A Note on Least-Squares Fitting of Signal Waveforms

SK Mishra
Dept. of Economics
North-Eastern Hill University
Shillong, Meghalaya (India)
E-mail: mishrasknehu@yahoo.com

Introduction: In their recently published paper Han et al. (2006) fit the function

$$s(t) = \sum_{j=1}^{j\max} \frac{a_j}{1-b_j\tau}\left[\exp(-b_j t) - \exp(-t/\tau)\right] \qquad \ldots \qquad (1)$$

to the data of signal waveforms associated with optogalvanic (OG) transitions recorded with a hollow cathode discharge tube containing a mixture of neon (Ne) and carbon monoxide (CO) gases. Before fitting the function to the actual experimental data, they validate the method of curve fitting with the simulated data consisting of 601 points of $[s(t),t]$, where t=0, 0.25, 0.50, …, 150 at equal intervals. They observe that the regular least squares fitting technique is unstable when used to fit exponential functions to signal waveforms, since such functions are highly correlated. Therefore, they devise a procedure based on Monte Carlo method, utilizing both search and random walk. They report that the proposed procedure gives a stable least squares fitting algorithm that converges very rapidly. For their simulated data the details are given as follows.

Table-1: Details of the Curve Fitting Exercise on the Simulated Data (Han et al.)

Parameters	τ	a_1	b_1	a_2	b_2	a_3	b_3	χ^2	Computation time (min)	CPU GHz
True	0.50	2.00	0.20	-2.00	0.10	0.50	0.05	-	-	-
Guessed	0.10	0.50	0.30	-0.40	0.05	0.10	0.02	-	-	-
Estimated	0.511	2.07	0.202	-1.90	0.107	0.334	0.0443	3.2E-07	≈ 20 minutes	1.73

A perusal of the results shows that Han et al. have been able to estimate the parameters that are quite close to the true ones (with which the simulated data were generated). For whatever little discrepancies remain, they hold that one reason is that some parameters (a_3 and b_3) are small, so that the small differences between their true values and the fitted values amount to a larger percentage difference. Further, "the correlations between the parameters also contribute to the fact that the fitted values do not equal the true values." They do not attribute the discrepancies arising out of the power of their algorithm. This is exactly what has attracted us to this work.

Estimation of Parameters by the Differential Evolution based Algorithm: We generate the data (601 points) and fit the equation (1) to it by an algorithm that obtains least squares by the Differential Evolution method of global optimization (Mishra, 2007). This method has shown a great power in fitting nonlinear curves to datasets given by NIST (National Institute of Standards and Technology, USA), CPC-X and others.

We rename $s(t)$ as $y = [y_i; i=1,2,...,601)$ and t as $x=[x_i; i=1,2,...,601]$. The (FORTRAN) program for generating the data is given below.

```fortran
      parameter (n=601, m=3)
      implicit double precision (a-h, o-z)
      character *20 fo
      dimension a(m),b(m), x(n),y(n)
      COMMON /RNDM/IU,IV
      fo='dat40'
      jmax = 3
      tau  = 0.5d0
      a(1) = 2.d0
      b(1) = 0.2d0
      a(2) = -2.d0
      b(2) = 0.1d0
      a(3) = 0.5d0
      b(3) = 0.05d0
      del=150.d0/600.d0
C     generate x(i), y(i), i=1,n
      open(7, file=fo)
      do i=1,n
      x(i)=(i-1)*del
      y(i)=0.d0
      do j=1,jmax
      v=(a(j)/(1.d0-b(j)*tau))*(dexp(-b(j)*x(i))-dexp(-x(i)/tau))
      y(i)=y(i)+v
      enddo
      write(7,1)i, x(i),y(i)
      enddo
      close(7)
    1 format(i5,F10.5,f16.12)
      write(*,*)'over'
      end
```

When we estimate the parameters of the model (equation –1) we obtain the results that are much more accurate than those obtained by Han et al. It may be noted that first we have run the program on a very slow computer (7.5 MHz). Even then, the solution has taken about 25 minutes only. Hahn et al. had obtained their results on a 1.7 GHz machine in 20 minutes. Further, our starting points are not as close to the true parameters as set by Han et al. We have given wider range as the domain of parameters. One set of our results is given in Table-2 below (time taken 25 min approx). When we run our program on a fast computer (Compaq SG3370IL – Core 2 Duo E4600 model of 2.4 GHz) we obtain in 3 minutes and 26 seconds the results that are equal to the true parameters (see table 3).

