AN ALGORITHM FOR FITTING
AN ARCHIMEDEAN SPIRAL TO EMPIRICAL DATA

SK Mishra
Dept. of Economics
NEHU, Shillong
India

1. Introduction : Let there be a set of (empirically obtained) points Z= $(z_1, z_2,\ldots,z_n)$: $n \geq 3$ and any $z_i = (x_i, y_i)$. Let an inspection of the pattern that these points suggest or a conjecture regarding the law governing the generation of these points indicate that they resemble the trace (Buck, pp. 313-318) of a spiral. Then there may arise a need to investigate the law generating such a spiral or, to begin with, fit a spiral to the empirical data.

In the Cartesian coordinate system a spiral (of Archimedes) is described by two parametric equations, viz.

$$x_i = r_i \cos(\theta_i + 360k) = r_i \cos(\theta_i)$$
$$y_i = r_i \sin(\theta_i + 360k) = r_i \sin(\theta_i) \qquad \ldots (1)$$

where, i =1, 2,…,n; $0^0 \leq \theta_i < 360^0$; $r_i = (x_i^2 + y_i^2)^{1/2}$; k is a non-negative integer; θ_i = $\arctan(y_i/x_i)$ for $x_i \neq 0$, otherwise $\theta_i = 90^0$ for $(x_i, y_i) = (0, > 0)$ and 270^0 for $(x_i, y_i) = (0, < 0)$, while for $(x_i, y_i) = (0, 0)$ θ_i is undefined.

In the polar coordinate system an Archimedean spiral is described by the relationship

$$r_i = \mathbf{a}\,(\theta_i + 360k);\ i=1, 2,\ldots,n \qquad \ldots(2)$$

where, **a** is a positive constant and θ_i and k are specified as in the relationship (1) above (Piskunov, pp. 27-28). In view of the relationship (2), the parametric equations of an Archimedean spiral may also be rewritten as

$$x_i = \mathbf{a}\,(\theta_i + 360k)\cos(\theta_i + 360k) = \mathbf{a}\,(\theta_i + 360k)\cos(\theta_i)$$
$$y_i = \mathbf{a}\,(\theta_i + 360k)\sin(\theta_i + 360k) = \mathbf{a}\,(\theta_i + 360k)\sin(\theta_i) \qquad \ldots(2a)$$

The usual procedure of curve-fitting by the method of Least Squares fails miserably in fitting a spiral to empirical data. The author has tried with several algorithms available for non-linear regression and non-linear optimization, viz. (i) Gauss-Newton, (ii) Powell, (iii) Nelder-Meade, (iv) Hooke-Jeeves, (v) Rosenbrock, (vi) Fletcher-Powell, (vii) Fletcher-Reeves, and (viii) Box algorithms. These algorithms and their FORTRAN codes are available in Kuester and Mize (1973). Failure of the available statistical software packages also is expected in fitting the spiral since these packages use the one or the other algorithm mentioned above.

The main reason for the failure of these algorithms is easily discernible. A spiral is a periodic function for which $f(\theta) = f(\theta + 360k)$ for any non-negative integer, k. Periodicity also results into multiple values of $f(\theta)$ for any given θ. The said algorithms are not designed for tackling such a situation since a good many genuine values of $f(\theta)$ are taken for errors by the procedure adopted by these algorithms.

The objective of this paper, therefore, is to devise an algorithm to fit an Archimedean spiral to empirical data. The algorithm would also be tested on numerical data.

2. Empirical Definitions : Empirical data, Z, are considered here as a sample drawn from the universe in which the relationship R = α (θ + 360k) holds. In the universe, X, Y and the related R and θ are real continuous variables. But sample (x, y) is a set of discrete points. For devising an algorithm proposed in this paper we need some empirical definitions which are facilitating ones and that is the only justification for enunciating them. These definitions are given as follows.

