# CONTIN: A GENERAL PURPOSE CONSTRAINED REGULARIZATION PROGRAM FOR INVERTING NOISY LINEAR ALGEBRAIC AND INTEGRAL EQUATIONS

# Stephen W. PROVENCHER

EMBL, Postfach 10.2209, D-6900 Heidelberg, Fed. Rep. Germany

Received 5 May 1982

### PROGRAM SUMMARY

Title of program: CONTIN

Catalogue number: AAOB (version 2DP), AAOC (version 2 SP)

Program obtainable from: CPC Program Library, Queen's University of Belfast, N. Ireland (see application form in this issue)

Computer: VAX 11/780; Installation: European Molecular Biology Laboratory (EMBL). The program is intended to be fully portable. After setting four installation dependent values, it should run without modification on most systems supporting 1966 ANSI standard Fortran IV. Version 2SP is recommended for machines (e.g., CRAY or CDC) with Fortran REAL representations of at least 60 bits and Version 2DP for all other machines

Operating system: VMS

Programming language used: Fortran IV (1966 ANSI standard, except for 1 in the rightmost DIMENSION specification of some dummy arrays)

High speed storage required: about 200 kbytes (depending on the size of the problem)

No. of bits in a byte: 8

Overlay structure: none

No. of cards in combined program and test deck: 5822

Card punching code: EBCDIC

Keywords: regularization, inequality constraints, ill-posed, inverse problems, integral equations, quadratic programming, deconvolution, information content, superresolution, photon correlation, constrained least squares

Nature of the physical problem

Many experiments are indirect in that the observed data are

linear integral (or matrix) transforms of the quantities to be estimated. These transforms typically arise because of the imperfect impulse response of the detection system or because of the indirect nature of the experiment itself (as with Fourier transforms in diffraction and Laplace transforms in relaxation experiments). The inversion of these linear operator equations are generally ill-posed problems in that there exists a large number of possible solutions (with arbitrarily large deviations from each other) all of which fit the data to within experimental error. Therefore straightforward inversion procedures cannot be used and statistical regularization techniques are necessary.

#### Method of solution

A general purpose constrained regularization method [1] finds the simplest (most parsimonious) solution that is consistent with prior knowledge and the experimental data. The problem is formulated as a weighted least squares problem with an added quadratic form, the regularizor, which imposes parsimony (typically smoothness) or statistical prior knowledge. Numerically stable orthogonal decomposition and quadratic programming algorithms [2] are used to obtain the unique global solution subject to any linear equality or inequality constraints imposed by prior knowledge (e.g., nonnegativity). The regularization parameter can be automatically chosen on the basis of an *F*-test and confidence regions.

#### Restrictions on the complexity of the problem

Part of the computation time is proportional to the cube of the number of parameters used to represent the solution. Computations with no more than about 100 parameters can be done economically. This is usually more than adequate for solutions in one dimension, but not for two- or three-dimensional solutions.

## Typical running time

Execution times depend on the complexity of the solution, constraints and operator. The test data sets in this paper each took about 180 s on the VAX 11/780. When least squares weights do not have to be calculated from a fit to the data, the time is reduced by a factor of two.

0010-4655/82/0000-0000/\$02.75 © 1982 North-Holland

#### Unusual features of the program

CONTIN has been designed to be easily adaptable to a wide variety of problems, but still easy to use. It consists of a fixed core of 53 subprograms plus 13 simple and thoroughly documented "USER" subprograms that define nearly all aspects of the problem. These USER subprograms usually do not have to be changed, but they can be easily modified to define non-standard integral equations, data preprocessing, simulations, constraints, regularizors, statistical weighting, output, etc.

# References

- [1] S.W. Provencher, Comput. Phys. Commun. 27 (1982) 213.
- [2] C.L. Lawson and R.J. Hanson, Solving Least Squares Problems (Prentice-Hall, Englewood Cliffs, 1974).

#### LONG WRITE-UP

#### 1. Introduction

CONTIN implements the constrained regularization algorithm described in detail in ref. [1]. (In general the notation and terminology of ref. [1] will be used in this paper without redefinition.) CONTIN is a general purpose package for the automatic inversion of noisy linear algebraic and integral equations, eqs. (1.2) and (3.2) of ref. [1], subject to the optional linear equality and inequality constraints in eqs. (3.6) and (3.7) of ref. [1]. CONTIN has been designed so that the user can easily adapt it to his particular problem. This paper outlines the structure, usage and testing of CONTIN.

## 2. Structure of the code

Table 1 lists the 66 subprograms in CONTIN. A complete block diagram showing the calling dependencies of the subprograms is given in ref. [2].

All COMMON variables are in four COM-MON blocks, DBLOCK, SBLOCK, IBLOCK and LBLOCK, containing variables of type DOUBLE PRECISION (REAL in version 2SP), REAL, IN-TEGER and LOGICAL, respectively. These contain all the control variables (listed in table 4) plus a few others (listed in table 2). The four COM-MON blocks always appear together in a subprogram and the variable lists are always the same. In general, the values are set at the beginning and are not changed during the analysis of a data set.

Table 1 CONTIN subprograms

Name	Purpose
	Initialization and input
BLOCK DATA	Initializes control variables
INIT	Initializes other variables, computes rela-
	tive machine precision
INPUT	Reads all input data for one data set
STORIN	Checks and stores control variables
READYT	Reads $y_k$ , $t_k$ and $W_k$
USERIN	Preprocesses input data
	Problem setup
SETGRD	Sets up grid of $\lambda_m$ values in eq. (3.1) of ref. [1]
USERGR	Sets up special-purpose grid of $\lambda_m$ values
CQTRAP	Sets $c_m$ values for trapezoidal rule in eq. (3.1) of ref. [1]
USERTR	Defines $h(\lambda)$ for $\lambda$ -grid (see section 4.2 of ref. [1])
USERSI	Computes simulated noisy $y_k$ values
USEREX	Delivers noise-free simulated $y_k$ to USERSI
RGAUSS	Computes pseudorandom normal deviates
RANDOM	Computes uniformly distributed pseudo- random deviates
USERSX	Computes extra curve to be plotted with the solution
USERNQ	Sets up inequality constraints in eq. (3.6) of ref. [1]
SETNNG	Sets up nonnegativity constraints in eq. (3.6) of ref. [1]
SETWT	Computes $W_k$ in section 4.4 of ref. [1]
USERWT	Computes special-purpose $W_k$ for SETWT
SETREG	Sets up $R$ and $r$ in eq. (3.10) of ref. [1]
USERRG	Delivers special-purpose R and r to SETREG
USEREQ	Sets up $E$ and $e$ in eq. (3.7) of ref. [1]
SETSGN	Sets up monotonic regions (see section 4.7 of ref. [1])
ANALYZ	Controls complete series of solutions in the appendix of ref. [1]

