub>	2.080 <sub>-1</sub>	2.306
相対誤差	$m$		7.450 <sub>-8</sub>	2.392 <sub>-6</sub>	3.378 <sub>-4</sub>	3.702 <sub>-3</sub>	2.192 <sub>-2</sub>
	$a$		5.008 <sub>-8</sub>	1.810 <sub>-6</sub>	2.849 <sub>-4</sub>	3.122 <sub>-3</sub>	1.781 <sub>-2</sub>
	$f$		5.182 <sub>-8</sub>	1.846 <sub>-6</sub>	2.892 <sub>-4</sub>	3.178 <sub>-3</sub>	1.804 <sub>-2</sub>
log $K - \log z$			5.107	3.636	2.149	6.521 <sub>-1</sub>	8.502 <sub>-1</sub>
- log $m$			7.128	5.621	3.471	2.432	1.659
- log $f$			7.285	5.734	3.539	2.498	1.744
HITAC-5020			5.007 <sub>-8</sub>	8.148 <sub>-8</sub>	1.377 <sub>-3</sub>	7.306 <sub>-2</sub>	6.460 <sub>-1</sub>

表4-2  $H_N$ : CLIV2D

FORTRANH, OPT-0

		$N$	5	6	7	8	9
絶対誤差	$m$		3.457 <sub>-13</sub>	2.125 <sub>-11</sub>	4.480 <sub>-9</sub>	4.044 <sub>-6</sub>	3.739 <sub>-7</sub>
	$a$		8.053 <sub>-14</sub>	4.702 <sub>-12</sub>	8.211 <sub>-10</sub>	6.027 <sub>-7</sub>	4.893 <sub>-8</sub>
	$f$		1.278 <sub>-13</sub>	8.044 <sub>-12</sub>	1.498 <sub>-9</sub>	1.167 <sub>-6</sub>	1.017 <sub>-7</sub>
相対誤差	$m$		5.130 <sub>-15</sub>	1.413 <sub>-13</sub>	1.380 <sub>-11</sub>	3.982 <sub>-10</sub>	4.386 <sub>-11</sub>
	$a$		4.409 <sub>-15</sub>	1.203 <sub>-13</sub>	1.139 <sub>-11</sub>	3.254 <sub>-10</sub>	2.471 <sub>-11</sub>
	$f$		4.487 <sub>-15</sub>	1.225 <sub>-13</sub>	1.155 <sub>-11</sub>	3.298 <sub>-10</sub>	2.735 <sub>-11</sub>
log $K - \log z$			12.68	11.19	9.686	8.179	6.670
- log $m$			14.29	12.85	10.86	9.400	10.36
- log $f$			14.35	12.91	10.94	9.482	10.56
HITAC-5020			1.279 <sub>-13</sub>	8.822 <sub>-12</sub>	5.808 <sub>-10</sub>	1.491 <sub>-8</sub>	1.350 <sub>-6</sub>

表4-3  $P_N^{(2)}$ : CLIV2S

FORTRAN-D

		$N$	10	50	100	200	300	400
絶対誤差	$m$		5.960 <sub>-8</sub>	1.043 <sub>-7</sub>	1.043 <sub>-7</sub>	2.235 <sub>-7</sub>	2.086 <sub>-7</sub>	2.384 <sub>-7</sub>
	$a$		7.214 <sub>-9</sub>	2.438 <sub>-9</sub>	1.323 <sub>-9</sub>	1.198 <sub>-9</sub>	8.961 <sub>-10</sub>	8.346 <sub>-10</sub>
	$f$		1.352 <sub>-8</sub>	7.434 <sub>-9</sub>	4.573 <sub>-9</sub>	8.081 <sub>-9</sub>	8.048 <sub>-9</sub>	7.491 <sub>-9</sub>
相対誤差	$m$		1.490 <sub>-7</sub>	7.153 <sub>-7</sub>	1.758 <sub>-6</sub>	3.874 <sub>-6</sub>	6.020 <sub>-6</sub>	7.033 <sub>-6</sub>
	$a$		3.387 <sub>-8</sub>	6.894 <sub>-8</sub>	8.604 <sub>-8</sub>	1.119 <sub>-7</sub>	9.799 <sub>-8</sub>	1.592 <sub>-7</sub>
	$f$		4.759 <sub>-8</sub>	9.191 <sub>-8</sub>	1.113 <sub>-7</sub>	1.447 <sub>-7</sub>	1.403 <sub>-7</sub>	2.443 <sub>-7</sub>
時間 sec			—	0.2 ×	1.5 1.9	12.2 15.9	40.8 55.7	× 137.8
log $K - \log z$			6.785	6.119	5.822	5.524	5.348	5.224
- log $m$			6.827	6.146	5.755	5.412	5.220	5.153
- log $f$			7.322	7.037	6.953	6.840	6.853	6.612
HITAC-5020 sec			1.505 <sub>-6</sub>	6.519 <sub>-6</sub> 2	1.441 <sub>-5</sub> 15	4.008 <sub>-5</sub> 126		

表 4-4  $P_N^{(N+1)}$ ; CLIV2S

FORTRAN-D

		$N$	10	50	100	200	300	400
絶対誤差	$m$		5.588 <sub>-9</sub>	1.863 <sub>-9</sub>	9.313 <sub>-10</sub>	8.149 <sub>-10</sub>	6.985 <sub>-10</sub>	2.328 <sub>-10</sub>
	$a$		5.249 <sub>-10</sub>	3.392 <sub>-11</sub>	9.536 <sub>-12</sub>	4.843 <sub>-12</sub>	2.957 <sub>-12</sub>	7.967 <sub>-13</sub>
	$f$		1.169 <sub>-9</sub>	1.396 <sub>-10</sub>	5.479 <sub>-11</sub>	4.386 <sub>-11</sub>	3.288 <sub>-11</sub>	7.227 <sub>-12</sub>
相対誤差	$m$		1.192 <sub>-7</sub>	2.384 <sub>-7</sub>	2.686 <sub>-7</sub>	3.278 <sub>-7</sub>	3.725 <sub>-7</sub>	3.874 <sub>-7</sub>
	$a$		3.820 <sub>-8</sub>	6.628 <sub>-8</sub>	6.782 <sub>-8</sub>	7.179 <sub>-8</sub>	8.926 <sub>-8</sub>	1.169 <sub>-7</sub>
	$f$		4.586 <sub>-8</sub>	8.206 <sub>-8</sub>	8.254 <sub>-8</sub>	9.610 <sub>-8</sub>	1.115 <sub>-7</sub>	1.349 <sub>-7</sub>
時間	sec		—	0.2 ×	1.5 1.9	12.2 15.9	40.7 55.6	× 137.3
log $K - \log \lambda$			7.526	7.526	7.526	7.526	7.526	7.526
- log $m$			6.924	6.623	6.681	6.484	6.429	6.412
- log $f$			7.339	7.086	7.083	7.017	6.953	6.870

表 4-5  $B_N$ ; CLIV2S

FORTRAN-D

		$N$	10	35	50	100	200	300
絶対誤差	$m$		2.384 <sub>-7</sub>	5.722 <sub>-6</sub>	1.621 <sub>-5</sub>	9.203 <sub>-5</sub>	1.343 <sub>-3</sub>	4.452 <sub>-3</sub>
	$a$		5.368 <sub>-8</sub>	2.466 <sub>-6</sub>	5.419 <sub>-6</sub>	3.889 <sub>-5</sub>	5.011 <sub>-4</sub>	1.847 <sub>-3</sub>
	$f$		7.204 <sub>-8</sub>	2.919 <sub>-6</sub>	6.782 <sub>-6</sub>	4.648 <sub>-5</sub>	6.245 <sub>-4</sub>	2.251 <sub>-3</sub>
相対誤差	$m$		1.043 <sub>-7</sub>	1.386 <sub>-6</sub>	2.027 <sub>-6</sub>	8.687 <sub>-6</sub>	4.715 <sub>-5</sub>	1.269 <sub>-4</sub>
	$a$		5.093 <sub>-8</sub>	8.611 <sub>-7</sub>	1.311 <sub>-6</sub>	5.220 <sub>-6</sub>	3.226 <sub>-5</sub>	8.225 <sub>-5</sub>
	$f$		6.008 <sub>-8</sub>	9.069 <sub>-7</sub>	1.398 <sub>-6</sub>	5.508 <sub>-6</sub>	3.395 <sub>-5</sub>	8.645 <sub>-5</sub>
時間	sec		—	—	0.2	1.4	11.5	38.9
log $K - \log \lambda$			6.142	5.107	4.804	4.210	3.613	3.262
- log $m$			6.982	5.858	5.693	5.061	4.327	3.897
- log $f$			7.221	6.042	5.854	5.259	4.469	4.063

表 4-6  $B_N^2$ ; CLIV2S

FORTRAN-D

		$N$	10	17	40	50	80	100
絶対誤差	$m$		1.941 <sub>-4</sub>	9.495 <sub>-3</sub>	1.169	4.857	1.729 <sub>+2</sub>	9.072 <sub>+2</sub>
	$a$		9.092 <sub>-5</sub>	4.265 <sub>-3</sub>	5.003 <sub>-1</sub>	2.105	7.202 <sub>+1</sub>	3.744 <sub>+2</sub>
	$f$		1.052 <sub>-4</sub>	5.026 <sub>-3</sub>	6.029 <sub>-1</sub>	2.534	8.776 <sub>+1</sub>	4.578 <sub>+2</sub>
相対誤差	$m$		8.091 <sub>-6</sub>	9.052 <sub>-5</sub>	1.000 <sub>-3</sub>	2.261 <sub>-3</sub>	1.909 <sub>-2</sub>	5.100 <sub>-2</sub>
	$a$		6.596 <sub>-6</sub>	7.676 <sub>-5</sub>	8.288 <sub>-4</sub>	1.845 <sub>-3</sub>	1.588 <sub>-2</sub>	4.271 <sub>-2</sub>
	$f$		6.733 <sub>-6</sub>	7.734 <sub>-5</sub>	8.370 <sub>-4</sub>	1.858 <sub>-3</sub>	1.600 <sub>-2</sub>	4.303 <sub>-2</sub>
時間	sec		—	—	0.1	0.2	0.7	1.5
log $K - \log \lambda$			4.458	3.595	2.161	1.782	9.775 <sub>-1</sub>	5.941 <sub>-1</sub>
- log $m$			5.092	4.043	3.000	2.646	1.719	1.292
- log $f$			5.172	4.112	3.077	2.731	1.796	1.366