(i) ***Trace of an Archimedean Spiral*** : A non-empty finite set, **S**, of points (x_i, y_i); i=1, 2,…,n is called the trace of an Archimedean spiral if $r_i = (x_i^2 + y_i^2)^{1/2} = \mathbf{a}\ (\theta_i + 360k)$ for all i, where **a** > 0 is a constant; $0^0 \le \theta_i < 360^0$; $\theta_i = \arctan(y_i/x_i)$ for $x_i \neq 0$, otherwise $\theta_i = 90^0$ for $(x_i, y_i) = (0, > 0)$ and 270^0 for $(x_i, y_i) = (0, < 0)$; k is a non-negative integer. Moreover, $r_i \neq r_j$ for i≠j and $r_i = r_j$ implies $(x_i, y_i) = (x_j, y_j)$. The point $(x_i, y_i) = (0, 0)$ for which θ_i is undefined does not belong to **S**.

(ii) ***Iso-periodic Subset of the Trace of a Spiral*** : A non-empty sub-set $\mathbf{s_j}$ of set **S** is called an iso-periodic subset of order j if each and every point belonging to $\mathbf{s_j}$ satisfies the equation $r_i = a\ (\theta_i + 360j)$ for some fixed j, where j ≤ k and j, k ∈ (0, 1, 2,…,k). All points in $\mathbf{s_j}$ are iso-periodical. The set **S** is a union of all such subsets. Against this definition, one should note that in the universe each iso-periodic subset of the trace of a spiral is a collection of infinitely many points and the universe is a union of infinitely many such iso-periodic subsets.

(iii) ***Quadrant Index (IQ)*** : If a point $z_i = (x_i, y_i) \in \mathbf{S}$ lies in the quadrant q (q = 1, 2, 3 or 4) then $IQ_i = q$.

(iv) ***Iso-periodical Index (ISO)*** : If $r_i = \mathbf{a}\ (\theta_i + 360k)$ where k is a non-negative integer, then $ISO_i = k$. Numerically, it may be obtained as follows: Let there be two distinct points (r_i, θ_i) and (r_j, θ_j) in the polar co-ordinate system, then,

$$r_i/r_j = [\theta_i + 360(k_1)]/[\theta_j + 360(k_2)] \Leftrightarrow r_j[\theta_i + 360(k_1)] = r_i[\theta_j + 360(k_2)] \quad \text{…(3)}$$

Now, there exist two non-negative integers, k_1 and k_2, for which the equality (3) or its numerical approximation,

$$\mathrm{ABS}\{(r_j\theta_i - r_i\theta_j) - 360(r_ik_2 - r_jk_1)\} \le e \quad \text{…(4)}$$

for some small (positive) e, holds. Here ABS(.) means the absolute value of (.). Iteratively, the values of k_1 and k_2 may be found out from (4). Then k_1 and k_2 are the iso-periodical indices of i^{th} and j^{th} points, respectively. That is to say that $ISO_i = k_1$ and $ISO_j = k_2$.

3. The Algorithm : Given the data set $Z = (z_1, z_2, …, z_n)$; $n \ge 3$

(i) Step 1: Find r_i for all i.

(ii) Step 2: Arrange r_i (and associated z_i) in an ascending order of their magnitudes. Delete the observation if $r_i = 0$. Delete the observations for which $r_i = r_j$ (i≠j of course). Replace n by n-nd, where nd is the number of points thus deleted.

(iii) Step 3: Find IQ_i for all i as described in (iii) of section 2.

(iv) Step 4: Find $\theta_i\ \forall$ i by the following formula (where int(.) is integer value of (.)):

$$\begin{aligned}\theta_i &= [\{\mathrm{Arctan}(y_i/x_i)\} + \{\mathrm{int}(IQ_i/2)\}*180] \text{ for } x_i, y_i \neq 0\\ &= 90 \text{ for } (x_i, y_i) = (0, > 0)\\ &= 270 \text{ for } (x_i, y_i) = (0, < 0). \quad \text{…(5).}\end{aligned}$$

(v) Step 5: Find ISO_i $\forall$ i as described in the definition of ISO_i in (iv) of section 2.