Table 1 (continued)

Name	Purpose
	Preliminary computations
SETSCA	Sets up internal scaling of variables
SEQACC	Does sequential accumulation of A to C in eq. (A.1) of ref. [1]
GETROW	Delivers one row of $A$ in eq. (3.2) of ref.
USERK	Computes $F_k(\lambda_m)$ in eq. (3.1) of ref. [1]
USERLF	Computes $L_{ki}$ in eq. (3.1) of ref. [1]
H12	Performs a Householder transformation [6]
ELIMEQ	Computes $K$ and $x_1$ in eqs. (A.2) and (A.4) of ref. [1]
LH1405	Computes $\eta$ - $CK_1x_1$ in eq. (A.5) of ref. [1]
SVDRS2	Does a singular value decomposition of a matrix [6]
QRBD	Does a singular value decomposition of a bidiagonal matrix [6]
G1	Constructs a Givens rotation matrix [6]
G2	Applies a Givens rotation matrix [6]
DIFF	Delivers the difference of two numbers
DIAREG	Computes $CK_2ZH_1^{-1}$ and $K_2ZH_1^{-1}$ in eq.
	(A.13) of ref. [1] Computes $\gamma$ and $K_2 Z H_1^{-1} W$ in eqs. (A.19)
DIAGA	and (A.29) of ref. [1] Computes $DK_2ZH_1^{-1}W$ in eq. (A.28) of
SETGA1	ref. [1]
SETVAL	Core computations  Computes $\tilde{S}_{jj}^{-1}$ and $K_1x_1 + K_2ZH_1^{-1}$ $\times [W\tilde{S}^{-1}\tilde{\gamma} + r_1]$ in eqs. (A.22) and (A.29) of ref. [1]
LDPETC	Controls the solution of eqs. (A.27) and (A.28) of ref. [1] and its output
LDP	Solves the least distance programming problem [6]
NNLS	Solves the nonnegative least squares prob- lem [6]
CVNEQ	Computes error estimates and $N_{DF}$ in sections 3.5 and A.2 of ref. [1]
GETYLY	Computes the weighted residuals of the fit to the data
FISHNI	Computes the Fisher F-distribution for noninteger degrees of freedom
BETAIN	Computes the incomplete beta function
GAMLN	Computes the logarithm of the gamma function
MOMENT	Computes the moments of the solution
RUNRES	Controls the plots of the weighted residuals and the fit to the data
GETPRU	Performs autocorrelation and runs tests on the weighted residuals
PGAUSS	Computes the normal probability integral
ANPEAK	Controls peak-constrained solutions

Table 1 (continued)

Name	Purpose					
UPDSGN	Updates inequality constraints for peak- constrained solutions					
UPDDON	Skips combinations of monotonic regions that have already been analyzed					
FFLAT	.TRUE. if peak-constraints have artificially imposed a flat plateau					
UPDLLS	Updates boundaries of monotonic regions					
	Output					
WRITIN	Outputs control variables					
WRITYT	Outputs $y_k$ , $t_k$ and $W_k$ arrays					
PLPRIN	Plots a solution or a fit to the data					
USEROU	Does special-purpose output					
MOMOUT	Outputs the moments of a solution					
PLRES	Plots the weighted residuals					
ERRMES	Outputs error messages					

Table 2 COMMON variables that are not control variables

COMMON block	Name (DIMENSION spec.)	Purpose
DBLOCK	PRECIS	Approx. 10× (relative machine precision)
	RANGE	A very large number (see section 3.1)
SBLOCK	SRANGE	A very large number (see section 3.1)
	EXMAX	ALOG(SRANGE)
IBLOCK	IAPACK(6)	Program version name (see section 3.2)
	ITITLE(80)	Title in card 1 (see section 3.3)
	NGL	$N_{\rm x}$ in eq. (3.3) of ref.
	NGLP1	NGL+1
	NIN	Input unit device num- ber
	NINEQ	$N_{\text{ineq}}$ in eq. (3.6) of ref. [1]
	NOUT	Output unit device number
	NY	$N_{\rm y}$ in eq. (1.1) of ref. [1]

#### 3. Instructions for the user

### 3.1. Necessary changes

Every attempt was made to adhere to 1966 ANSI standard Fortran IV, with the common exception that some dummy arrays have a 1 in their rightmost DIMENSION specification. It is therefore intended that CONTIN will run without modification on most machines after the following four installation dependent variables have been set in BLOCK DATA.

RANGE should be a few orders of magnitude smaller than BIG, where BIG is the largest number such that BIG does not overflow and 1/BIG does not cause an underflow. RANGE and BIG are DOUBLE PRECISION in version 2DP and REAL in version 2SP.

SRANGE is the same as RANGE except that SRANGE (and BIG) is REAL in both version 2DP and 2SP. The default values for RANGE and SRANGE are 10<sup>35</sup>.

NIN is the input device unit number (default = 5). NOUT is the output device number (default = 6).

In many parts of CONTIN, floating underflows are expected to occur and be replaced by zero. With a few compilers it may be necessary to set a switch or call a subprogram to suppress underflow diagnostics. With some systems, input and output files may have to be opened with control cards or in the main subprogram (at places indicated with COMMENT cards).

### 3.2. Problem-dependent changes

The DIMENSION specifications of the large arrays need only be set in the main subprogram, which passes them along as dummy arguments to the rest of the subprograms. Therefore the arrays can be easily expanded to handle larger problems or reduced to save high speed storage, as explained in detail in the COMMENT cards and in ref. [2]. In BLOCK DATA the default values of the control variables can be reset to commonly used values to avoid having to input them for every run (see section 3.3).

Most important, the USER subprograms can be

easily modified to redefine nearly every aspect of the particular problem being solved. They are outlined in section 4 of ref. [1] and section 4 of this paper and fully documented in ref. [2] and in their COMMENT cards. The array IAPACK contains a six-character name that is output as part of the heading. IAPACK can be reset in BLOCK DATA to uniquely identify a modified version of CONTIN.