表4-7  $B_N^2$ ; CLIV2D

FORTRAN-H, OPT-0

	$N$	50	80	100	150	200
	絶対誤差	$m$	3.159 <sub>-11</sub>	9.276 <sub>-10</sub>	4.711 <sub>-9</sub>	5.058 <sub>-8</sub>
$a$		1.284 <sub>-11</sub>	3.921 <sub>-10</sub>	1.930 <sub>-9</sub>	2.083 <sub>-8</sub>	1.711 <sub>-7</sub>
$f$		1.572 <sub>-11</sub>	4.770 <sub>-10</sub>	2.363 <sub>-9</sub>	2.549 <sub>-8</sub>	2.095 <sub>-7</sub>
相対誤差	$m$	1.252 <sub>-14</sub>	1.039 <sub>-13</sub>	2.602 <sub>-13</sub>	8.738 <sub>-13</sub>	3.038 <sub>-12</sub>
	$a$	1.093 <sub>-14</sub>	8.660 <sub>-14</sub>	2.195 <sub>-13</sub>	7.211 <sub>-13</sub>	2.519 <sub>-12</sub>
	$f$	1.102 <sub>-14</sub>	8.726 <sub>-14</sub>	2.211 <sub>-13</sub>	7.265 <sub>-13</sub>	2.536 <sub>-12</sub>
時間 sec		0.5	2.0	4.0	13.3	31.6
$\log K - \log x$		12.32	11.51	11.13	10.43	9.935
	$-\log m$	13.90	12.98	12.58	12.06	11.52
	$-\log f$	13.96	13.06	12.66	12.14	11.60
HITAC-5020 sec		1.759 <sub>-11</sub> 3	1.213 <sub>-10</sub> 9	3.003 <sub>-10</sub> 18	1.403 <sub>-9</sub> 60	

のサブルーチンで得られた解に誤差が生じるのは、まるめ誤差に起因するわけだから、上述の逐次法が実際に収束するという、アプリオリな判定はないわけである。その収束のための条件が、 $A$ の条件数があまり大きくないこと、及び $\gamma^{(s)}$ を倍精度で計算することである。

このとき、 $Ax = b$ の計算された解は、 $(A + \delta A)x = b$ の正確な解になっており、 $\delta A$ は $b$ に無関係に有界である。即ち

$$\|\delta A\|_2 < M; M \text{ indep. of } b \quad (5.1.2)$$

このことから、

$$\|A^{-1} \delta A\|_2 < M \cdot \|A^{-1}\|_2 = \alpha < 1 \quad (5.1.3)$$

であれば

$$\|x - x^{(s+1)}\|_2 < \frac{\alpha}{1-\alpha} \|x - x^{(s)}\|_2 \quad (5.1.4)$$

となって、 $\alpha < 1/2$ であれば収束する。

逆行列の逐次計算による精密化は、

$$X^{(m+1)} = X^{(m)} + X^{(m)} (E - AX^{(m)}) \quad (5.1.5)$$

に従って行われる。ここでも、収束のための条件は、 $A$ の条件数があまり大きくないことと、 $E - AX^{(m)}$ を倍精度で計算することである。

(5.1.5) -式を変形すると

$$E - AX^{(m+1)} = (E - AX^{(m)})^2 \quad (5.1.6)$$

となり、初期近似 $X^{(0)}$ が

$$\|E - AX^{(0)}\|_2 = s < 1 \quad (5.1.7)$$

を満足すれば

$$\|X^{(m)} - A^{-1}\|_2 \leq \|X^{(0)}\|_2 \cdot \frac{s^{2m}}{1-s}$$

となって、2乗のオーダーで収束する。

### 5-2 演算回数

#### 1) $Ax = b$ を解く場合

ACLSLSの1ループにつき、 $2N^2 + N$ 回の乗除、及び $2N^2$ 回の加減が必要となる。従って、右辺が $M$ コのと

き、 $L$ 回ループをまわったとすると、

$$N \text{ 回の開平} \\ N^3/6 + MN^2 + ML(2N^2 + N) \text{ 回の乗除} \\ N^3/6 + M(N^2 + 2LN^2) \text{ 回の加減}$$

が必要である。

#### 2) 逆行列を求める場合

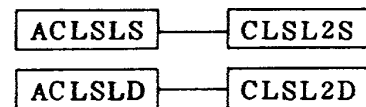
ACLIVSの1ループにつき、 $2N^3$ 回の乗除と加減が必要となる。従って、 $L$ 回ループをまわったとすると、全部で

$$N \text{ 回の開平} \\ N^3/2 + 2LN^3 \text{ 回の乗除と加減}$$

が必要である。

### 5-3 サブルーチンの説明

逐次法の式からわかるように、この方法では、元の $A$ 及び $B$ の値を必要とするので、第4章で述べたサブルーチンのうち、IIもしくはNを使用することになる。実際には、ACLSLSの中で、CLSL2Sをコールしているだけである。このプログラムの構成は、



となっている。

内積を倍精度で計算する部分をアセンブラで書けば、元のFORTRANプログラムのステップ数は格段に減少するし、消費時間も大巾に短縮される。

サブルーチン V

(1) ACLSLS, ACLSLD

逐次式 (5.1.1) のうち

$$\begin{aligned} r^{(s)} &= b - Ax^{(s)} && \text{倍精度計算} \\ (LL^T)d^{(s)} &= r^{(s)} && \text{CLSLS2Sを使用} \\ x^{(s+1)} &= x^{(s)} + d^{(s)} \end{aligned}$$

という3つの過程を繰り返して計算する。

このサブルーチンを使用する前に、CLDT2Sによって、係数行列  $A$  をコレスキー分解しておく。

インプットデータの値は変化しない。

呼び出し形式

ACLSLS (  $N, NR, A, KA, DIL, B, KB, X,$   
 $KX, ITS, BB, KBB, EPS, ILL$  )

インプット

$N$  : 行列  $A$  の次数,  $N > 2$

整数型変数名または整数定数

$NR$  : 方程式  $AX=B$  の右辺の数,  $NR > 1$

整数型変数名または整数定数

$A$  : 係数行列  $A$  の下三角部分及び、 $A$  をコレスキー分解して得られる  $L$  の非対角要素が、図 5-1 のようにストアされている。

大きさ  $KA$  の 2次元アレイ

$KA$  : アレイ  $A$  の大きさを指定,  $KA > N$

整数型変数名または整数定数

$DIL$  :  $L$  の対角要素の逆数がストアされている

大きさ  $N$  の 1次元アレイ

$B$  : 方程式  $AX=B$  の右辺,  $N \times NR$  行列

大きさ  $KB$  の 2次元アレイ

$KB$  : アレイ  $B$  の大きさを指定,  $KB > N$

整数型変数名または整数定数

$KX$  : アレイ  $X$  の大きさを指定,  $KX > N$

整数型変数名または整数定数

$KBB$  : アレイ  $BB$  の大きさを指定,  $KBB > N$

整数型変数名または整数定数

$EPS$  : 収束判定値

標準的には、単精度で  $10^{-5}$ 、倍精度で  $10^{-14}$

詳細説明参照

アウトプット

$X$  : 方程式  $AX=B$  の解がストアされている。

$N \times NR$  行列

大きさ  $KX$  の 2次元アレイ

$ITS$  : イタレーションの回数

整数型変数名

$BB$  :  $(LL^T)d^{(s)}=r^{(s)}$  の解  $d^{(s)}$  がストアされてい

る。

ワーキングエリアとして使用

$N \times NR$  行列

大きさ  $KBB$  の 2次元アレイ

$ILL$  : イタレーションが収束して、必要な程度の精密解が得られたときには、 $ILL=0$  がセットされている。

精密化が行えなかったときには、 $ILL=2$  がセットされている。

エラー処理

$N < 1$  であれば、何も計算せずにそのままリターンする。 $ILL$  値のセットやメッセージの印字は行わない。

解の精密化に失敗したときには、 $ILL=2$  がセットされ、改頁して、“SUB ACLSLS IS FAIL” と印字される。それまでに要したイタレーションの回数も併せて印字する。 $X$  には最終近似解がストアされている。

詳細説明

$EPS$

仮数部  $t$  桁の浮動小数点表示の場合

$$EPS > 2^{1-t}$$

であるが、 $EPS$  が小さすぎると、無駄なイタレーションをしたり、或は十分近似解が得られていたのに、もう一度イタレーションループに入ってしまった、次には  $ILL=2$  がセットされてしまった、という事態が生じることがある。

収束及び  $ILL$  判定

$$\alpha^{(s)} = \max_{\substack{1 < i < N \\ 1 < j < NR}} |a_{ij}^{(s)}|, \quad \beta^{(s)} = \max_{\substack{1 < i < N \\ 1 < j < NR}} |x_{ij}^{(s)}|$$