(vi) Step 6: Find $\mathbf{a_i} = r_i/(\ \theta_i\ + 360*ISO_i\ )$ and $\bar{\mathbf{a}} = (\Sigma \mathbf{a_i})/n$. This $\bar{\mathbf{a}}$ is the first approximation of α in the universe from which the sample was drawn.

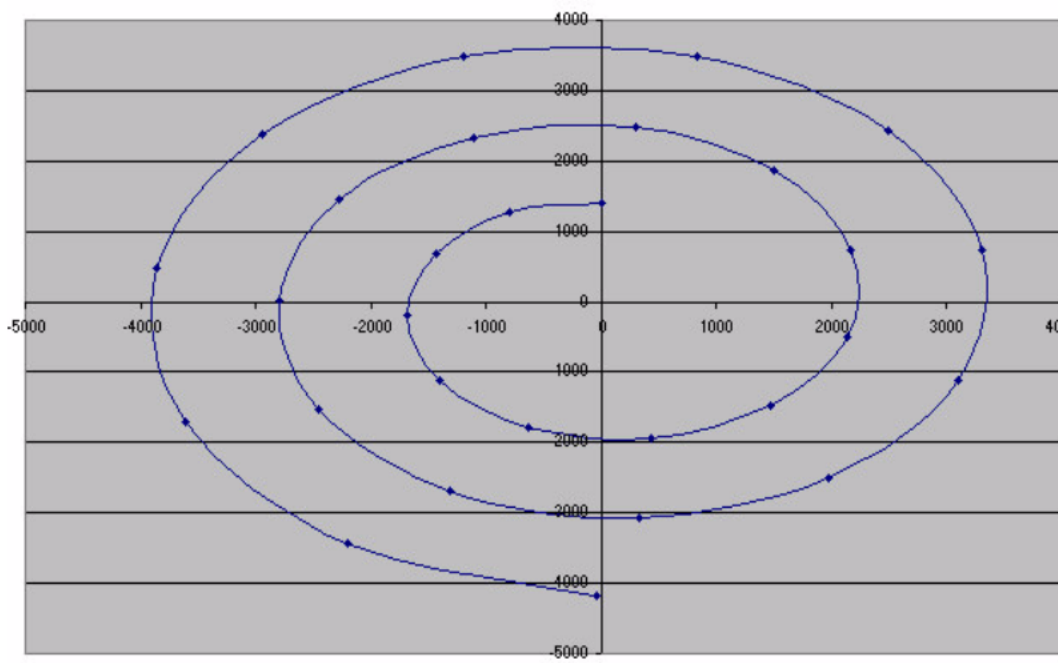
(vii) Step 7: Using Least squares principle estimate $\hat{\mathbf{a}}$ and c in $r_i = \hat{\mathbf{a}}\ (\theta_i + 360*ISO_i) + c$. It may be so that $\bar{\mathbf{a}}$ and $\hat{\mathbf{a}}$ are identical. Here c is the constant of displacement (of the pole of the spiral). This estimation is reliable if c is very close to zero (statistically, not significantly different from zero). When errors in x or y do not disturb the true iso-periodical index of the point, the estimator is quite reliable and c is negligibly small.

4. A Numerical Example : To test the functioning of the algorithm, a sample of 30 points was generated using a computer program (written in BASIC).

