IBM **Application Program** H20-0544-0

System/360 Scientific Subroutine Package (PL/I)

Application Description

The System/360 Scientific Subroutine Package (SSP) (PL/I) is a collection of mathematical and statistical subroutines (or procedures) written in the PL/I language. It provides the PL/I user with most of the basic capabilities in earlier FORTRAN versions of SSP/360. It also has the same basic characteristics as the FORTRAN versions, in that it consists of input/output-free computational building blocks, written completely in PL/I, which may be combined with a user's input, output, or computational routines as needed. The package may be applied to the solution of many problems in industry, science, and engineering.

This Application Description presents an introduction to the program, a list of the capabilities of the package, rules of usage, machine configuration, programming systems, and a list of reference material.

CONTENTS

Note: This programming package is being developed with the cooperation and assistance of IBM Germany and IBM France.

First Edition

Significant changes or additions to the specifications contained in this publication will be reported in subsequent revisions or Technical Newsletters.

Copies of this and other IBM publications can be obtained through IBM branch offices. Address comments concerning the contents of this publication to IBM, Technical Publications Department, 112 East Post Road, White Plains, N. Y. 10601

INTRODUCTION

The Scientific Subroutine Package/360 (PL/I) is a set of basic computational subroutines intended to help the user develop his own PL/I subroutine library. The PL/I version of SSP/360 is similar to the FORTRAN version, from which it is derived, and it includes most of the capabilities of the FORTRAN version.

Areas of Application

Individual subroutines or a combination of them can be used for the following general areas:

<u>Mathematics</u>

- Matrix operations
  - Elementary
  - Linear equations
  - Eigenvalues
- Polynomial operations
  - Orthogonal polynomials
  - Polynomial economization
  - Polynomial roots
- Numerical Quadrature
  - Tabulated functions
  - Nontabulated functions
- Numerical differentiation
  - Tabulated functions
  - Nontabulated functions
- Interpolation of tabulated functions
- Approximation of tabulated functions
- Smoothing of tabulated functions
- Roots and extrema
- Ordinary differential equations
- Special functions

<u>Statistics</u>

- Data screening and analysis
- Elementary statistics
- Correlation and regression analysis
  - Correlation
  - Multiple linear regression
  - Stepwise multiple regression
  - Canonical correlation
- Analysis of variance
- Discriminant analysis
- Principal components analysis
- Nonparametric statistics
- Distribution functions

Characteristics

Some of the characteristics of SSP/360 (PL/I) are as follows:

- All subroutines are free of input/output statements.
- All subroutines are written in OS/360 PL/I (F).
- Most of the subroutines provide a double-precision option.
- The use of certain subroutines (or groups of them) is illustrated in the program documentation by sample main programs with input/output.
- All subroutines are documented uniformly.

An example of a sample main program that uses several of the subroutines is the statistical function called principal components analysis. It uses five separate subroutine capabilities, as follows:

- Computation of means, standard deviations and correlation matrix
- Computation of eigenvalues and eigenvectors of the correlation matrix
- Selection of eigenvalues
- Computation of factor matrix
- Varimax rotation of the factor matrix

This is one of the sample main programs to be included in the program documentation.

FUNCTIONAL CAPABILITIES OF SSP/360 (PL/I)

The functions performed by SSP/360 (PL/I) are listed below, grouped into related functional areas. In some cases, several subroutines are required to perform the function described, as in the case of principal components analysis, referred to above.

Mathematics

MATRIX OPERATIONS

Elementary Operations

Matrix product: Standard matrix multiplication of two matrices which may be either general or symmetric.

Matrix permutation: Permuting the rows or columns according to a given permutation vector.

Storage conversion: Converting storage mode from compressed symmetric to general and vice versa.

Linear Equations and Related Topics

Triangular factorization: A given nonsingular matrix (general, symmetric, or band) is factored into the product of an upper and lower triangular matrix.

Division by a matrix: A general matrix B is divided by a matrix A (symmetric, band, or general), where A has the triangular factorization A=LU. This factorization may be obtained by another subroutine. A particular application is the solution of a set of simultaneous equations, AX=B, which is done in two steps: A = LU, then $X = U^{-1} L^{-1} B$.