#### 3.3. Input data

Table 3 shows the structure of a data set. As explained in section 4.1 of ref. [1], several data sets can be placed in succession in an input deck for analysis in one run. An example of this is given in the test run input at the end of this paper. Card 1 contains a title, which then appears in various headings in the output.

Card set 2 is only necessary if some of the control variables are to be changed from their default values, which are initially set in BLOCK DATA. During a run, the values of the control variables are preserved from the analysis of one data set to the next. They can be reset with card set 2. Table 4 lists all the control variables. Their usage is outlined in section 4 of ref. [1] and described in detail in ref. [2]. Most control variables are input with one card, with FORMAT

$$(1X, 6A1, I5, E15.6).$$
 (1)

The 6A1 field contains the control variable name, left justified and filled with blanks. The I5 field contains the single subscript of the control variable, if it has one. The E15.6 field specifies the value. If the control variable is of type INTEGER, then the value (converted to REAL) must be in the E15.6 field. For LOGICAL control variables, only the values 1.0 (for .TRUE.) and -1.0 (for .FALSE.) are allowed. Seven control variables must be input with two successive cards. The first card has the same FORMAT as (1), but only the 6A1 field is used (for the name). The FORMAT specification for the second card is given in table 4.

Card 3 has the FORMAT in (1), but only the 6A1 field is used, with the characters 'END', left justified and filled with blanks.

Card set 4 provides a convenient way of speci-

Table 3
Composition of an input data set

Card	FORMAT	Input variables	Necessary when	
1	80A1	ITITLE(K), K = 1, 80	always	
set 2	(see section 3.3)	Control variables	never	
3	1X,6A1	'END' card	always	
set 4	1X,6A1,I5,2E15.6	'NSTEND', NT, TSTART, TEND (NINTT cards)	NINTT>0	
5a	1X,6A1,I5	'NY',NY	NINTT≤0	
set 5b	IFORMT	T(K), K = 1, NY	NINTT≤0	
set 6	IFORMY	Y(K), K = 1, NY	SIMULA = .FALSE	
set 7	IFORMW	W(K), K = 1, NY	IWT = 4	

Table 4 Control variables

Name	Type (DIMENSION spec.)	FORMAT (2nd card)	Default value	Purpose
Input control				
LAST	L		T	=T for last test data set
NINTT	I		0	No. of equally spaced sets of $t_k$ in eq. (1.4) of ref. [1]
IFORMT	I(70)	<b>1X</b> ,70A1 <b>1X</b> ,70A1	(5E15.6)	Input FORMAT for the $t_k$ in eq. (1.4) of ref. [1]
IFORMY	I(70)	<b>1X</b> ,70A1	(5E15.6)	Input FORMAT for the $y_k$ in eq. (1.2) of ref. [1]
IFORMW	I(70)	1X,70A1	(5E15.6)	Input FORMAT for the $W_k$ in section 4.4 of ref. [1]
DOUSIN	L	,	Т	=T to call USERIN (section 4.1 of ref. [1])
User Arrays				
RUSER	R(100)		100 * 0.	REAL USER array (see table 5)
IUSER	I(50)		(see section 4)	INTEGER USER array (see table 5)
LUSER	L(30)		30 * F	LOGICAL USER array (see table 5)
Formation of	linear equations			
NG	I		31	$N_{\rm g}$ in eq. (3.1) of ref. [1]
GMNMX	R(2)		0.,0.	$\overline{GMNMX}(1) = \lambda_1$ , $\overline{GMNMX}(2) = \lambda_{N_g}$ in eq. (3.1) of ref. [1]
IQUAD	I		3	Quadrature method (section 4.2 of ref. [1])
IGRID	I		2	$\lambda$ -grid (section 4.2 of ref. [1])
NLINF	I		0	$N_{\rm L}$ in eq. (3.1) of ref. [1]
Constraints				
DOUSNQ	L		F	=T to call USERNQ for inequality constraints
NONNEG	L		T	= T to constrain $x_j \ge 0$ , $j = 1,, N_g$ , in eq. (3.2) of ref. [1]
NEQ	I		0	$N_{\rm eq}$ in eq. (3.7) of ref. [1]
Least-squares	weights			
IWT	I		1	Weighting scheme (section 4.4 of ref. [1])
NERFIT	I		10	No. of points used for safety margin [2]
Regularizor (s	ection 4.5 of ref.	[1])		
NORDER	I		2	$\geq 0$ for order of regularizor, $< 0$ for call to USERRG
NENDZ	I(2)		2,2	No. of external boundary zeroes
α-grids (see re	ef. [2])			
NQPROG	I(2)		6,6	No. of points in $\alpha$ -grids
RSVMNX	R(2,2)	4E10.3	1.,1.,0.,0.	Range of $\alpha$ -grids
DFMIN	R		2.	Min. reasonable $N_{\rm DF}$ in eq. (3.15) of ref. [1]
ALPST	R(2)		0.,0.	$>0$ to replace grid by the specified $\alpha$ value

Table 4 (continued)

Name	Type (DIMENSION spec.)	FORMAT (2nd card)	Default value	Purpose
Miscellaneous	7			
SIMULA	L		F	=T to simulate data (section 4.6 of ref. [1])
IUNIT	I		-1	>0 to specify device no. for scratch file
ICRIT	I(2)		1,1	Criterion for choosing $\alpha$ [2]
PLEVEL	R(2,2)	4F5.2	4 * 0.5	Probability level for choosing $\alpha$ [2]
Peak constrai	ints (see ref. [2])			
NNSGN	I(2)		0,0	No. of peak-constrained analyses
NSGN	I(4)		4 * 0	No. of monotonic regions
LSIGN	I(4,4)	16I5	16 * 0	Starting locations of monotonic regions
NFLAT	I(4,2)	815	8 * 0	No. of attempts to eliminate plateaus
SRMIN	R		0.01	Relative threshold for defining plateaus
MQPITR	I		35	Max. iterations for a peak-constrained analysis
Output contro	ol			
LINEPG	I		60	Lines per page of printed output
NEWPG1	L		F	=T to start new page at the start of a run
MIOERR	I		5	Abort after MIOERR input diagnostics
PRY	L		T	= T to output $y_k$ and $t_k$ in eq. (1.4) of ref. [1]
PRWT	L		T	= T to output the computed $W_k$ in section 4.4 of ref. [1]
IPRINT	I(2)		4,4	Frequency and spacing of the output of the solutions
ONLY1	L		Т	= F to plot a second curve, computed in USERSX, with the solution
IPLRES	I(2)		2,2	Frequency of plots of the weighted residuals [2]
IPLFIT	I(2)		2,2	Frequency of plots of the fit to the data [2]
IUSROU	I(2)		0,0	Frequency of calls to USEROU for special-purpose output
DOMOM	Ĺ		T	= T to output moments of the solution
MOMNMX	I(2)		-1,3	Min, and max, degrees of moments
MPKMOM	I		5	Max. no. of peaks for which moments are computed
DOCHOS	L		Т	=T to print solution chosen by eq. (3.24) of ref. [1] once again at end