Table 4.1: Data Generated by the Computer Program

x	y	x	y	x	y
4	1393	1500	1859	3307	727
-791	1266	292	2471	2484	2444
-1434	695	-1110	2338	826	3488
-1684	-185	-2269	1443	-1189	3487
-1407	-1114	-2789	16	-2939	2384
-633	-1785	-2456	-1522	-3857	467
434	-1946	-1315	-2685	-3605	-1698
1472	-1487	324	-3072	-2205	-3439
2133	-500	1969	-2508	-40	-4185
2166	744	3094	-1110	2243	3650

The Simulated Archimedean Spiral (Table 4.1)



The inputs given were : n=30, **a** = 3.1, seed for generating the random number = 3211, nr = 50 (that is, 50 uniformly distributed random numbers are generated and averaged to obtain a single normally distributed random number, using the Central Limit theorem), sdx = 1, sdy = 3, theta = 32.13, delta = 13, output file F$ = spir.rn to save the data thus generated, using FORM$ = ####### so that the nearest integral part of the generated data was stored. Then the program for fitting the spiral was run with parameters n=30; F$=spir.rn and KMAX = 100. The program ran successfully. It goes without saying that while developing the program several experiments under varying conditions were carried out.

The results are: Mean alpha = 3.100045, Least Squares alpha = 3.100814 and $c = -0.640279$. Thus the estimated alpha was very close to the parameter with which the data (sample) was generated. So, the algorithm works well.

5. Results of a Monte Carlo Experiment : A Monte Carlo experiment was carried out to evaluate the efficiency and unbiasedness of the estimator obtained by the proposed algorithm. A Monte Carlo experiment (Froberg, pp. 306-308) simulates the process of estimating parameters, e.g. alpha (α) in the spiral R = $\alpha(\theta + 360k)$ in the present case, using a controlled setting in which the true parameter values are known. Details of the Monte Carlo experimental methodology are available in Hamersley and Handscomb (1964) and Naylor (1969).

Various criteria are used in the comparison of estimates to the true values of parameters. Let α be the known true value of a parameter and $\mathbf{a}_i$ be the estimated at the i^{th} replication (of the experiment), then the mean estimate of α is given by $\bar{\mathbf{a}} = (\Sigma a_i)/m$, where m is the number of replicates in the experiment. The bias, b, is then estimated numerically as b = ($\bar{\mathbf{a}}$ - α), where $\bar{\mathbf{a}}$ is taken as a sample measure of expectation of **a**, i.e., $\bar{\mathbf{a}} = E(\mathbf{a})$. The bias is an important criterion to evaluating the small sample properties of estimators. Another criterion is the variance, v (or standard deviation, σ), defined as $v = (\Sigma(a_i - \bar{\mathbf{a}})^2)/m$ (or $\sigma = v^{½}$), measuring the dispersion of the estimates about their mean value. A third criterion is the root mean square of errors, RMS, defined as RMS = $[\Sigma(a_i - \alpha)^2\}/m\}]^{½}$, measuring the dispersion of the estimates about the true value. It may be noted that since RMS = $(v + b^2)^{½}$, only two of these three criteria are independent (Intriligator, pp. 416-420).

The results of Monte Carlo experiments are given below. In case of table 5.1, the specifications are: Seed for random number generation = 3211; No. of replications (m) = 30; No. of points generated in each sample replicate (n) = 10, 30 and 100; No. of uniformly distributed random numbers to obtain (Gillett, pp. 518-519) normally distributed random numbers (nr) = 30; α = 3.1; delta = 13; and the angle (θ) = 32.13. The errors were normally distributed with zero mean and 1 and 3 standard deviations for x and y respectively.

Table 5.1: Results of Monte Carlo Experiments - I

Sample Size	n = 10		n = 30		n = 100	
Estimator/Criterion	Bias	RMS	Bias	RMS	Bias	RMS
Mean Alpha	0.00	0.000	0.00	0.00	0.00	0.00
LS Alpha	0.01	0.005	0.00	0.00	0.00	0.00
c of displacement	-2.51	2.506	-0.65	0.65	-0.05	0.05

For table 5.2, the values are: $\alpha = 2.0$; delta = 10; angle (θ) =23; and standard deviation = 2 and 4 for x and y respectively. Other values are fixed as for table 5.1.