Determination of rank and dependencies: The rank of a given general rectangular matrix is determined using the Gaussian elimination technique. The procedure also gives the linear dependencies among rows and columns.

Inversion of a matrix: The inverse of a matrix is calculated from its triangular factorization. Separate procedures are provided for general and symmetric matrices. There is an additional procedure for the inverse of a general matrix by the standard Gauss-Jordan method with complete pivoting.

Solution of overdetermined systems of equations: The least squares solution is obtained using Householder's unitary transformations. This procedure may also be applied to calculate the solution of a system of equations.

Eigenvalues and Related Topics

Transformation to almost-triangular (Hessenberg) form: A general matrix is transformed to almost-triangular form using either unitary transformations or an elimination technique. A third procedure reduces a symmetric matrix to symmetric tridiagonal form.

Eigenvalues: The eigenvalues of an almost-triangular matrix are obtained using Francis' QR method. A modification of the QR method, which is due to Kaiser and Ortega, is used in the case of a symmetric tridiagonal matrix.

Bounds on eigenvalues: Bounds on the eigenvalues of a symmetric matrix are calculated using Laguerre's transformation of the point at infinity.

Eigenvectors: The eigenvector corresponding to a given eigenvalue of an almost-triangular or a tridiagonal matrix is obtained using inverse iteration.

Back transformation of eigenvectors: A transformation is applied to eigenvectors of an almost-triangular or a tridiagonal matrix, which gives the eigenvectors of the associated original matrix (see "Transformation to almost-triangular form" above).

Jacobi's method: The eigenvalues and eigenvectors of a symmetric matrix are calculated simultaneously by means of Jacobi's method.

POLYNOMIAL OPERATIONS

Values of orthogonal polynomials: The values of the first n orthogonal polynomials of Chebyshev, Legendre, Laguerre, or Hermite are calculated from the three-term recurrence relation.

Value of series expansion in orthogonal polynomials: The value of a series expansion in orthogonal polynomials of Chebyshev, Legendre, Laguerre, or Hermite is calculated using a backward iteration scheme.

Transformation of an orthogonal polynomial expansion: A given series expansion in orthogonal polynomials of Chebyshev, Legendre, Laguerre, or Hermite is transformed to an ordinary polynomial. A linear transformation of the range is performed simultaneously.

Polynomial economization: A given polynomial is replaced by a polynomial of lower degree within a specified error bound by means of telescoping. The range may be (-a, a) or (0, a). This procedure also gives the transformation of an ordinary polynomial to an expansion in Chebyshev polynomials.

Roots of polynomials: The roots of a polynomial are calculated using a combination of Newton's, Bairstow's and Nickel's methods.

NUMERICAL QUADRATURE

Quadrature of Tabulated Functions

Trapezoidal rule: A table of values of an integral is calculated using the trapezoidal rule.

Simpson's rule: A table of values of an integral is calculated using Simpson's rule combined with Newton's 3/8 rule.

Hermite formulas: A table of values of an integral is calculated from a given function and its derivatives using Hermite formulas of first or second order.

Quadrature of Nontabulated Functions

Romberg's method: The value of the integral is calculated by Romberg's extrapolation method on successive approximations obtained by the trapezoidal rule with refined stepsize.

Gaussian quadrature: The value of the integral $\int_a^b f(x)dx$ is obtained from a weighted sum of function values using Gaussian quadrature formulas.

Gauss-Laguerre quadrature: The value of the integral $\int_0^{\infty} e^{-x} f(x)dx$ is obtained from a weighted sum of function values using Gauss-Laguerre quadrature formulas.

Gauss-Hermite quadrature: The value of the integral $\int_{-\infty}^{+\infty} e^{-x^2} f(x)dx$ is obtained from a weighted sum of function values using Gauss-Hermite quadrature formulas.

Gauss-Laguerre/Hermite quadrature: The value of the integral $\int_{-\infty}^{+\infty} e^{-x^2} f(x^2)dx = \int_0^{+\infty} e^{-x} \frac{fx}{\sqrt{x}}dx$ is obtained from a weighted sum of function values using Gauss-Laguerre/Hermite quadrature formulas.