としたとき、

$$\alpha^{(s)} < 2 \times EPS \times \beta^{(s+1)}$$

となったときに、収束と判定する。また

$$\alpha^{(s)} / \beta^{(s+1)} > 1/2 \times \alpha^{(s-1)} / \beta^{(s)}$$

$$\text{但し、} \alpha^{(0)} / \beta^{(1)} = 3$$

となったときに、 $ILL$  と判定する。

インプット  $A$  の配列の様子

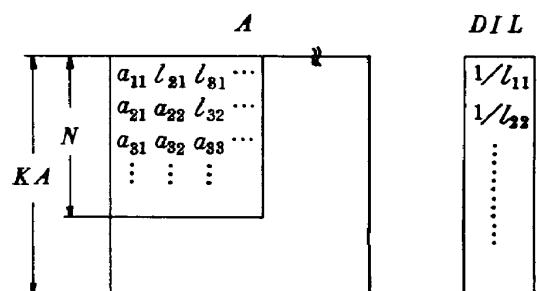


図 5-1

ACLSD を用いるときには、 $A, DIL, B, X, BB$

EPSを倍精度実数型として指定する。

仮引数BとXの実引数を同一のものにしてはならない。(CLSL2Sを参照。逐次式の形から、Bの値を保存する必要のあることがわかる。)

## (2) ACLIVS, ACLIVD

このサブルーチンは、正値対称行列Aの近似逆行列 $X^{(0)}$ を、逐次式(5.1.5)に従って精密化するものである。そのため、前もってなんらかの手続に従って、近似逆行列 $X^{(0)}$ を求めておく。(CLIV2SやSMIV2S等を用いる)

呼び出し形式

ACLIVS( N, A, KA, ITS, B, KB, EPS,  
ILL )

インプット

N : 行列Aの次数,  $2 \leq N \leq 350$

整数型変数名または整数値

A : 行列Aの要素がアレイAの下三角部分に、近似逆行列の要素がアレイAの上三角部分にストアされている。図5-2を参照

大きさKAの2次元アレイ

KA : アレイAの大きさを指定,  $KA \geq N$

整数型変数名または整数値

KB : アレイBの大きさを指定,  $KA \geq N$

整数型変数名または整数値

EPS : 収束判定値

標準的には、単精度で $10^{-5}$ 、倍精度で $10^{-14}$

詳細説明参照

アウトプット

A : 逐次法で精密化された逆行列要素が、アレイAの上三角部分にストアされている。

ITS : イタレイションの回数

整数型変数名

B : 最終剰余  $B = E - AX^{(s)}$  がストアされている。

ワーキングエリアとして使用

$N \times N$  行列

大きさKBの2次元アレイ

ILL : イタレイションが収束した場合には、ILL=0がセットされている。

精密化が行えなかったときには、ILL=2がセットされている。

エラー処理

$N < 1$  または  $N > 350$  であれば、何も計算せずにそのままリターンする。ILLのセットやメッセージの印字は

行わない。

解の精密化に失敗したときには、ILL=2がセットされ、改頁して、“SUB ACLIVS IS FAIL”と印字される。それまでに要したイタレイションの回数も併せて印字する。

### 詳細説明

EPSについては前項のサブルーチンの詳細説明を参照。

収束及びILL判定

$$X^s = (x_{ij}^{(s)}), \quad Z^s = (z_{ij}^{(s)}) = X^s (E - AX^s)$$

$$\alpha^{(s)} = \max_{1 \leq i, j < N} |x_{ij}^{(s)}|, \quad \beta^{(s)} = \max_{1 \leq i, j < N} |z_{ij}^{(s)}|$$

としたとき

$$\alpha^{(s)} \leq 2 \times EPS \times \beta^{(s)}$$

となったときに収束と判定する。また

$$\alpha^{(s)} / \beta^{(s)} > \frac{1}{2} \times \alpha^{(s-1)} / \beta^{(s-1)}$$

$$\text{但し, } \alpha^{(0)} / \beta^{(0)} = 1$$

となったときに、ILLと判定する。

近似逆行列を  $X = (x_{ij})$  としたとき、インプットアレイAの配列の仕方。

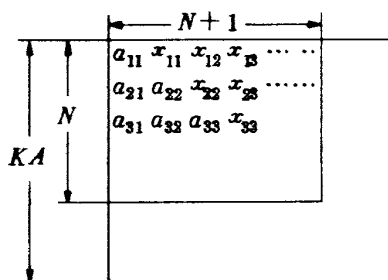


図5-2

イタレイションが収束した場合には、収束解は同じくアレイの右上部分にストアされている。

行列の次数Nに上限があるのは、サブルーチン内で1次元のワーキングエリアZを使用しているためである。ACLIVDを用いるときには、この上限が250になっていることに注意する。

ACLIVDを用いるときには、A, B, EPSを倍精度実数型として指定する。

5-4 数値実験及び結果

ここでのテストに使用したサブルーチンは、ACLIVS, ACLIVDである。又、テストマトリックスとしては、なるだけ条件の悪い — 条件数の大きい — 行列が好ましいので、 $H_N$ を使用した。

表中の  $d_L$  は、

$$d_L = \max_{1 < i, j < N} |x_{ij}^{(L)}| / \max_{1 < i, j < N} |z_{ij}^{(L)}|$$

$$d_0 = 1/2$$

$$X^{(L)} = (x_{ij}^{(L)}), Z^{(L)} = (z_{ij}^{(L)}) = X^{(L)}(E - AX^{(L)})$$

である。

第0近似というのは、ACLIVSへのインプットとしての初期近似逆行列である。(今回はCLIV2Sを使用した)。第L近似は、逐次過程をL回まわった後の近似値である。(表5.1, 5.2)

表からわかるように、概ね  $L=2\sim 4$  で収束しており、実際の  $L$  の値よりも1だけ少ない所で殆んど収束解に一致している。これは、収束判定のために一度ループをまわる必要があるからである。

6章 帯状行列

6-1 方法の基礎

行列  $A$  が

$$a_{ij} = 0 \quad |i-j| \geq m \quad 1 < i, j < n$$

なるときには、 $A$  をコレスキー分解した  $L = (l_{ij})$  として、

$$l_{ij} = 0 \quad i-j > m, i < j \quad 1 < i, j < n$$

なる下三角帯状行列が存在する。この  $m$  のことを、行列  $A$  の帯巾と云うことにする。

$m < n$  のような場合、4のサブルーチンをそのまま用いると、0要素のためにストアを大量に消費し、また0を加えるという演算が何度も繰り返され、時間のロスが大きい。

従って、あらかじめ0であることがわかっている要素は計算しないようにし、さらに  $A, L$  の配列を後述のように特殊化して、時間とメモリの節約を図ったのがこのアルゴリズムである。

6-2 数値的議論

帯巾  $m$  の帯状行列  $A$  に対しては、積和を作る操作が  $m-1$  回に減るので、4章のときと同様な評価については、

$$LL^T = A + F$$

$$\|F\|_2 \leq k m^2 2^{-t} \max |a_{ij}| \leq k m^2 2^{-t} \|A\|_2$$

が成立する。また積和を倍精度で計算すると、

$$\|F\|_2 \leq k m 2^{-t} \max |a_{ij}| \leq k m 2^{-t} \|A\|_2$$

と改善される。(ここで、 $t$  は浮動小数点表示における

仮数部の桁数である。)

6-3 演算回数

コレスキー分解のためには、 $\frac{1}{2} NM(M+1)$  回の乗除と、 $\frac{1}{2} NM(M-1)$  回の加減、及び  $N$  回の開平が必要である。