fying NINTT groups of equally spaced  $t_k$  by setting the control variable NINTT>0. Each card contains the characters 'NSTEND' in columns 2–7 and NT, TSTART, TEND as specified in table 3. For each card, NT equally spaced values of  $t_k$  from TSTART to TEND are computed.

Card 5a and card set 5b are used when NINTT  $\leq 0$  to read in the  $t_k$  directly. The 6A1 field in card 5a must contain the characters 'NY', left justified and filled with blanks. NY =  $N_y$  and T(K) =  $t_k$  in eqs. (1.1) and (1.4) of ref. [1].

Card set 6 contains the  $Y(K) = y_k$  in eq. (1.1) of ref. [1]. Card set 7 contains the  $W(K) = W_k$  in section 4.4 of ref. [1]. The relative positions of possible input to USER subprograms is given in ref. [2].

# 4. USER subprograms

Table 5 lists all the USER array elements that are used in the default versions of the USER subprograms. These subprograms can be easily modified for a wide variety of applications. However, in doing so, care should be taken not to use the same USER array element for more than one purpose. The default values of all elements of IUSER are zero except for IUSER(10) = 2 and IUSER(18) = 50. The default versions of USERIN, USERK, USERSI, USEREX and USERWT illustrate applications to the inversion of photon correlation spectroscopy (PCS) and Laplace transform data. In this section we outline their use. Full details are given in ref. [2] and the test data in this

Table 5
Default usage of USER arrays

Array	Subscript	Subprogram	Purpose
RUSER	1	USEREQ	$s(\lambda_1)$
	2	USEREQ	$s(\lambda_{N_{\mathbf{v}}})$
	3	USERSI	Sets noise level
	6	USEREQ	Integral of $s(\lambda)$
	8	USERSX	$\lambda^{R_8} e^{-\lambda} / \Gamma(R_8 + 1)$ is plotted
	10	USERIN	Background (section 4.2)
	14	USEREX	Storage for normalization constant
	15	USERIN	Refractive index
	16	USERIN	Wavelength (in nm) (section 4.2)
	17	USERIN	Scattering angle (in degrees)
	18	USERIN	Proportionality constant or absolute temperature (section 4.2)
	19	USERIN	Viscosity (in centipoise)
	20	USERIN	Storage for magnitude of scattering vector (section 4.2)
	21-23	USERK	See eq. (2) and section 4.2
	24	USERK	Wall thickness (in cm) of hollow spheres (section 4.1)
	25,27,,39	USEREX	Positions of $\delta$ -functions
	26,28,,40	USEREX	Integrated amplitudes of $\delta$ -functions
	41,42,,50	USEREX	Simulated $\beta_i$ in eq. (3.1) of ref. [1]
	$51,52,,50+N_{\rm g}$	USERK	Storage for form factors
IUSER	1	USERRG	Starting location of regularizor in RUSER [2]
	2	USERLF	No. of data points preceding second data set [2]
	3	USERSI	Starting integer for pseudorandom generator
	10	USERK	Selects form of kernel (section 4.2)
	11	USEREX	No. of simulated $\delta$ -functions
	18	USERK	Form factor averaged over $2I_{18+1}$ points (section 4.1)
LUSER	1	USERRG	Internal flag
	3	USERK	=T to use form factors (section 4.1)
	4	USERK	Internal flag

paper provides examples of their use. We abbreviate USER array elements such as RUSER(23) and IUSER(10) by  $R_{23}$  and  $I_{10}$ , etc.

USERK evaluates kernels of the general form

$$F(\lambda_m, t_k) = f_m^2 \lambda_m^{R_{23}} \exp\left(-R_{21} t_k \lambda_m^{R_{22}}\right),\tag{2}$$

which includes Laplace transforms.  $R_{21}$  and  $R_{22}$  must be nonzero. The form factors,  $f_m$ , are described in section 4.1. Convenient options for automatically computing  $R_{21}$  for PCS are described in section 4.2.

# 4.1. Form factors

If  $L_3$  = .FALSE., then all the  $f_m$  = 1. Otherwise, the  $f_m$  are the squared Rayleigh-Debye form fac-

tors for hollow spheres [3], where  $R_{24}$  is the wall thickness in cm. However,  $R_{24} \le 0$  will cause the form factors for solid spheres to be computed. An  $I_{18} > 0$  causes the squared form factor to be averaged over  $2I_{18} + 1$  equally spaced points on the interval centred at the grid point and extending halfway to its nearest neighbors. This is necessary if the form factors rapidly oscillate. The default value of  $I_{18} = 50$  is recommended.

# 4.2. Data preprocessing in USERIN

The default value of DOUSIN = .TRUE., which causes USERIN to be called, is necessary. A non-zero  $R_{10}$  can be used to convert the  $y_k$  input in card set 6 to the PCS normalized first-order corre-

lation function. If  $R_{10} > 0$ , then the  $y_k$  will be replaced by  $+(y_k/R_{10}-1)^{1/2}$ , as appropriate when the second-order correlation function (with background  $R_{10}$ ) is input. When  $R_{10} < 0$ , the input  $y_k$  will be replaced by  $+y_k^{1/2}$ , as appropriate when the normalized second-order correlation function minus 1.0 is input. In both cases if x, the argument in the square root, is negative, then  $-(-x)^{1/2}$  is used (rather than zero) to avoid bias toward positive noise components. If  $R_{10} = 0$ , then the  $y_k$  are not changed. When IWT = 5, USERWT computes statistical weights for the normalized first-order correlation function assuming the second-order correlation function follows Poisson statistics.