Table 5.2: Results of Monte Carlo Experiments - II						
Sample Size	n = 10		n = 30		n = 100	
Estimator/Criterion	Bias	RMS	Bias	RMS	Bias	RMS
Mean Alpha	0.00	0.000	0.00	0.00	0.00	0.00
LS Alpha	0.02	0.002	0.00	0.00	0.00	0.00
c of displacement	8.55	8.556	0.87	0.88	0.07	0.07

A perusal of these tables reveals that: (1) Bias and RMS are negligibly small for α, even for small samples, and (2) Bias and RMS of the constant of displacement, c, are larger for small samples, but become smaller for larger samples, indicating that the estimator is consistent.

6. Concluding Remarks : Although it is possible to indirectly estimate the parameters of a spiral by fitting Fourier's polynomial (Piskunov, pp. 355-356), a direct algorithm to fit a spiral may have its own advantages. The algorithm proposed here fits an Archimedean spiral directly. It may be possible to fit other types of spiral (e.g. logarithmic, exponential, hyperbolic, etc.) by analogous procedures.

In practice, we may come across empirical data, which indicate a pattern resembling a family of ellipses in the positive orthant (first quadrant). In a developing economy, technological innovations over time may push up production possibility curves generating such a pattern. In biological investigations one may find a pattern that resembles a spiral. The algorithm developed here may find its application in such instances.

References

1. Buck, RC and Buck, EF (1965). *Advanced Calculus*. Mc-Graw Hill, New York.
2. Froberg, CE (1965). *Introduction to Numerical Analysis*. Addison Wesley, London.
3. Gillett, BE (1979). *Introduction to Operations Research: A Computer-Oriented Algorithmic Approach*. Tata McGraw Hill, New Delhi.
4. Hamersley, JM and Handscomb, DC (1964). *Monte Carlo Methods*. John Wiley and Sons. New York.
5. Intriligator, MD (1978). *Econometric Models, Techniques and Applications*. Prentice Hall, New Jersey.
6. Kuester, JL and Mize, JH (1973). *Optimization Techniques with FORTRAN IV*. McGraw Hill, New York.
7. Naylor, TH (1969). The Design of Computer Simulation Experiments. (ed) Duke Univ. Press, Durham, NC.
8. Piskunov, N (1974). *Differential and Integral Calculus. Vol. I & II*. Mir Publishers, Moscow.

APPENDIX-1 : BASIC COMPUTER PROGRAMS

This Basic Program generates data for Archimedean Spiral

```
10 RANDOMIZE
20 DEFINT I-N
30 INPUT"NUMBER OF POINTS TO BE GENERATED (N=SAMPLE SIZE) ?",N
40 DIM X(N), Y(N), ANGLE(N), R(N),IQ(N),THETA(N)
50 DIM IP(N), REXP(N), ISO(N), NIQ(4)
60 INPUT"FILE TO STORE GENERATED DATA ? ",F$
70 INPUT"FEED THE VALUE OF A ",A
80 INPUT"NO. OF RANDOM POINTS TO OBTAIN NORMAL DSTBN (NR) ? ",NR
90 REM  NR NUMBER OF UNIFORMLY DISTRIBUTED RANDOM NUMBERS ARE
100 REM  AVERAGED TO OBTAIN A NORMALLY DISTRIBUTED RANDOM NUMBER.
110 REM IF NR = 1, RECTANGULAR DISTRIBUTION OF ERRORS
120 REM IF NR > 30, NORMAL DISTRIBUTION OF ERRORS. USUALLY NR <100.
130 INPUT"STANDARD DEVIATION OF E(X) AND E(Y) ? ",SDX,SDY
140 INPUT"FEED DELTA FOR INCREMENT OR SPACING (DELTA)? ",DELTA
150 INPUT"FEED ANGLE (THETA) ? ",THETA
160 INPUT"FORMAT TO PRINT DATA (e.g. #####.###) ? ",FORM$
170 REM FOR SDX = 0 NO ERROR ADDED TO X
180 REM FOR SDY = 0 NO ERROR ADDED TO Y
190 PI=ATN(1)*4: FACT=PI/180
200 FOR I=1 TO N
210 RX=0:RY=0
220 FOR J=1 TO NR
230 RX=RND(RX):XR=RX
240 RX=RND(RX):YR=RX
250 NEXT J
260 YR=(YR/NR-.5)*SDY : XR=(XR/NR-.5)*SDX
270 ANGLE(I)=(I+DELTA)*THETA
280 X(I)=A*ANGLE(I)*COS(ANGLE(I)*FACT)+XR
290 Y(I)=A*ANGLE(I)*SIN(ANGLE(I)*FACT)+YR
300 NEXT I
310 OPEN"O",#1, F$
320 FOR I=1 TO N
330 PRINT #1, USING FORM$;X(I);Y(I)
340 REM PRINT X(I), Y(I), ANGLE(I)
350 NEXT I
360 CLOSE
370 END
```