Table-2:Details of the Curve Fitting Exercise on the Simulated Data (Present Study)

Parameters	τ	a_1	b_1	a_2	b_2	a_3	b_3
True	0.50	2.00	0.20	-2.00	0.10	0.50	0.05
Range	0 - 1	0 - 4	0 - 1	0 - 3	0 - 1	0 - 1	0 - 1
Estimated	0.5000374	1.9999816	0.2000151	-1.9993511	0.1000169	0.4993898	0.0499842

For the estimates reported in Table-2 we have obtained s^2 = 0.2961362927800704E-08 where $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{601} (y_i - \hat{y}_i)^2 = s^2$ which is a measure of the goodness of fit.

```fortran
      SUBROUTINE REGMODEL_40(M,X,F,NQ)!optogalvanic (OG) effct FUNCTION
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      COMMON /RNDM/IU,IV
      COMMON /REGDAT40/DMODEL40(601,2),NDM40
      INTEGER IU,IV
      DIMENSION X(*),ALIM(7), a(3), b(3)
      DATA (ALIM(I),I=1,7)/1, 4,  1, 3,  1,  1,  1/
      jmax = 3
      IF(NQ.NE.0) OPEN(11,FILE='REGRES')
       DO J=1,M
       IF(X(J).GT.ALIM(J).OR. X(J).LT.0.D0) THEN
       CALL RANDOM(RAND)
       X(J)=ALIM(J)*RAND
       ENDIF
       ENDDO
      SY=0.D0
      SSY=0.D0
      S=0.D0
      SER=0.D0
      N=NDM40 ! N IS THE NUMBER OF OBSERVATIONS IN DATA SET 1
C     REDEFINE PARAMETERS TO BE ESTIMATED
      tau=X(1)
      a(1)=X(2)
      b(1)=X(3)
      a(2)=-x(4)  !(negatively signed) Note this
      b(2)=x(5)
      a(3)=x(6)
      b(3)=x(7)
      DO I=1,N
      X1=DMODEL40(I,1)  ! REGRESSOR REDEFINED AZ X1
      Y=DMODEL40(I,2)  ! REGRESSAND REDEFINED AS Y; EXPECTED Y IS YX
C     computing the function
      yx=0.d0
      do j=1,jmax
      v=(a(j)/(1.d0-b(j)*tau))*(dexp(-b(j)*x1)-dexp(-x1/tau))
      yx=yx+v
      enddo
      IF(NQ.NE.0) WRITE(11,*) I,x1,Y,YX
      ER=Y-YX ! ERROR
      SER=SER+ER
      S=S+ER**2
      SY=SY+Y
      SSY=SSY+Y**2
      ENDDO
      RMS2=S/N
      V=SSY/N-(SY/N)**2
      F=-(1.D0-RMS2/V)   ! FUNCTION TO MINIMISE (IT MAXIMIZES R_SQUARE)
      f=s
      IF(NQ.NE.0) WRITE(11,1) (1.D0-RMS2/V),f,DSQRT(S/N)
    1 FORMAT('R-SQUARE = ',F20.15,' S-SQUARE = ', 2F25.15)
      IF(NQ.NE.0) CLOSE(11)
      RETURN
      END
```

Lest it causes confusion, it is to be noted that we have changed the sign of the parameter a_2 in our program so that all the parameters may become positive and lie in the ranges specified by us. This endeavor has saved a line or two in our program. Nevertheless, it was not necessary to do that. The subroutine that defines the OG function is given above (Subroutine REGMODEL_40). It is called by the Main/DE Program for function evaluation. The DE program may be obtained from the author.

Table-3: Sample Results of Differential Evaluation (Seed = 3311)
Time Taken : 3 min 26 sec

```
NAME OF OUTPUT FILE IN WHICH RESULTS WILL BE STORED
waveout.txt
 COMPUTING. PLEASE WAIT. LARGER PROBLEMS TAKE MORE TIME
 ******************************************************
 PROBLEM= 1
 NONLINEAR REGRESSION PROBLEM
 ************* INTERMEDIATE RESULTS ***************
 ESTIMATED PARAMETERS UP TO NOW
  0.508946197  1.84720373  0.213983383  1.57902016  0.106683223  0.24472686
  0.0391884307
 SUM OF SQUARED ERRORS UP TO NOW =       0.0015285182701856
 -----------------------------------------------------
 ESTIMATED PARAMETERS UP TO NOW
  0.499974632  2.00317083  0.199890766  2.00125949  0.100107552  0.498019836
  0.0499538737
 SUM OF SQUARED ERRORS UP TO NOW =       0.0000000178860683
 -----------------------------------------------------
 ESTIMATED PARAMETERS UP TO NOW
  0.500000001  1.99999993  0.200000002  1.99999999  0.0999999974  0.500000051
  0.0500000012
 SUM OF SQUARED ERRORS UP TO NOW =       0.0000000000000000
 -----------------------------------------------------
 ESTIMATED PARAMETERS UP TO NOW
  0.5  2.  0.2  2.  0.1  0.5  0.05
 SUM OF SQUARED ERRORS UP TO NOW =       0.0000000000000000
 -----------------------------------------------------
 COMPUTATION OVER. RESULTS STORED IN THE FOLLOWING FILE
 waveout.txt
 THANK YOU
```

Our results indicate that neither of the two reasons postulated by Han et al. is responsible for the discrepancies in the true parameters and the estimated parameters obtained by them. If it were not so, the DE based least squares algorithm would not have given us the results that are much more close (rather identical) to the true parameters.

References

- Han, XL, Pozdin, V, Haridass, C and Misra, P (2006) "Monte Carlo Least-Squares Fitting of Experimental Signal Waveforms", *Journal of Information & Computational Science,* 3(4), pp. 1-7. http://www.physics1.howard.edu/~pmisra/publications/137_ISICS06.pdf
- Mishra, SK (2007) "Performance of Differential Evolution Method in Least Squares Fitting of Some Typical Nonlinear Curves", SSRN http://ssrn.com/abstract=1010508