When  $R_{15}$ ,  $R_{16}$  and  $R_{17}$  are input as specified in table 5, the magnitude of the scattering vector (in cm<sup>-1</sup>) is automatically stored in

$$R_{20} = (4 \times 10^7 \pi R_{15} / R_{16}) \sin(R_{17} / 2). \tag{3}$$

If  $R_{16} = 0$ , then  $R_{20}$  is not computed and  $R_{21}$  in eq. (2) is set to 1.0, as appropriate for Laplace transforms.

 $I_{10}$  provides convenient options for automatically setting  $R_{21}$ ,  $R_{22}$  and  $R_{23}$  in eq. (2). When  $I_{10}=1$ , CONTIN sets  $R_{23}=1$  and  $R_{21}=R_{18}R_{20}^2$ . This is appropriate when  $s(\lambda)$  is a weight fraction molecular weight distribution and  $R_{18}$  and  $R_{22}$  are input so that the translational diffusion coefficient (in cm<sup>2</sup>/s), is related to the molecular weight,  $\lambda$ , by

$$D = R_{18} \lambda^{R_{22}}. (4)$$

When  $I_{10} = 2$ , CONTIN sets  $R_{23} = 0$ ,  $R_{22} = 1$  and  $R_{21} = R_{20}^2$ . This is appropriate for Laplace transforms (when  $R_{16} = 0$ ) or when  $s(\lambda)$  is a diffusion coefficient distribution. When  $I_{10} = 3$ , CONTIN sets  $R_{23} = 3$ ,  $R_{22} = -1$  and

$$R_{21} = k_{\rm B} R_{18} R_{20}^2 / (0.06 \pi R_{19}), \tag{5}$$

where  $k_B$  is the Boltzmann constant. This is appropriate when  $s(\lambda)$  is a weight fraction radius (in cm) distribution of spheres satisfying the Einstein-Stokes relation,  $R_{15}$ ,  $R_{16}$ ,  $R_{17}$  and  $R_{19}$  are input as specified in table 5, and  $R_{18}$  is the absolute temperature.  $I_{10} = 4$  is for the general case where the user inputs  $R_{21}$ ,  $R_{22}$  and  $R_{23}$  in eq. (2).

#### 4.3. Simulation

The default version of USEREX simulates the case that  $s(\lambda)$  is a sum of  $I_{11}$  Dirac  $\delta$ -functions. This can be very useful in studying the potential resolving power of an experimental method (see section 3.7 of ref. [1]). The positions and integrated amplitudes of the  $\delta$ -functions are specified respectively by  $R_j$  and  $R_{j+1}$  with  $j=25,27,...,23+2I_{11}$ . In addition the sum in eq. (1.2) of ref. [1] can be simulated by specifying the  $\beta_i$  in  $R_j$  with  $j=41,42,...,40+N_L$ .

In USERSI the noisy data are simulated by adding pseudorandom normally distributed noise to the noise-free values,  $Y_k$ , produced in USEREX. The standard deviations of the noise components are  $R_3$  when IWT = 1,  $R_3Y_k^{1/2}$  when IWT = 2,  $R_3Y_k$  when IWT = 3, and  $R_3/W_k^{1/2}$  when IWT = 4. These are consistent with the noise statistics assumed in specifying IWT in section 4.4 of ref. [1]. When IWT = 5, USEREX, USERSI, USERIN and USERWT simulate a second-order PCS correlation function following Poisson statistics,

$$y_k = B[1 + \gamma^2 \{g(t_k) + \Delta_2\}^2],$$
 (6)

where B is the background,  $\gamma$  and  $\Delta_2$  are unknown parameters [4], and  $g(t_k)$  is the first-order correlation function to be analyzed. The user need only input  $R_3 = B^{-1/2}$ ,  $R_{41} = \gamma \Delta_2$ , and  $R_{25}$ ,  $R_{26}$ ,..., so that USEREX simulates  $\gamma g(t_k)$ . This is illustrated in test data set 2 at the end of this paper.

# 5. Output and diagnostics

The output includes the input data; the final values of the control variables; a table of the minimum and maximum elements of each column of A in eq. (3.2) of ref. [1] and the corresponding internal scale factors;  $\rho$ , the internal scale factor for  $\alpha$ ;  $s_j/\rho$ , the scaled singular values in section 3.5 of ref. [1]; and optional plots of the solutions, weighted residuals and fits to the data. When appropriate, there is also optional output of the least squares weights, the scattering form factors, or the noise components in the simulated data.

The test run output at the end of this paper

illustrates the output of the solutions. The first line contains ITITLE from card 1, table 3. The next two lines comprise a table, with ALPHA =  $\alpha/\rho$ , ALPHA/S(1) =  $\alpha/s_1$ , OBJ. FCTN. =  $V(\alpha)$  in eq. (3.10) of ref. [1]; VARIANCE = V(0) in eq. (3.9) of ref. [1]; STD. DEV. =  $\hat{\sigma}$  in eq. (3.22) of ref. [1], DEG FREEDOM =  $N_{\rm DF}$  in eq. (3.15) of ref. [1] and PROB1 TO REJECT = PROBl( $\alpha$ ) in eq. (3.23) of ref. [1]. PROB2 TO REJECT should be ignored [2]. Under the heading ORDINATE are printed (and plotted horizontally) the solution values, i.e., the  $x_i$  in eq. (3.2) of ref. [1] if IQUAD = 1 and the  $s(\lambda_m)$  in eq. (3.1) otherwise. The  $\lambda_m$  are printed under ABSCISSA and the error estimates (see section 3.7 of ref. [1]) are printed under ERROR and shown as error bars with dots above and below the plotted values of the solution. The LIN-EAR COEFFICIENTS are the  $\beta_i$  in eq. (3.1) of

The last two pages of the test run output illustrate how the moments are optionally computed for each peak as well as for the entire curve. The moments,

$$M_{j} = \sum_{m=m_{\min}}^{m=m_{\max}} c_{m} \lambda_{m}^{j} x_{m}, \tag{7}$$

are a quadrature approximation to

$$M_j \approx \int_{\lambda_{\min}}^{\lambda_{\max}} \lambda^j s(\lambda) \, d\lambda$$
 (8)

when IQUAD  $\neq 1$  and eq. (3.1) of ref. [1] is used. The ranges  $\lambda_{max}$  and  $\lambda_{min}$  are given in the table. The table of ratios,  $M_i/M_{i-1}$ , are useful weighted averages of  $\lambda$ . For example, in the first two pages of the test run output,  $s(\lambda)$  is a weight fraction molecular weight distribution and the ratios printed under M(J)/M(J-1) are respectively the number-, weight-, z- and (z + 1)-average molecular weights. In the last two pages,  $s(\lambda)$  is a weight fraction particle radius distribution, and the number- and weight-average radii are therefore given by  $M_{-2}/M_{-3}$  and  $M_1/M_0$ , respectively. The error estimates are especially useful because the weaker scattering and shorter-lived correlations of the smaller particles makes the accuracy of the estimates very uneven.