次に、 $(LL^T)x = b$  を解くには、 $2N(M+1)$  回の乗除と  $2NM$  回の加減が必要である。

従って、右辺が  $NR$  個ある方程式系  $AX = B$  を解くのは、全部で

$N$  回の開平

$NM(M+1)/2 + 2NR \cdot N(M+1)$  回の乗除

$NM(M-1)/2 + 2NR \cdot NM$  回の加減

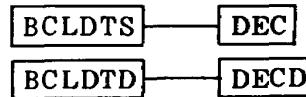
が必要になる。

6-4 サブルーチンの説明

ここには、 $LL^T$  分解を用いたサブルーチンだけしか挙げてない。

又、帯状行列の逆行列は、必ずしも帯状行列とは限らないので、逆行列を直接計算するサブルーチンは作っていない。

BCLDTS では、サブルーチン内で“DEC”をコールしている。即ち、サブルーチンの構成が



となっている。

表 5 - 1  $H_N$  : ACLIVS FORTRAN-H, OPT-0 EPS =  $2^{-48}$

N	3			4			5			6			7				
	第0近似	収束解	第1近似	第0近似	収束解	第1近似	第0近似	収束解	第1近似	第0近似	収束解	第1近似	第0近似	収束解	第1近似	第2近似	収束解
絶対誤差	$m$	1.192 <sub>-7</sub>	0.0	2.337 <sub>-6</sub>	0.0	0.0	2.185 <sub>-3</sub>	5.722 <sub>-6</sub>	0.0	5.497 <sub>-1</sub>	1.900 <sub>-3</sub>	0.0	0.0	6.901	1.289 <sub>-1</sub>	4.578 <sub>-5</sub>	0.0
	$a$	7.140 <sub>-8</sub>	0.0	7.175 <sub>-6</sub>	0.0	0.0	5.108 <sub>-3</sub>	1.511 <sub>-6</sub>	0.0	1.216 <sub>-1</sub>	4.208 <sub>-4</sub>	0.0	0.0	1.265	2.362 <sub>-2</sub>	8.164 <sub>-6</sub>	0.0
	$f$	8.615 <sub>-8</sub>	0.0	1.006 <sub>-6</sub>	0.0	0.0	8.095 <sub>-8</sub>	2.335 <sub>-6</sub>	0.0	2.080 <sub>-1</sub>	7.197 <sub>-4</sub>	0.0	0.0	2.306	4.307 <sub>-2</sub>	1.483 <sub>-5</sub>	0.0
相対誤差	$m$	7.450 <sub>-8</sub>	0.0	2.392 <sub>-6</sub>	0.0	0.0	3.378 <sub>-4</sub>	1.118 <sub>-7</sub>	0.0	3.702 <sub>-3</sub>	1.278 <sub>-6</sub>	0.0	0.0	2.192 <sub>-2</sub>	4.091 <sub>-4</sub>	1.490 <sub>-7</sub>	0.0
	$a$	5.008 <sub>-8</sub>	0.0	1.810 <sub>-6</sub>	0.0	0.0	2.849 <sub>-4</sub>	8.596 <sub>-8</sub>	0.0	3.122 <sub>-3</sub>	1.083 <sub>-6</sub>	0.0	0.0	1.781 <sub>-2</sub>	3.326 <sub>-4</sub>	1.193 <sub>-7</sub>	0.0
	$f$	5.182 <sub>-8</sub>	0.0	1.846 <sub>-6</sub>	0.0	0.0	2.892 <sub>-4</sub>	8.761 <sub>-8</sub>	0.0	3.178 <sub>-3</sub>	1.102 <sub>-6</sub>	0.0	0.0	1.804 <sub>-2</sub>	3.369 <sub>-4</sub>	1.208 <sub>-7</sub>	0.0
$\Delta L$			4.098 <sub>-8</sub>		1.512 <sub>-6</sub>	2.576 <sub>-9</sub>		3.073 <sub>-4</sub>	8.643 <sub>-8</sub>		3.456 <sub>-3</sub>	1.195 <sub>-5</sub>	1.087 <sub>-8</sub>		1.864 <sub>-2</sub>	3.481 <sub>-4</sub>	1.254 <sub>-7</sub>
$\Delta L-0$			5.000 <sub>-1</sub>		5.000 <sub>-1</sub>	7.561 <sub>-7</sub>		5.000 <sub>-1</sub>	1.536 <sub>-4</sub>		5.000 <sub>-1</sub>	1.728 <sub>-3</sub>	5.974 <sub>-6</sub>		5.000 <sub>-1</sub>	9.321 <sub>-3</sub>	1.740 <sub>-4</sub>
2×EPS			2.384 <sub>-7</sub>		2.384 <sub>-7</sub>			2.384 <sub>-7</sub>			2.384 <sub>-7</sub>		2.384 <sub>-7</sub>				2.384 <sub>-7</sub>

表 5 - 2  $H_N$  : ACLIVD FORTRAN-H, OPT-0 EPS =  $2^{-68}$

N	7			8			9			10			
	第0近似	第1近似	収束解	第0近似	第1近似	収束解	第0近似	第1近似	収束解	第0近似	第1近似	第2近似	収束解
絶対誤差	$m$	4.480 <sub>-9</sub>	0.0	4.044 <sub>-6</sub>	2.220 <sub>-16</sub>	0.0	3.739 <sub>-7</sub>	0.0	0.0	9.165 <sub>-4</sub>	5.621 <sub>-11</sub>	0.0	0.0
	$a$	8.211 <sub>-10</sub>	0.0	6.027 <sub>-7</sub>	7.144 <sub>-18</sub>	0.0	4.893 <sub>-8</sub>	0.0	0.0	1.222 <sub>-4</sub>	7.495 <sub>-12</sub>	0.0	0.0
	$f$	1.498 <sub>-9</sub>	0.0	1.167 <sub>-6</sub>	3.731 <sub>-17</sub>	0.0	1.017 <sub>-7</sub>	0.0	0.0	2.621 <sub>-4</sub>	1.608 <sub>-11</sub>	0.0	0.0
相対誤差	$m$	1.380 <sub>-11</sub>	0.0	3.982 <sub>-10</sub>	6.471 <sub>-19</sub>	0.0	4.386 <sub>-11</sub>	0.0	0.0	7.098 <sub>-8</sub>	4.348 <sub>-16</sub>	0.0	0.0
	$a$	1.139 <sub>-11</sub>	0.0	3.254 <sub>-10</sub>	6.381 <sub>-20</sub>	0.0	2.471 <sub>-11</sub>	0.0	0.0	5.776 <sub>-8</sub>	3.544 <sub>-16</sub>	0.0	0.0
	$f$	1.155 <sub>-11</sub>	0.0	3.298 <sub>-10</sub>	1.838 <sub>-19</sub>	0.0	2.735 <sub>-11</sub>	0.0	0.0	5.859 <sub>-8</sub>	3.594 <sub>-16</sub>	0.0	0.0
$\Delta L$			1.210 <sub>-11</sub>		3.429 <sub>-10</sub>	1.205 <sub>-19</sub>		3.743 <sub>-11</sub>		2.511 <sub>-19</sub>	6.130 <sub>-8</sub>	3.760 <sub>-16</sub>	1.572 <sub>-19</sub>
$\Delta L-1$			5.000 <sub>-1</sub>		5.000 <sub>-1</sub>	1.714 <sub>-10</sub>		5.000 <sub>-1</sub>		1.872 <sub>-11</sub>	5.000 <sub>-1</sub>	3.065 <sub>-8</sub>	1.880 <sub>-16</sub>
2×EPS			2.220 <sub>-16</sub>		2.220 <sub>-16</sub>		2.220 <sub>-16</sub>		2.220 <sub>-16</sub>		2.220 <sub>-16</sub>		2.220 <sub>-16</sub>

サブルーチン VI

(1) BCLDTS, BCLDTD

このサブルーチンは、2次元アレイで指定された帯状の正値対称行列  $A$  を、 $A=LL^T$  とコレスキー分解し、副産物として  $A$  の行列式の値をも与える。

行列のストアの仕方は特別な形になっている。

元の  $A$  の値は保存される。(詳細説明参照)

呼び出し形式

BCLDTS (  $N$ ,  $MBW$ ,  $A$ ,  $KA$ ,  $AL$ ,  $KAL$ ,  $DET$ ,  $MDET$ ,  $ILL$  )

インプット

$N$  : 行列  $A$  の次数,  $2 \leq N$

整数型変数名または整数定数

$MBW$  :  $6-1$  で定義した行列  $A$  の帯巾

整数型変数名または整数定数

$2 < MBW < N$

$A$  :  $N \times N$  行列  $A$  の要素

大きさ  $KA$  の 2次元アレイ

配列の仕方については図 6-1 を参照

$KA$  : アレイ  $A$  の大きさを指定,  $KA \geq N$

整数型変数名または整数定数

$KAL$  : アレイ  $AL$  の大きさを指定,  $KAL \geq N$

整数型変数名または整数定数

アウトプット

$AL$  :  $A$  をコレスキー分解して得られる下三角行列  $L$  の要素がストアされている。

図 6-2 を参照

大きさ  $KAL$  の 2次元アレイ

$DET$ ,  $MDET$

$A$  の行列式の値を、 $DET \times 2^{MDET}$  の形で与える。

$DET$  は実数型で、 $\frac{1}{2} \leq DET < 1$

$MDET$  は整数型

$ILL$  : コレスキー分解が正常に終了したときには、 $ILL=0$  がセットされている。

エラーが生じたときには、0 以外の値がセットされている。

エラー処理

$N < 1$  または  $MBW \leq 1$ ,  $N < MBW$  であれば、なんらの計算も行わない。改頁して、その旨の印字をする。

分解途中で、 $l_{ii}^2 \leq 0$  となって、コレスキー分解ができなくなったときには、 $ILL=1$  がセットされ、そこから先の計算は行わない。改頁して、“SUB BCLDTS IS FAIL” と印字し、 $l_{ii}^2 \leq 0$  となったときの  $i$  の値

も併せ印字する。このときの出力は意味のあるものではない。

詳細説明

$A$  及び  $AL$  の配列の仕方

$N=5$ ,  $MBW=3$  として例示する。

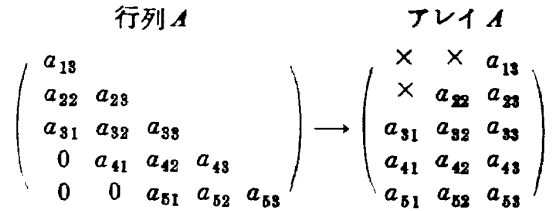


図 6-1

× の所の値は任意

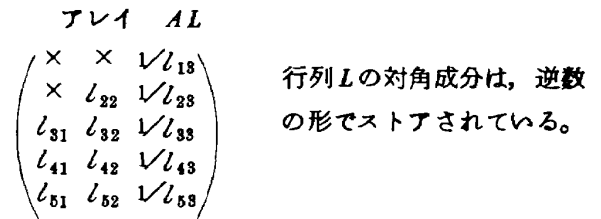


図 6-2

通常行列表示を  $A=(a_{ij})$ , アレイ  $A$  の要素を  $\bar{a}_{ij}$  とすると、

$\bar{a}_{i,m-j} = a_{i,i-j} \quad 1 \leq i \leq n, 0 \leq j \leq m-1$  という関係にある。

$A$  の値は保存できるが、仮引数  $A$  と  $AL$  の実引数を同じものにすれば、結果は  $A$  にストアされるので、 $A$  の値は失われる。

BCLDTD を用いるときには、 $A$ ,  $AL$ ,  $DET$  を倍精度実数型として指定する。

(2) BCLSL S, BCLSLD

このサブルーチンは、帯状の正値対称行列  $A$  を係数行列とする方程式  $AX=B$  を解くもので、 $A=LL^T$  と分解し (BCLDTS), ついで  $LY=B$ ,  $L^T=Y$  を解いて  $X$  を求める (BCLSL S)。

前項のサブルーチン(1)からのアウトプット  $AL$  が、本サブルーチンのインプット  $AL$  になるので、このサブルーチンを使用する前に、必ず(1)のサブルーチンを使用する。もしくは、それと同等な手続に従って、 $AL$  に図 6-2 のような情報をセットする。一度そのようにセットされれば、このサブルーチンを何度使用してもよい。