This Basic Program estimates parameters from the generated data

```
10 REM INITIALIZATION AND DEFINITIONS
20 DEFINT I-N
30 INPUT"NUMBER OF OBSERVATIONS (N) ? ",N
40 IF(N<3) THEN PRINT "No. Give valid value of N =>3 ": GOTO 30
50 DIM R(N), REXP(N), THETA(N), X(N), Y(N)
60 DIM IP(N), IQ(N), ISO(N), NIQ(4)
70 PI=ATN(1)*4 : FACT = PI/180 : EPS = .00001
80 INPUT"FILE IN WHICH DATA ARE STORED ? ",F$
90 IF(F$="") THEN PRINT "Give valid File name " : GOTO 80
100 INPUT"MAXIMUM ORDER OF SPIRAL ? No idea ? THEN 100 ",KMAX
110 IF(KMAX<=0) THEN KMAX=100
120 REM READING (X, Y) DATA FROM THE INPUT FILE
130 OPEN"I",#1, F$
```

```
140 FOR I=1 TO N
150 INPUT#1, X(I), Y(I)
160 NEXT I
170 CLOSE
180 REM FINDING RADIUS, QUADRANT INDEX AND ANGLE OF (X, Y)
190 FOR I=1 TO N
200 R(I) = SQR(X(I)^2 + Y(I)^2)
210 IQ(I)=1
220 IF(X(I) < 0 AND Y(I) >= 0) THEN IQ(I)=2
230 IF(X(I) < 0 AND Y(I) < 0) THEN IQ(I)=3
240 IF(X(I) => 0 AND Y(I) < 0) THEN IQ(I)=4
250 IF(X(I) <> 0) THEN TH = Y(I)/X(I) : IL = 1 ELSE IL = 0
260 IF(IL = 1) THEN THETA(I)=ATN(TH)/FACT+INT(IQ(I)/2)*180
270 IF(X(I) = 0 AND Y(I) > 0) THEN THETA(I)= 90
280 IF(X(I) = 0 AND Y(I) < 0) THEN THETA(I) = 270
290 NEXT I
300 REM ARRANGING DATA IN ASCENDING ORDER OF RADIUS
310 FOR I=1 TO N-1
320 FOR II=I+1 TO N
330 IF(R(I)<>R(II)) THEN GOTO 350
340 PRINT"TWO RADII FOR  ";I;II;" EQUAL. " : GOTO 890
350 IF(R(I) < R(II)) GOTO 410
360 AA=R(I) : R(I)=R(II) : R(II)=AA
370 AA=X(I) : X(I)=X(II) : X(II)=AA
380 AA=Y(I) : Y(I)=Y(II) : Y(II)=AA
390 AA=IQ(I): IQ(I)=IQ(II):IQ(II)=AA
400 AA=THETA(I): THETA(I)=THETA(II):THETA(II)=AA
410 NEXT II
420 NEXT I
430 PRINT"NUMBER OF POINTS IN Ist TO IVth QUADRANT ARE :"
440 FOR J=1 TO 4
450 NIQ(J)=0
460 FOR I=1 TO N
470 IF(IQ(I)=J) THEN NIQ(J)=NIQ(J)+1
480 NEXT I
490 PRINT NIQ(J);
500 NEXT J
510 PRINT
520 REM INITIALIZATION OF ISO-PERIODICAL INDEX