The last line on pages 1 and 3 of the test run

output summarizes the results of statistical tests on the number of runs (PRUNS) and the autocorrelation with lags 1,...,5 (PUNCOR) in the residuals of the fit to the data [5]. Large values, say  $\gtrsim 0.1$ , are indications of well scattered residuals, probably due to random noise only. Small values, say  $\leq 0.01$ , are a warning that there may be systematic errors or effects not properly accounted for by the model or that  $\alpha$  is too large. The CHOSEN SOLUTIONs on pages 2 and 4 of the test run output are simply being output again at the end of each analysis because their PROB1( $\alpha$ ) was closest to 0.5 (see section 3.6 of ref. [1]). The corresponding PRUNS and PUNCOR, as well as plots of the residuals and fits to the data have already been output previously (not shown).

CONTIN makes extensive tests for error conditions during input and throughout the analysis. There are more than 50 diagnostic messages in a standard format uniquely identifying the error and its location. Explanations and possible remedies for each error are given in detail elsewhere [2].

#### 6. Testing the code

The test run input shown below contains two data sets. The first illustrates the estimation of a molecular weight distribution from real data (of relatively poor quality). The second illustrates the estimation of a particle radius distribution from simulated data corresponding to a mixture of equal weights of particles of radii 50 and 150 nm. Eight of the cards in the second data set are unnecessary since they are the same as in the first data set and values of control variables are passed from one set to the next. The cards are included anyway so that the second data set can be run independently.

CONTIN is meant to be fully portable. However, the discrete grid of  $\alpha$  values used depends on the relative machine precision, which is automatically computed by CONTIN. This quantity is also used to make the regularizor full rank. This means that the CHOSEN SOLUTIONs on pages 2 and 4 of the test run output will only be machine independent to typically one or two times the error estimates. This can best be checked with the values of M(J)/M(J-1). Pages 1 and 3 of the test run

output correspond to the smallest values of  $\alpha$  used, and since these are practically zero, the solutions should be nearly machine independent, except for the values of ALPHA and ALPHA/S(1).

# References

- [1] S.W. Provencher, Comput. Phys. Commun. 27 (1982) 213.
- [2] S.W. Provencher, CONTIN Users Manual, EMBL Techni-

- cal Report DA05, European Molecular Biology Laboratory (1982).
- [3] M. Kerker, The Scattering of Light and Other Electromagnetic Radiation (Academic Press, New York, 1969) p. 482.
- [4] S.W. Provencher, J. Hendrix, L. De Maeyer and N. Paulussen, J. Chem. Phys. 69 (1978) 4273.
- [5] S.W. Provencher, J. Chem. Phys. 64 (1976) 2772.
- [6] C.L. Lawson and R.J. Hanson, Solving Least Squares Problems (Prentice-Hall, Englewood Cliffs, 1974).

# **TEST RUN INPUT**

TEST DATA S	ET 1 (MOLECULAR WEIGHT	DISTRIBUTION)	TEST DATA	SET 2	(SIMULATED RADIUS	DISTRIBUTION)
LAST	-1.		LAST		1.	
GMNMX 1	5.E+2		IWT		5.	
GMNMX 2	5.E+6		NERFIT		0.	
IWT	5.		NINTT		1.	
NERFIT	0.		NLINF		1.	
NINTT	3.		DOUSNQ		1.	
NLINF	1.		IUSER	10	3.	
IFORMY			MOMNMX	1	-3.	
(6F8.6)			RUSER	15	1.43	
DOUSNQ	1.		RUSER	16	488.	
IUSER 10	1.		RUSER	17	60.	
RUSER 15	1.43		RUSER	18	298.	
HUSER 16	488.		RUSER	19	.8937	
RUSER 17	60.		SIMULA		1.	
RUSER 18	1.37E-4		LUSER	3	1.	
RUSER 22	5		IUSER	11	2.	
RUSER 10	-1.		1 US ER	3	30171.	
END			RUSER	3	.0002	
NSTEND 17	5.E-6	85.E-6	RUSER	25	1.5E-5	
NSTEND 16	95.E-6	245.E-6	RUSER	26	- 35	
NSTEND 4	265.E-6	325.E-6	RUSER	27	.5E-5	
.450999 .41	0113 .372522 .340069 .3	310318 .283569	RUSER	28	• 35	
.258853 .23	6028 .216811 .199376 .1	181524 .165491	RUSER	41	.035	
.153746 .13	9687 .128724 .117704 .1	109878 .094114	GMNMX	1	10.E-7	
.080559 .06	8725 .058679 .053363 .0	045275 .039581	GMNMX	2	1000.E-7	
	1586 .027971 .023976 .0		END			
	6487 .017212 .016077 .0		NSTEND	88	8.E-5	704.E-5

# **TEST RUN OUTPUT**

(FOR ALPHA/S(1) = 1.86E-16) PRUNS = 0.5018

TEST DATA SET 1 (MOLECULAR WEIGHT DISTRIBUTION)						PRELIMINARY UNWEIGHTED ANALY				
ALPHA * 2.09E-10	ALPHA/S(1) 1.86E-16	OBJ. FCTN. 2.83804E-04	VARIANCE 2.83804E-04	STD. DEV. 2.889E-03	DEG FREEDOM 3.000	PROB1 TO REJECT 0.000	PROB2 TO REJECT 1.000			
ORDINATE 0.000E+00	8.6D-29 9.24 9.3D-29 1.77 2.9D-28 1.77 2.4D-28 3.15 2.7D-28 3.15 3.6D-28 7.35 3.6D-28 7.35 3.6D-28 1.97 3.4D-29 3.61 3.5D-28 8.27 3.4D-29 3.61 3.5D-28 8.27 3.4D-29 3.61 3.5D-28 1.97 3.4D-29 3.61 3.5D-28 5.83 3.6D-28 5.83 3.	DE+0 2X DE+0 2X UE+0 2X UE+0 2X UE+0 2X UE+0 3X E+0 3X E+0 3X UE+0 3X UE+0 3X UE+0 3X UE+0 3X UE+0 3X UE+0 4X UE+0 5X UE+0 6X	x					х		
PEAK 1 GOES	FROM 5.000E+0	02 TO 5.000E+06	-1 1.9509 X 0 3.4570 X 1 6.1951 X 2 1.1257 X	(10** -11) (10** -6) (10** -1) (10** 5) (10** 10)	PERCENT ERROR 2.8E+00 1.6E+00 2.8E-01 2.0E+00 4.3E+00	M(J)/M(J-1) 1.7720E+05 1.7921E+05 1.8171E+05 1.8475E+05	PERCENT ERROR 4.4E+00 1.9E+00 2.3E+00 6.3E+00	J 0 1 2 3		