インプット  $B$  の値は保存される。

呼び出し形式

BCLSL S (  $N$ ,  $MBW$ ,  $NR$ ,  $AL$ ,  $KAL$ ,  $B$ ,  $KB$ ,



$X, KX)$

インプット

$N$ : 行列  $A$  の次数,  $N > 2$

整数型変数名または整数定数

$MBW$ : 行列  $A$  の帯巾,  $2 \leq MBW \leq N$

整数型変数名または整数定数

$NR$ : 方程式  $AX = B$  の右辺の数,  $NR > 1$

整数型変数名または整数定数

$AL$ : 係数行列  $A$  をコレスキー分解して得られる  $L$  の要素が図 6-2 のようにストアされている。

大きさ  $KAL$  の 2次元アレイ

$KAL$ : アレイ  $AL$  の大きさを指定,  $KAL > N$

整数型変数名または整数定数

$B$ : 方程式  $AX = B$  の右辺,  $N \times NR$  行列

大きさ  $KB$  の 2次元アレイ

$KB$ : アレイ  $B$  の大きさを指定,  $KB \geq N$

整数型変数名または整数定数

$KX$ : アレイ  $X$  の大きさを指定,  $KX \geq N$

整数型変数名または整数定数

アウトプット

$X$ : 方程式  $AX = B$  の解  $X$  がストアされている。

$N \times NR$  行列

大きさ  $KX$  の 2次元アレイ

エラー処理

$N \leq 1$  または  $MBW \leq 1$ ,  $N < MBW$  であれば, なんらの計算もせずに, そのままだターンする。ILL 値のセツトやメッセージの印字は行わない。

詳細説明

$B$  及び  $X$  は 2次元アレイだから, たとえ  $NR$  が 1 であっても, 2次元で指定する。

$B$  の値は保存されて, 解は  $X$  にストアされる。しかし, 仮引数  $B$  及び  $X$  の実引数を同じものにする, 解は  $B$  にストアされ,  $B$  の値は失われる。

BCLSLD を用いるときには,  $AL, B, X$  を倍精度実数型として指定する。

## 6-5 用途

用途としては, 常微分方程式や偏微分方程式を差分化して解く場合などが考えられる。この場合には,  $m \ll n$  が十分成立する。

対称帯状行列に対する RUTISHAUSER の LR アルゴリズム (固有値の計算) では, コレスキー分解が本質的な役割を果たしている。

また積和を倍精度で計算すれば, 5章と同じ逐次式

$$\begin{aligned} A &= LL^T, \quad x^{(0)} = 0 \\ r^{(s)} &= b - Ax^{(s)}, \quad (LL^T)d^{(s)} = r^{(s)} \\ x^{(s+1)} &= x^{(s)} + d^{(s)} \end{aligned}$$

に従って, 解の精密化が行える。この際, 積和を倍精度で計算する必要があるのは, 本質的には  $r^{(s)}$  の計算に対してだけである。

## 6-6 数値実験及び結果

各サブルーチンの比較は, 逆行列を計算することによって種々の数値を比べている。一方, ここのサブルーチンには, 逆行列を直接計算するものがないので,  $AX = E$  を解いて逆行列を求めた。従って, 逆行列要素のすべてが計算されるので, 最大値  $m$ , 算術平均  $a$ , 2乗平均  $f$  の 3つの値も, これら  $N^2$  の全要素にわたって計算してある。

なお, 表 6-1~6-3 の「時間」の欄は, BCLSLS に要する時間である。BCLDTS に要する時間は, 今回の実験範囲では, 高々 10msec のオーダーである。

6-2 の演算回数からわかるように, ( $M=2$  or  $3$ ,  $NR=N$ ) BCLDTS の主要項は  $N$ , BCLSLS のそれは  $N^3$  である。従って, その消費時間も  $N$ , もしくは  $N^3$  に依存することが期待されるが, そのことは表 6-4 によく現われている。

$B_N^2$  を用いて, CLDT2S・CLSL2S との消費時間の比較を表 6-4 に挙げておく。

表 6-1  $B_N$ : BCLDTS, BCLSLs

FORTRAN-D

		$N$	10	35	50	100	200	300
絶対誤差	$m$		1.788 <sub>-7</sub>	5.722 <sub>-6</sub>	1.597 <sub>-5</sub>	9.251 <sub>-5</sub>	1.336 <sub>-3</sub>	4.459 <sub>-3</sub>
	$a$		4.595 <sub>-8</sub>	2.420 <sub>-6</sub>	5.315 <sub>-6</sub>	3.872 <sub>-5</sub>	5.001 <sub>-4</sub>	1.845 <sub>-3</sub>
	$f$		6.093 <sub>-8</sub>	2.866 <sub>-6</sub>	6.653 <sub>-6</sub>	4.629 <sub>-5</sub>	6.232 <sub>-4</sub>	2.249 <sub>-3</sub>
相対誤差	$m$		1.043 <sub>-7</sub>	1.431 <sub>-6</sub>	2.027 <sub>-6</sub>	8.687 <sub>-6</sub>	4.712 <sub>-5</sub>	1.269 <sub>-4</sub>
	$a$		4.441 <sub>-8</sub>	8.635 <sub>-7</sub>	1.303 <sub>-6</sub>	5.237 <sub>-6</sub>	3.233 <sub>-5</sub>	8.238 <sub>-5</sub>
	$f$		5.222 <sub>-8</sub>	9.068 <sub>-7</sub>	1.387 <sub>-6</sub>	5.521 <sub>-6</sub>	3.400 <sub>-5</sub>	8.655 <sub>-5</sub>
時間	sec		—	—	0.1	0.3	1.0	2.2
$\log K - \log \lambda$			6.142	5.107	4.804	4.210	3.613	3.262
$-\log m$			6.982	5.845	5.693	5.061	4.327	3.897
$-\log f$			7.282	6.042	5.858	5.258	4.469	4.063
HITAC-5020	sec		4.783 <sub>-6</sub> ×	2.730 <sub>-5</sub> ×	4.155 <sub>-5</sub> ×	2.446 <sub>-4</sub> 3	1.036 <sub>-3</sub> 9	2.207 <sub>-3</sub>

表 6-2  $B_N^2$ : BCLDTS, BCLSLs

FORTRAN-D

		$N$	10	17	40	50	80	100
絶対誤差	$m$		1.945 <sub>-4</sub>	9.487 <sub>-3</sub>	1.169	4.856	1.729 <sub>+2</sub>	9.072 <sub>+2</sub>
	$a$		8.915 <sub>-5</sub>	4.223 <sub>-3</sub>	4.977 <sub>-1</sub>	2.096	7.182 <sub>+1</sub>	3.735 <sub>+2</sub>
	$f$		1.027 <sub>-4</sub>	4.967 <sub>-3</sub>	5.994 <sub>-1</sub>	2.522	8.750 <sub>+1</sub>	4.567 <sub>+2</sub>
相対誤差	$m$		8.091 <sub>-6</sub>	9.067 <sub>-5</sub>	1.002 <sub>-3</sub>	2.263 <sub>-3</sub>	1.910 <sub>-2</sub>	5.101 <sub>-2</sub>
	$a$		6.641 <sub>-6</sub>	7.737 <sub>-5</sub>	8.316 <sub>-4</sub>	1.850 <sub>-3</sub>	1.591 <sub>-2</sub>	4.278 <sub>-2</sub>
	$f$		6.768 <sub>-6</sub>	7.791 <sub>-5</sub>	8.396 <sub>-4</sub>	1.863 <sub>-3</sub>	1.602 <sub>-2</sub>	4.309 <sub>-2</sub>
時間	sec		—	—	—	0.1	0.2	0.3
$\log K - \log \lambda$			4.458	3.595	2.161	1.782	9.775 <sub>-1</sub>	5.941 <sub>-1</sub>
$-\log m$			5.092	4.043	2.999	2.645	1.719	1.292
$-\log f$			5.170	4.108	3.076	2.730	1.795	1.366
HITAC-5020	sec		5.386 <sub>-5</sub> ×	1.492 <sub>-4</sub> ×	5.298 <sub>-3</sub> ×	1.733 <sub>-2</sub> 1	7.086 <sub>-2</sub> 1.5	1.947 <sub>-1</sub> 2.5

表 6-3  $B_N^2$ : BCLDTD, BCLSLD

FORTRANH, OPT-0

		$N$	50	80	100	150	200
絶対誤差	$m$		3.163 <sub>-11</sub>	9.276 <sub>-10</sub>	4.712 <sub>-9</sub>	5.057 <sub>-8</sub>	4.113 <sub>-7</sub>
	$a$		1.279 <sub>-11</sub>	3.910 <sub>-10</sub>	1.925 <sub>-9</sub>	2.080 <sub>-8</sub>	1.709 <sub>-7</sub>
	$f$		1.564 <sub>-11</sub>	4.756 <sub>-10</sub>	2.358 <sub>-9</sub>	2.545 <sub>-8</sub>	2.092 <sub>-7</sub>
相対誤差	$m$		1.252 <sub>-14</sub>	1.040 <sub>-13</sub>	2.601 <sub>-13</sub>	8.738 <sub>-13</sub>	3.038 <sub>-12</sub>
	$a$		1.096 <sub>-14</sub>	8.675 <sub>-14</sub>	2.198 <sub>-13</sub>	7.218 <sub>-13</sub>	2.521 <sub>-12</sub>
	$f$		1.105 <sub>-14</sub>	8.740 <sub>-14</sub>	2.214 <sub>-13</sub>	7.272 <sub>-13</sub>	2.538 <sub>-12</sub>
時間			0.1	0.4	0.6	1.3	2.3
$\log K - \log \lambda$			12.32	11.51	11.13	10.43	9.935
$-\log m$			13.90	12.98	12.58	12.06	11.52
$-\log f$			13.96	13.06	12.65	12.14	11.60

表 6-4

単位 msec

		N					
		10	17	40	50	80	100
CLDT2S	H-0	1	5	62	121	486	935
	H-2	1	2	19	34	133	258
CLSL2S	H-0	6	28	350	682	2784	5448
	H-2	2	10	98	186	723	1416
BCLDTS	H-0	—	1	2	3	5	7
	H-2	—	1	2	3	5	6
BCLSLS	H-0	4	13	77	121	312	488
	H-2	3	11	64	100	258	406