530 FOR I=1 TO N
540 ISO(I)=-999
550 NEXT I
560 REM FIND ISO-PERIODICAL INDEX FOR EACH POINT
570 EPS1=EPS : EPSMAX=EPS
580 FOR K1=0 TO KMAX
590 FOR K2=K1 TO KMAX
600 R1=(R(N)*THETA(1)-R(1)*THETA(N))
610 R2=(K2*R(1)-K1*R(N))*360
620 IF(ABS(R1-R2) < EPS) GOTO 680
630 NEXT K2,K1
640 REM EPS IS TOO SMALL. ACCURACY IN DATA DOES NOT PERMIT
650 REM THIS SMALL EPS. SO EPS IS NOW INCREASED.
660 EPS=EPS*10
670 GOTO 580
680 ISO(1)=K1 : ISO(N)=K2
690 IF(EPSMAX < EPS) THEN EPSMAX = EPS
700 FOR I=2 TO N-1
```

```
710 K1=ISO(1) : RR=R(1) : TH = THETA(1) : EPS = EPS1
720 FOR K2=K1 TO ISO(N)
730 R1=(R(I)*TH-RR*THETA(I))
740 R2=(K2*RR-K1*R(I))*360
750 IF(ABS(R1-R2) < EPS) GOTO 790
760 NEXT K2
770 EPS=EPS*10 : IF(EPSMAX < EPS) THEN EPSMAX=EPS
780 GOTO 720
790 ISO(I)=K2
800 NEXT I
810 PRINT"MAXIMUM EPS USED IS ";
820 PRINT USING "######.######";EPSMAX
830 REM CHECK FOR DYFUNCTIONALITY
840 NISO=0
850 FOR I=1 TO N
860 IF(ISO(I) > 0) THEN NISO=NISO+1
870 NEXT I
880 IF(NISO <> 0) GOTO 920
890 PRINT "PROGRAM FAILED. POSSIBLY DATA ARE DEFECTIVE."
900 PRINT"CHECK DATA; DELETE (X, Y)=(0, 0), IF ANY"
910 PRINT"DELETE OBSERVATIONS FOR WHICH RI = RJ" : STOP
920 ASUM=0 : ASSUM=0 : NA=0
930 REM FIND MEAN VALUE OF ALPHA
940 FOR I=1 TO N
950 AL=(THETA(I)+360*ISO(I))
960 IF(AL=0) GOTO 1000
970 ALP=R(I)/AL
980 ASUM=ASUM+ALP : ASSUM=ASSUM+ALP^2
990 NA=NA+1
1000 NEXT I
1010 ALPHA=ASUM/NA
1020 PRINT "MEAN ALPHA IS ";ALPHA
1030 REM FIND LEAST SQUARES VALOE OF ALPHA
1040 AX=0 : AXX=0 : AXY=0 : AY=0
1050 FOR I=1 TO N
1060 AL=(THETA(I)+360*ISO(I))
1070 AX=AX+AL : AXX=AXX+AL^2
1080 AXY=AXY+AL*R(I) : AY=AY+R(I)
1090 NEXT I
1100 DET=N*AXX-AX^2
1110 D1=AXX*AY-AX*AXY
1120 D2=N*AXY-AY*AX
1130 ALPHA=D2/DET : C=D1/DET
1140 PRINT "LEAST SQUARES ALPHA IS = ";ALPHA
1150 PRINT"THE CONSTANT OF DISPLACEMENT IS = ";C
1160 END
```