PUNCOR = 0.1876 0.7394 0.0205 0.8123 0.5931

CONTIN VERSI	ON 2DP (APR 19	82) ( PCS-1 PAC	KAGE) ++++++		++++++++ CHO:	SEN SOLUTION ++++++	++++++++++++++	++++++
TE	ST DATA SET 1	(MOLECULAR WEIG	HT DISTRIBUTION	)				
AL PHA 1.63E-06	ALPHA/S(1) 1.31E-12	OBJ. FCTN. 3.57888E-05	VARIANCE 3.31443E-05	STD. DEV. 9.893E-04	DEG FREEDOM 3.137	PROB1 TO REJECT 0.785	PROB2 TO REJECT 1.000	
ORDINATE 0.000E+00 1.000E+00 0.000E+00	3.6D-29 6.80 3.1D-29 9.24 2.0D-29 1.71 9.1D-30 2.35 1.7D-29 3.15 1.7D-29 4.29 2.7D-29 5.33 3.0D-29 1.08 7.8D-29 1.08 7.8D-29 1.90 1.8D-28 2.71 2.3D-28 3.60 1.4D-12 6.80 1.7D-12 9.25 1.2D-12 1.71 8.4D-13 2.32 4.6D-13 1.26 1.2D-12 1.71 8.4D-13 2.32 4.6D-13 7.92 1.2D-26 5.83 4.3D-29 1.46 8.2D-29 1.46 8.2D-29 1.99 5.3D-29 2.71 8.6D-30 3.68 1.4D-29 5.00	E+02X E+02X E+03X E+03X E+03X E+03X E+03X E+03X E+03X E+043X E+043X E+044 E+044 E+044 E+045 E+045 E+05 E+05 E+05 E+05 E+05 E+05 E+05X E+05X E+06X E+06X E+06X E+06X E+06X	X X		x	 x	x	x
		2 TO 5.000E+06	J MON -1 2.2100 1 0 3.4989 1 1 6.2598 1 2 1.2516 1	MENT(J) ( (10** ~11) ( (10** ~6) ( (10** -1) ( (10** 5) ( (10** 10)	PERCENT ERROR 3.1E+00 1.1E+00 2.8E-01 1.8E+00 3.9E+00	M(J)/M(J-1) 1.5832E+05 1.7891E+05 1.9994E+05 2.2070E+05	PERCENT ERROR  4.1E+00 1.4E+00 2.0E+00 5.7E+00	J 0 1 2 3

PRELIMINARY UNWEIGHTED ANALYSIS

TEST DATA SET 2 (SIMULATED RADIUS DISTRIBUTION)

DEG FREEDOM PROB2 TO REJECT ALPHA ALPHA/S(1) OBJ. FCTN. VARIANCE STD. DEV. PROB1 TO REJECT 1.160E-03 0.000 \* 1.35E-10 5,000 1.86E-16 1.11779E-04 1.11779E-04 ORDINATE ERROR ABSCISSA 1.00E-06X 0.000E+00 7.9D+02 0.000E+00 8.9D+03 1.17E-06X 0.0008+00 3.3D+04 1.36E-06X 0.000E+00 9.2D+03 1.0D+04 2.8D+04 0.000E+00 1.85E-06X 2.15E-06X 0.000E+00 0.000E+00 2.6D+04 2.51E-06X 0.000E+00 4.6D+04 2.93E-06X 0.000£+00 5.1D+04 3.41E-06X 0.000E+00 3.2D+04 3.98E-06X χ,..... 1.892E+21 1.3D+21 4.64E-06 1.056E+20 4.5D+20 5.41E-06....X..... 0.000E+00 2.40+05 6.31E-06X 1.4D+05 7.36E-06X 0.000E+00 0.000E+00 8.6D+04 8,58E-06X 0.000E+00 1.8D+05 1.00E-05X 0.000E+00 1.17E-05X 7.0D+04 ..X. 1.521E+20 3.6D+19 1.36E-05 1.58E-05 ...X... 5.6D+19 3.839E+20 1.85E-05X 0.000E+00 1.6D+05 2.15E-05X 2.51E-05X 0.000E+00  $1.10 \pm 0.5$ 0.000E+00 8.9D+04 0.000E+00 1.3D+05 2.93E-05X 1.8D+03 3.41E-05X 0.0008+00 0.000E+00 1.2D+04 3.98E-05X 0.000E+00 2.30+04 4.64E-05X 0.000E+00 1.5D+04 5.41E-05X 0.000E+00 5.10+04 6.31E-05X 7.36E-05X 1.2D+04 0.000E+00 0.000£+00 2.1D+04 8.58E-05X 0.000£+00 1.5D+04 1.00E-04X LINEAH COEFFICIENTS = 3.4824E-02 +- 6.9D-04 MOMENT(J) 9.7277 X (10\*\* 30) 4.5720 X (10\*\* 25) M(J)/M(J-1) PERCENT ERROR PEAK 1 GOES FROM 1.000E-06 TO 1.166E-05 PERCENT ERROR 3.0E+01 2.4E+01 4.7000E-06 5.4E+01 2.1529 X (10\*\* 20) 1.0159 X (10\*\* 15) 4.7088E-06 4.2E+01 1.8E+01 4.7189E-06 2.9E+01 1.1E+01 4.8056 X (10\*\* 9) 2.2793 X (10\*\* 4) 1.0843 X (10\*\* -1) 4.4E+00 4.7302E-06 1.5E+01 4.7430E-06 1.2E+01 7.9E+00 1.7E+01 4.7573E-06 2.5E+01 PERCENT ERROR M(J)/M(J-1) PERCENT ERROR PEAK 2 GOES FROM 1.359E-05 TO 1.000E-04 MOMENT(J) 3.2492 X (10\*\* 29) 4.7695 X (10\*\* 24) 5.3E+00 3.8E+00 1.4679E-05 9.1E+00 -2 7.0425 X (10\*\* 2.4E+00 1.4766E-05 6.2E+00 19) 1.0459 X (10\*\* 1.5622 X (10\*\* 2.3461 X (10\*\* 3.5419 X (10\*\* 0 0 15) 1.0E+00 1.4852E-05 3.4E+00 10) 6.3E-01 1.4936E-05 1.6E+00 1.5018E-05 5) 1.9E+00 2.5E+00 1,5097E-05 5.0E+00 3 3.1E+00 PERCENT ERROR M(J)/M(J-1)PERCENT ERROR MOMENTS OF ENTIRE SOLUTION MOMENT(J) 1.0053 X (10\*\* 5.0490 X (10\*\* 2.9E+01 5.0225E-06 5.1E+01 25) 2.2E+01 -2 2.8571 X (10\*\* 1.4E+01 5.6589E-06 3.6E+01 20) -1 2.0618 X (10\*\* 2.0428 X (10\*\* 2.5741 X (10\*\* 0 15) 5.5E+00 7.2165E-06 1.9E+01 0 10) 9.9074E-06 1.2E+00 5) 1.8E+00 1.2601E-05 3.0E+00 3.6504 X (10\*\* 1.4181E-05 4.9E+00 3.1E+00 PUNCOR = 0.0865 0.2558 0.1253 0.6963 0.9809 (FOR ALPHA/S(1) = 1.86E-16) PRUNS = 0.2973