## 7章 ガウス・ジョルダン法

### 7-1 方法の基礎

方程式系を

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n & (7.1.1) \\ y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n & (7.1.2) \\ \vdots \\ y_n = a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n & (7.1.n) \end{cases}$$

とすると、 $a_{11} \neq 0$ であれば、(7.1.1)式より、 $x_1$ を  
 $x_1 = a_{11}^{(1)}y_1 + a_{12}^{(1)}x_2 + \dots + a_{1n}^{(1)}x_n$   
 と求め、この $x_1$ を逐次方程式(7.1.2)～(7.1.n)に代入して、 $x_1$ が消去できる。

ついで、(7.1.2)から $x_2$ を求め、その $x_2$ を残りの方程式に代入すると、右辺から $x_2$ が消去できる。

順次この手続を進めていくと、第  $n$  回目の手続が終わったときには、 $x_1 \sim x_n$  が  $y_1 \sim y_n$  を用いて表わされている。このとき、その係数行列が、始めの方程式系(7.1.1)～(7.1.n)の係数行列の逆行列となっている。

この手続を進めていく際に、第  $k$  ステップ目で、 $a_{kk}^{(k-1)}$  で割算を行う必要がある。この  $a_{kk}^{(k-1)}$  は、 $A$  の  $k$  次主小行列式の値  $A_k$  を用いて、

$$a_{kk}^{(k-1)} = A_k / A_{k-1}$$

ただし  $A_0 = 1, a_{11}^{(0)} = a_{11}$

と表わされる。(ガウス消去法の場合と同じである) 従って、 $A$  の主小行列式がすべて0でなければ、この手続は、最後まで行える。

今回は、行列  $A$  には、正定値行列という制限をつけることにした。そうすると、その主小行列式はすべて正になるので、上式の  $a_{kk}^{(k-1)}$  もまたすべて正である。このことによって、上記手順が最後まで行えることがわかる。また、 $a_{kk}^{(k-1)}$  — ピボットと云う — の符号を調べることによって、エラーの検出ができる。(途中でピボットが負または零になれば、それは、まるめ誤差等の影響で、

行列が正定値でなくなったことを意味する)

ガウス・ジョルダン法自体は、その対象とする行列が必ずしも正値対称行列であることを必要としない。ここではただこれを正値対称行列用に特殊化しているだけである。

さらに、正定値行列  $A$  にガウス・ジョルダン法を適用して、その第  $k$  ステップが終わったあとの行列を、

$$\begin{pmatrix} A_k - C_k^T \\ C_k & B_k \end{pmatrix}$$

とブロック分割すると、 $B_k$  も正定値になり、第  $k+1$  ステップは、この  $B_k$  の (1, 1) 要素で割り算することから始まる。一方、「正定値行列では、その絶対値が最大の要素は対角成分中にあり、当然その要素は正である。」このことを利用すると、トータルピボットングも行える。

### 7-2 演算回数

$N^2(N+1)/2$  回の乗除と、 $N^2(N-1)/2$  回の加減が必要である。コレスキー法と比較して、 $N$  回の開平演算部分だけ減少している。

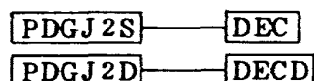
### 7-3 サブルーチンの説明

ここに掲げたサブルーチンは、専ら逆行列自身を求めるためのものである。

プログラミング技術上、1ステップごとに変数を、 $x_1 \rightarrow x_n, x_i \rightarrow x_{i-1} \quad 2 \leq i \leq n$  に代えている。そうすると、毎回消去されるのは  $x_1$  であり、 $N$  回のステップが終了すると、変数は元通りになっている。

サブルーチン名の第5文字目の「1」或は「2」は、それぞれ  $A$  が1次元アレイ、或は2次元アレイとして扱われていることを示す。又第6文字目の「S」は単精度を、「D」は倍精度を意味する。

この2つのサブルーチンは、共に行列式の値をも求めているので、サブルーチンの中で“DEC”(倍精度のときは“DECD”)をコールしている。構成は下図の通り。



サブルーチン VI

(1) PDGJ1S, PDGJ1D

このサブルーチンは、1次元アレイで指定された正値対称行列に、ガウス・ジョルダン法を適用して、 $A$ の逆行列を求めるものである。副産物として、 $A$ の行列式の値をも与える。

$A$ の元の値は失われる。

呼び出し形式

PDGJ1S (  $N$ ,  $A$ ,  $DET$ ,  $MDET$ ,  $ILL$  )

インプット

$N$ : 行列 $A$ の次数  $2 \leq N < 700$

整数型変数名または整数

$A$ :  $N \times N$  行列 $A$ の要素

大きさは  $N(N+1)/2$  の1次元アレイ

配列の仕方については、図7-1を参照

アウトプット

$A$ :  $A^{-1}$ の要素がストアされている。

図7-1を参照

$DET$ ,  $MDET$

$A$ の行列式の値を、 $DET \times 2^{MDET}$ の形で与える。

$DET$ は実数型で、 $1/2 \leq DET < 1$

$MDET$ は整数型

$ILL$ : 計算が正常に終了したときには、 $ILL=0$ がセットされている。

エラーが生じたときには、0以外の値がセットされている。

エラー処理

$N < 1$  または  $700 < N$  であれば、 $ILL=10$ をセットして、なんらの計算も行わない。改頁して、その旨の印字をする。

計算途中で、ピボットが負または零になって、計算が意味をもたなくなったときには、 $ILL=1$ がセットされ、そこから先の計算は行わない。改頁して、“SUB PDGJ1S IS FAIL”と印字し、ピボットが負または零になったときの $I$ の値も併せ印字する。このときの出力は意味のあるものではない。

詳細説明

インプット $A$ の配列

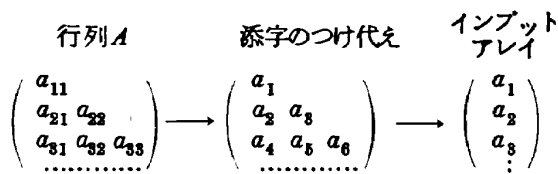


図7-1

一般に、行列 $A$ の第 $(I, J)$ -要素は、1次元アレイ $A$ の第 $I * (I-1) / 2 + J$ 番目の要素となっている。

アウトプットの逆行列も同じ配列の仕方になっている。

行列の次数 $N$ に上限があるのは、サブルーチン内で1次元のワーキングエリア $H$ を使用しているためである。PDGJ1Dを用いるときには、この上限が500になっていることに注意する。

PDGJ1Dを用いるときには、 $A$ ,  $DET$ を倍精度実数型として指定する。

(2) PDGJ2S, PDGJ2D

このサブルーチンは、2次元アレイで指定された正値対称行列 $A$ に、ガウス・ジョルダン法を適用して、 $A$ の逆行列を求めるものである。副産物として、 $A$ の行列式の値をも与える。

元の $A$ の値は保存される。

呼び出し形式

PDGJ2S (  $N$ ,  $A$ ,  $KA$ ,  $DIA$ ,  $DET$ ,  $MDET$ ,  $ILL$  )

インプット

$N$ : 行列 $A$ の次数  $2 < N < 500$

整数型変数名または整数

$A$ :  $N \times N$  行列 $A$ の要素

配列の仕方は通常の行列表示と同じ。

図7-2を参照

大きさ $KA$ の2次元アレイ

$KA$ : アレイ $A$ の大きさを指定、 $KA > N$

整数型変数名または整数

アウトプット

$A$ :  $A^{-1}$ の要素がストアされている。

図7-2を参照

$DIA$ : インプット $A$ の対角要素がストアされている。

図7-2を参照

大きさ $N$ の1次元アレイ

$DET$ ,  $MDET$

$A$ の行列式の値を、 $DET \times 2^{MDET}$ の形で与える。

$DET$ は実数型で、 $1/2 \leq DET < 1$

$MDET$ は整数型

$ILL$ : 計算が正常に終了したときには、 $ILL=0$ がセットされている。

エラーが生じたときには、0以外の値がセット

されている。

エラー処理

$N < 1$  または  $500 < N$  であれば,  $ILL = 10$  をセットして, なんらの計算も行わない。改頁して, その旨の印字をする。

計算途中で, ピボットが負または零になって, 計算が意味をもたなくなったときには,  $ILL = 1$  がセットされ, そこから先の計算は行わない。改頁して, “SUB PDG J2S IS FAIL” と印字し, ピボットが負または零になったときの  $I$  の値も併せ印字する。このときの出力は, 意味のあるものではない。

詳細説明

インプット, アウトプットの  $A$  及び  $DIA$

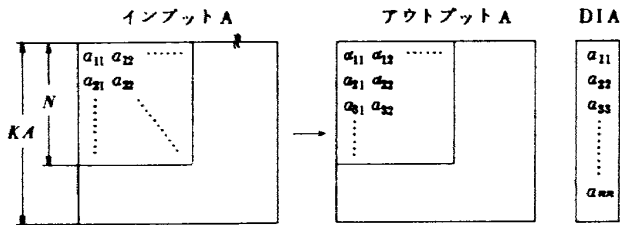


図 7-2

CLIV2S や SMIV2S のアウトプットとは配列の仕方が異なることに注意する。

行列の次数  $N$  に上限があるのは, サブルーチン内で 1 次元のワーキングエリア  $H$  を使用しているためである。PDGJ2D を用いるときには, この上限が 350 になっていることに注意する。

PDGJ2D を用いるときには,  $A$ ,  $DIA$ ,  $DET$  を倍精度実数型として指定する。

7-4 数値実験及び結果

数値実験

第 4 章との違いは, CLIV2S を用いる代わりに PDGJ2S を用いることだけで, その他の違いはない。「時間」も, PDGJ2S にだけ要する時間である。(表 7.1 ~ 表 7.7)