		(SIMULATED RADI			, , , <del>, , , , , , , , , , , , , , , , </del>	+++++++++ CNOS	EN SOLUTION +++++	····	++++++++
AL PHA	ALPHA/S(1)	OBJ. FCTN.			TD. DEV.	DEG FREEDOM	PROB1 TO REJECT	PROB2 TO REJECT	т
2.58E-05	2.77E-11	4.15143E-06	4.0651	5E-06 2	.213E-04		0.349	1.000	
ORDINATE	ERHUR ABSC	A22T							
	6.8D+02 1.00								
	2.20+02 1.17								
	1.10+03 1.36								
	7.1D+02 1.58								
0.000E+00	9.10+02 1.85	E-06X							
1.921E+19	2.0D+19 2.15	E-06x							
7.353£+19	4.0D+19 2.51	E-06 .		X					
1.553E+20						X			
	4,6D+19 3.41							x	•
	3.3D+19 3.98								X
	1.9D+19 4.64								X
	1.6D+19 5.41							X	
1.490E+20	2.0D+19 6.31 1.7D+19 7.36	E-05	v			X			
	6.3D+03 8.58		X	•					
0.000E+00									
6.822E+19				x.,					
	8.2D+18 1.36					x			
	5.5D+18 1.58					x			
	7.8D+18 1.85				х	****			
	1.3D+19 2.15		.x.,,						
	9.3D+18 2.51								
	3.9D+03 2.93								
	2.0D+03 3.41								
0.000E+00	3.4D+03 3.98								
0.000E+00	1.7D+03 4.64								
	2.1D+03 5.411								
0.000E+00 0.000E+00	1.5D+03 6.311								
	3.0D+03 7.36i 1.0D+03 8.58i								
	4.8D+02 1.00								
LINEAR COEFF	ICIENTS = 3.2	2474E-02 +- 8.2	2D-04						
PEAK 1 GOES	FROM 1.000E-0	6 TO 1.000E-05	J	MOMENT (J	1)	PERCENT ERROR	M(J)/M(J-1)	PERCENT ERROR	J
				1.5770 X (10≝		2.1E+01			
				5.9417 X (10≝		1.5E+01	3.7676E-06	3.5E+01	-2
				2.4172 X (10		9.2E+00	4.0682E-06	2.4E+01	-1
				1.0604 X (10*		4.5E+00	4.3868E-06	1.4E+01	0
				1.9951 X (10		1.6E+00	4.7107E-06	6.1E+00	1
				2.5115 X·(10≝ 1.3383 X (10≝		4,4E+00	5.0278E-06	6.1E+00	2
			3	1.5505 X (10-	-1)	8.0E+00	5.3287E-06	1.2E+01	3
PEAK 2 GOES 1	FROM 1.166E-05	5 TO 1.000E-04	J	MOMENT (J	1)	PERCENT ERROR	M(J)/M(J-1)	PERCENT ERROR	J
				3.5338 X (10		1.7E+00	11(0)/11(0-1)	I ENCENT ENTON	J
				.2366 X (10♥		1.6E+00	1.4819E-05	3.3E+00	<b>-</b> 2
				7.9767 X (10*		3.2E+00	1.5233E-05	4.8E+00	-2 -1
			0 1	1.2513 X (10≝	<b>*</b> 15)	5.8E+00	1.5687E-05	9.0E+00	0
				2.0240 X (10≝	<b>*</b> 10)	9.3E+00	1.6175E-05	1.5E+01	1
				3.3777 X (10≝		1.4E+01	1.6689E-05	2.3E+01	2
			3 5	3.8158 X (10 <b>*</b>	• 0)	1.9E+01	1.7218E-05	3.3E+01	3
	MOMENTS OF	ENTIRE SOLUTION	J	MOMENT(J	)	PERCENT ERROR	M(J)/M(J-1)	PERCENT ERROR	1
				.6124 X (10*		2.0E+01	n(0 // n(0=1 /	I EUCENI EUNON	J
				.4653 X (10*		1.3E+01	4.0098E-06	3.4E+01	-2
				3.2149 X (10*		6.9E+00	4.9725E-06	2.0E+01	-2 -1
				.3117 X (10*		3.8E+00	7.1905E-06	1.1E+01	0
				2.5235 X (10*		7.4E+00	1.0916E-05	1.1E+01	1
				3.6289 X (10 <b>*</b>		1.3E+01	1.4381E-05	2.0E+01	2
			3 5	5.9497 X (10 <b>*</b>	* 0)	1,9E+01	1.6395E-05	3.2E+01	3