表 7 と表 4 とを比較すると, ガウス・ジョルダン法もコレスキー法も, 精度ではさほどの差はみられない。一方消費時間については, 行列を 2 次元アレイとして扱った場合には, ガウス・ジョルダン法の方が少ないが, 1 次元アレイとして扱った場合には, 逆にコレスキー法の方が早くなっている ( $P_N^{(2)}$ ,  $P_N^{(N+1)}$ )。これは, これらの行列については, 230/75 FORTRAN-D でテストしたため, オプティマイゼーションの効果の現われ方がでたためと思われる。これは, FORTRAN-H OPT-0 (オプティマイゼーションなし) でテストすると, 結果は逆になる。

H-0 と H-2 による消費時間の違いを,  $P_N^{(3)}$  を用いてテストした結果を表 7-8 に掲げておく。

8 章 バイマトリックスに関する若干の考察

第 3 章で述べたように, バイマトリックス  $P_N^d$  は,

$$(P_N^d)_{ij} = \begin{cases} d & i = j \\ 1 & i \neq j \end{cases} \quad (8.1)$$

と与えられる。これは又,  $d = c + 1$  とすると

$$(P_N^d)_{ij} = 1 + c \delta_{ij} \quad (8.1')$$

$\delta_{ij}$  はクロネッカーの  $\delta$

とも表わされる。

$P_N^d$  の固有値の 1 つは  $c$  であることが容易にわかる。さらに,  $P_N^d - cI$  ( $I$  は単位行列) のランクは  $N-1$  だから,  $c$  は  $N-1$  重の固有値である。残る 1 つの固有値は, trace との関係から  $c + N$  であることがわかる。以上より, 条件数・行列式の値はそれぞれ

$$\kappa(P_N^d) = 1 + N/c \quad (3.2)$$

$$\det(P_N^d) = c^{N-1} (c + N) \quad (8.3)$$

となる。又,  $d > 1$  即ち  $c > 0$  であれば  $P_N^d$  は正定値行列となる。一方  $P_N^d$  の逆行列を  $Q_N^d$  とすれば

$$(Q_N^d)_{ij} = \begin{cases} (c + N - 1)/c(c + N) & i = j \\ -1/c(c + N) & i \neq j \end{cases} \quad (8.4)$$

となる。

次に,  $P_N^d$  に対して, ガウス消去法の第  $k$  ステップ目が終わった行列を  $A_k$ , ガウス・ジョルダン法の第  $k$  ステップ目が終わった行列を  $B_k$  とし, 次のようにブロック分割する。

表 7-1  $H_N$ ; PDGJ2S

FORTRAN-H, OPT-0

		$N$	3	4	5	6	7
絶対誤差	$m$		7.153 <sub>-7</sub>	9.155 <sub>-5</sub>	9.718 <sub>-3</sub>	6.045 <sub>-1</sub>	1.461 <sub>+1</sub>
	$a$		4.005 <sub>-7</sub>	2.752 <sub>-5</sub>	2.313 <sub>-3</sub>	1.343 <sub>-1</sub>	2.681
	$f$		5.105 <sub>-7</sub>	3.969 <sub>-5</sub>	3.641 <sub>-3</sub>	2.295 <sub>-1</sub>	4.895
相対誤差	$m$		2.384 <sub>-7</sub>	6.370 <sub>-6</sub>	1.686 <sub>-4</sub>	4.103 <sub>-3</sub>	4.436 <sub>-2</sub>
	$a$		2.103 <sub>-7</sub>	5.521 <sub>-6</sub>	1.376 <sub>-4</sub>	3.474 <sub>-3</sub>	3.684 <sub>-2</sub>
	$f$		2.142 <sub>-7</sub>	5.596 <sub>-6</sub>	1.389 <sub>-4</sub>	3.532 <sub>-3</sub>	3.739 <sub>-2</sub>
log $K - \log \lambda$			5.107	3.636	2.149	6.521 <sub>-1</sub>	-8.502 <sub>-1</sub>
- log $m$			6.623	5.196	3.773	2.387	1.353
- log $f$			6.669	5.252	3.857	2.452	1.427
HITAC-5020			1.907 <sub>-6</sub>	4.745 <sub>-5</sub>	1.322 <sub>-3</sub>	1.213 <sub>-2</sub>	4.646 <sub>-1</sub>

表 7-2  $H_N$ ; PDGJ2D

FORTRAN-H, OPT-0

		$N$	5	6	7	8	9
絶対誤差	$m$		1.851 <sub>-13</sub>	7.982 <sub>-12</sub>	6.197 <sub>-10</sub>	7.916 <sub>-7</sub>	8.958 <sub>-6</sub>
	$a$		4.437 <sub>-14</sub>	1.760 <sub>-12</sub>	1.135 <sub>-10</sub>	1.178 <sub>-7</sub>	1.238 <sub>-6</sub>
	$f$		6.966 <sub>-14</sub>	3.012 <sub>-12</sub>	2.070 <sub>-10</sub>	2.284 <sub>-7</sub>	2.531 <sub>-6</sub>
相対誤差	$m$		3.370 <sub>-15</sub>	5.240 <sub>-14</sub>	1.935 <sub>-12</sub>	7.581 <sub>-11</sub>	1.004 <sub>-9</sub>
	$a$		2.705 <sub>-15</sub>	4.484 <sub>-14</sub>	1.587 <sub>-12</sub>	6.262 <sub>-11</sub>	8.335 <sub>-10</sub>
	$f$		2.730 <sub>-15</sub>	4.568 <sub>-14</sub>	1.609 <sub>-12</sub>	6.355 <sub>-11</sub>	8.461 <sub>-10</sub>
log $K - \log \lambda$			12.68	11.19	9.686	8.179	6.670
- log $m$			14.47	13.28	11.71	10.12	8.998
- log $f$			14.56	13.34	11.79	10.20	9.073
HITAC-5020			9.684 <sub>-14</sub>	6.120 <sub>-12</sub>	3.929 <sub>-10</sub>	5.254 <sub>-9</sub>	2.613 <sub>-7</sub>

表 7-3  $P_N^{(2)}$ ; PDGJ2S

FORTRAN-D

		$N$	10	50	100	200	300	400
絶対誤差	$m$		2.980 <sub>-8</sub>	4.470 <sub>-8</sub>	5.960 <sub>-8</sub>	2.086 <sub>-7</sub>	2.086 <sub>-7</sub>	2.384 <sub>-7</sub>
	$a$		4.369 <sub>-9</sub>	2.026 <sub>-9</sub>	1.184 <sub>-9</sub>	1.256 <sub>-9</sub>	9.745 <sub>-10</sub>	7.164 <sub>-10</sub>
	$f$		8.181 <sub>-9</sub>	5.659 <sub>-9</sub>	3.832 <sub>-9</sub>	9.303 <sub>-9</sub>	9.274 <sub>-9</sub>	8.513 <sub>-9</sub>
相対誤差	$m$		4.470 <sub>-8</sub>	3.576 <sub>-7</sub>	6.855 <sub>-7</sub>	1.490 <sub>-6</sub>	2.325 <sub>-6</sub>	2.980 <sub>-6</sub>
	$a$		1.707 <sub>-8</sub>	5.929 <sub>-8</sub>	7.716 <sub>-8</sub>	1.112 <sub>-7</sub>	9.420 <sub>-8</sub>	9.927 <sub>-8</sub>
	$f$		2.098 <sub>-8</sub>	6.777 <sub>-8</sub>	8.320 <sub>-8</sub>	1.342 <sub>-7</sub>	1.218 <sub>-7</sub>	1.225 <sub>-7</sub>
時間 sec			-	0.2 ×	1.2 2.3	9.3 18.2	31.0 61.2	× 145.2
log $K - \log \lambda$			6.785	6.119	5.822	5.524	5.348	5.224
- log $m$			7.350	6.447	6.164	5.827	5.634	5.526
- log $f$			7.678	7.169	7.080	6.872	6.914	6.912
HITAC-5020			6.855 <sub>-7</sub>	3.740 <sub>-6</sub>	8.356 <sub>-6</sub>	1.608 <sub>-5</sub>		
sec			-	×	8 16	65 126		

表 7-4  $P_N^{(N+1)}$ ; PDGJ2S

FORTRAN-D

		$N$	10	50	100	200	300	400
絶対誤差	$m$		1.863 <sub>-9</sub>	9.313 <sub>-10</sub>	9.313 <sub>-10</sub>	8.149 <sub>-10</sub>	5.821 <sub>-10</sub>	1.746 <sub>-10</sub>
	$a$		1.355 <sub>-10</sub>	2.777 <sub>-11</sub>	9.333 <sub>-12</sub>	4.728 <sub>-12</sub>	2.905 <sub>-12</sub>	7.225 <sub>-13</sub>
	$f$		4.406 <sub>-10</sub>	1.013 <sub>-10</sub>	5.468 <sub>-11</sub>	4.198 <sub>-11</sub>	3.245 <sub>-11</sub>	6.547 <sub>-12</sub>
相対誤差	$m$		5.960 <sub>-8</sub>	2.086 <sub>-7</sub>	2.086 <sub>-7</sub>	2.980 <sub>-7</sub>	3.129 <sub>-7</sub>	3.278 <sub>-7</sub>
	$a$		9.483 <sub>-9</sub>	5.655 <sub>-8</sub>	6.754 <sub>-8</sub>	7.034 <sub>-8</sub>	8.542 <sub>-8</sub>	1.034 <sub>-7</sub>
	$f$		1.831 <sub>-8</sub>	7.222 <sub>-8</sub>	7.875 <sub>-8</sub>	9.158 <sub>-8</sub>	1.066 <sub>-7</sub>	1.250 <sub>-7</sub>
時間 sec			—	0.2 ×	1.2, 2.3	9.3, 18.2	31.0, 61.2	× 145.2
$\log K - \log x$			7.526	7.526	7.526	7.526	7.526	7.526
- $\log m$			7.225	6.681	6.681	6.526	6.505	6.484
- $\log f$			7.737	7.141	7.104	7.038	6.972	6.903

表 7-5  $B_N$ ; PDGJ2S

FORTRAN-D

		$N$	10	35	50	100	200	300
絶対誤差	$m$		2.980 <sub>-7</sub>	1.001 <sub>-5</sub>	2.885 <sub>-5</sub>	1.841 <sub>-4</sub>	1.385 <sub>-3</sub>	4.364 <sub>-3</sub>
	$a$		9.462 <sub>-8</sub>	4.107 <sub>-6</sub>	1.174 <sub>-5</sub>	7.604 <sub>-5</sub>	5.479 <sub>-4</sub>	1.791 <sub>-3</sub>
	$f$		1.129 <sub>-7</sub>	4.951 <sub>-6</sub>	1.428 <sub>-5</sub>	9.291 <sub>-5</sub>	6.740 <sub>-4</sub>	2.198 <sub>-3</sub>
相対誤差	$m$		1.341 <sub>-7</sub>	2.146 <sub>-6</sub>	4.277 <sub>-6</sub>	1.474 <sub>-5</sub>	5.411 <sub>-5</sub>	1.195 <sub>-4</sub>
	$a$		8.480 <sub>-8</sub>	1.401 <sub>-6</sub>	2.883 <sub>-6</sub>	9.767 <sub>-6</sub>	3.584 <sub>-5</sub>	7.884 <sub>-5</sub>
	$f$		9.104 <sub>-8</sub>	1.472 <sub>-6</sub>	3.028 <sub>-6</sub>	1.025 <sub>-5</sub>	3.761 <sub>-5</sub>	8.268 <sub>-5</sub>
時間 sec			—	—	0.1	1.1	8.7	29.2
$\log K - \log x$			6.142	5.107	4.804	4.210	3.613	3.262
- $\log m$			6.873	5.668	5.369	4.832	4.267	3.923
- $\log f$			7.041	5.832	5.519	4.989	4.425	4.083

表 7-6  $B_N^2$ ; PDGJ2S

FORTRAN-D

		$N$	10	17	40	50	80	100
絶対誤差	$m$		5.865 <sub>-5</sub>	3.366 <sub>-3</sub>	1.515	6.622	1.424 <sub>+2</sub>	7.339 <sub>+2</sub>
	$a$		2.771 <sub>-5</sub>	1.556 <sub>-3</sub>	6.432 <sub>-1</sub>	2.793	5.988 <sub>+1</sub>	3.060 <sub>+2</sub>
	$f$		3.196 <sub>-5</sub>	1.824 <sub>-3</sub>	7.770 <sub>-1</sub>	3.385	7.284 <sub>+1</sub>	3.733 <sub>+2</sub>
相対誤差	$m$		2.369 <sub>-6</sub>	3.418 <sub>-5</sub>	1.258 <sub>-3</sub>	2.883 <sub>-3</sub>	1.596 <sub>-2</sub>	4.213 <sub>-2</sub>
	$a$		2.006 <sub>-6</sub>	2.829 <sub>-5</sub>	1.058 <sub>-3</sub>	2.418 <sub>-3</sub>	1.323 <sub>-2</sub>	3.504 <sub>-2</sub>
	$f$		2.029 <sub>-6</sub>	2.853 <sub>-5</sub>	1.065 <sub>-3</sub>	2.435 <sub>-3</sub>	1.332 <sub>-2</sub>	3.528 <sub>-2</sub>
時間 sec			—	—	—	0.1	0.6	1.1
$\log K - \log x$			4.458	3.595	2.161	1.782	9.775 <sub>-1</sub>	5.941 <sub>-1</sub>
- $\log m$			5.625	4.466	2.900	2.540	1.797	1.375
- $\log f$			5.693	4.545	2.973	2.614	1.875	1.452
HITAC-5020 sec			6.481 <sub>-5</sub>	4.512 <sub>-4</sub>	1.127 <sub>-2</sub>	2.644 <sub>-2</sub>	1.521 <sub>-1</sub>	2.989 <sub>-1</sub>
			—	—	—	1	5	9

表 7-7  $B_N^2$ ; PDGJ 2D

FORTTRAN-H, OPT-0

		$N$	50	80	100	150	200
絶対誤差	$m$		1.868 <sub>-10</sub>	5.025 <sub>-9</sub>	2.568 <sub>-8</sub>	5.026 <sub>-7</sub>	3.817 <sub>-6</sub>
	$a$		7.961 <sub>-11</sub>	2.093 <sub>-9</sub>	1.066 <sub>-8</sub>	2.060 <sub>-7</sub>	1.560 <sub>-6</sub>
	$f$		9.623 <sub>-11</sub>	2.551 <sub>-9</sub>	1.302 <sub>-8</sub>	2.527 <sub>-7</sub>	1.916 <sub>-6</sub>
相対誤差	$m$		8.334 <sub>-14</sub>	5.526 <sub>-13</sub>	1.465 <sub>-12</sub>	8.453 <sub>-12</sub>	2.728 <sub>-11</sub>
	$a$		6.922 <sub>-14</sub>	4.610 <sub>-13</sub>	1.220 <sub>-12</sub>	7.085 <sub>-12</sub>	2.285 <sub>-11</sub>
	$f$		6.969 <sub>-14</sub>	4.642 <sub>-13</sub>	1.228 <sub>-12</sub>	7.136 <sub>-12</sub>	2.301 <sub>-11</sub>
時間	sec		0.5	2.1	4.1	13.9	32.8
$\log K - \log \lambda$			12.32	11.51	11.13	10.43	9.935
$-\log m$			13.08	12.26	11.83	11.07	10.56
$-\log f$			13.16 <sub>1</sub>	12.33	11.91	11.15	10.64
HITAC-5020	sec		1.960 <sub>-11</sub> 1	1.458 <sub>-10</sub> 5	3.582 <sub>-10</sub> 9	1.792 <sub>-9</sub> 32	

表 7-8

単位 sec

		$N$	100	200	300	400
CLIV 2S	H-0		3.0	25.0	86.3	
	H-2		0.9	7.4	26.6	
CLIV 1S	H-0		2.5	21.4	75.5	184.6
	H-2		1.8	15.2	53.6	131.7
PDGJ 2S	H-0		3.4	27.0	90.5	
	H-2		1.0	7.7	25.5	
PDGJ 1S	H-0		2.3	17.9	60.0	141.3
	H-2		1.0	7.6	25.3	59.5

$$A_k = \begin{pmatrix} \widehat{A}_{11}^k & \widehat{A}_{12}^{N-k} \\ 0 & \widehat{A}_{22}^k \end{pmatrix} \quad B_k = \begin{pmatrix} B_{11} & -B_{21}^T \\ B_{21} & B_{22} \end{pmatrix}$$

$$(P_N^d)^{-1} = \begin{pmatrix} C_{11} & C_{21}^T \\ C_{21} & C_{22} \end{pmatrix}$$

そうすると、 $A_{22} = B_{22} = C_{22}^{-1}$ であることが容易に確かめられる。このことから、(8.4)を用いて

$$(A_{22})_{ij} = (B_{22})_{ij} = \begin{cases} c(c+k+1)/(c+k) & i=j \\ c/(c+k) & i \neq j \end{cases}$$

となる。 $c(c+k+1)/(c+k) > c/(c+k)$ だから、 $P_N^d$ の場合には、ガウス消去法もガウス・ジョルダン法もピボットティングをしなくとも、結果的にはトータルピボットティングをしたのと同じことになっている。

### 9章 結 語

本報告は、サブルーチンライブラリーに関する系統的

研究の最初のものであり、以後さらに拡充強化していく予定である。又、その性格上、常に手法の開発、プログラムの手直し等がなされねばならず、持続的研究が要求されるものである。

なお、この後一般の正方行列を対象とする、「1次方程式系の解法Ⅱ」、さらに overdetermined system を対象とする、「1次方程式系の解法Ⅲ」を順次刊行し、引き続き行列の固有値問題に進む予定である。

1次方程式系全体の見出しは、「1次方程式系の解法Ⅲ」<sup>(7)</sup>を参照されたい。

最後に、FORTRANプログラムの書換えの労をとってくれた富士通ファコムの中川信行氏に感謝の意を表す。

### 参 考 文 献

- 1) J. H. WILKINSON, C. REINISCH: Handbook for Automatic Computation Vol. II (1971) Springer-Verlag Berlin Heidelberg New York



- 2) M. NEWMAN, J. TODD: The evaluation of matrix inversion programs, J. Soc. Indust. Appl. Math. Vol. 6 (1958) pp. 466-476
- 3) M. H. LIETZKE, W. STOUGHTON, M. P. LIETZKE: A comparison of several methods for inverting large symmetric positive definite matrices, Math. Comp. Vol. 18 (1964) pp. 449-456
- 4) J. WALSH edited: Numerical Analysis: An introduction (1966) Academic Press  
高須達 監訳: 数値解析概論 (1970) 日本評論社
- 5) J. H. WILKINSON: The algebraic eigenvalue problem (1965) Clarendon Press, Oxford
- 6) J. H. WILKINSON: Error Analysis of Direct Methods of Matrix Inversion, J. Assoc. Comput. Mach. Vol. 8 (1961) pp. 281.
- 7) 木村武雄他: 1次方程式系の解法Ⅲ, 航技研報告, 近刊予定
- 8) 板垣芳雄: ベクトルのノルムと行列のノルム — 数値解法の収束条件その他への応用 —, 航技研資料 TM-53 (1965. 5)
- 9) ファージェーエフ, ファージェーエバ; 小国力訳 線型代数の計算法(上) (1970), 産業図書

---

## 航空宇宙技術研究所資料 277号

昭和50年5月発行

発行所 航空宇宙技術研究所  
東京都調布市深大寺町1880  
電話武蔵野三鷹(0422)47-5911(大代表)〒182

印刷所 株式会社 共 進  
東京都杉並区久我山4-1-7(羽田ビル)

---