NUMERICAL DIFFERENTIATION

Differentiation of Tabulated Functions

Three-point formulas: The table of the derivative values is obtained from the Lagrange interpolation polynomial passing through three consecutive points.

Five-point formulas: The table of the derivative values is obtained from the Lagrange interpolation polynomial passing through five consecutive points.

Differentiation of Nontabulated Functions

Extrapolation on central differences: The value of the derivative is computed using an extrapolation technique on successive central divided differences with decreasing stepsize.

Extrapolation on one-sided differences: The value of the derivative is computed using an extrapolation technique on successive one-sided divided differences with decreasing stepsize.

INTERPOLATION OF TABULATED FUNCTIONS

Aitken-Lagrange interpolation: Interpolation in the given table of function values is done using Aitken's scheme of Lagrange interpolation.

Aitken-Hermite interpolation: Interpolation in the given table of both function and derivative values is done using Aitken's scheme of Hermitian interpolation.

Continued fraction interpolation: Interpolation in the given table of function values is based on continued fraction interpolation utilizing inverted divided differences.

APPROXIMATION OF TABULATED FUNCTIONS

Fourier Analysis: The "fast" Fourier transform performs Fourier analysis and synthesis for the one-dimensional or multidimensional case, for real or complex data.

Least squares fit in Chebyshev polynomials: The normal equations for the linear least squares approximation in terms of Chebyshev polynomials are established from a given table of function values.

Least squares fit in general basic functions: The normal equations for the linear least squares approximation in terms of the specified basic functions are established from the given tables of function values and of fundamental functions.

Solution of normal equations: Using Choleski's method, all fits up to a specified order or accuracy are calculated from previously established normal equations. Optionally, the calculation of the low-order fits may be suppressed.

SMOOTHING OF TABULATED FUNCTIONS

Linear fit to three points: A smoothed table of function values is calculated from a local least-squares polynomial fit of degree 1 on successive groups of three points.

Linear fit to five points: A smoothed table of function values is calculated from a local least squares polynomial fit of degree 1 on successive groups of five points.

Third-degree fit to five points: A smoothed table of function values is calculated from a local least-squares polynomial fit of degree 3 on successive groups of five points.

Exponential smoothing: The triple exponentially smoothed series of a given series is found.

ROOTS AND EXTREMA OF FUNCTIONS

Root of a function: An estimate of the root is refined using either (1) linear (secant method), (2) quadratic (Mueller's method), or (3) hyperbolic (Halley's method) interpolation.

Root of a function — utilizing derivative: An estimate of the root is refined using either (1) linear (Newton's method), (2) inverse quadratic, or (3) hyperbolic interpolation. The latter two use an estimation of the second derivative.

Local minimum of a function: An estimate of the unconstrained minimum of a function of several variables whose values and gradients are given, is refined using the method of Davidon.

SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS

Rational extrapolation on midpoint rule: One integration step of suggested length is performed for a system of first-order differential equations. The rational extrapolation technique employed (which is applied to successive approximations by means of the midpoint rule) automatically adjusts the stepsize according to accuracy requirements.

SPECIAL MATHEMATICAL FUNCTIONS

Complete elliptic integrals: Calculation of complete elliptic integrals of the first and second kind is based on the process of the arithmetic-geometric mean.

Incomplete elliptic integrals: Calculation of incomplete elliptic integrals of the first and second kind is based on the process of the arithmetic-geometric mean combined with Landen's transformation.

Jacobi elliptic functions: Calculation of the three elliptic functions of Jacobi is based on the process of the arithmetic-geometric mean together with Gauss' transformation and with Jacobi's modulus transformation (if required).

Log of the gamma function: Calculation of the double-precision natural logarithm of the gamma function of a given double-precision argument.

Statistics

DATA SCREENING AND ANALYSIS

Subset: For a given series of observations on a fixed number of variables, this subroutine selects a subset vector indicating which observations satisfy given logical or arithmetic relationships involving the variables.

Tally: For a given series of observations, the totals, means, standard deviations, minima, and maxima are obtained for each variable.

Bounds: For a given series of observations, a count is made of the number of observations under, between, and over two given bounds for each variable.

One-variable tabulation: Given a series of observations on a given variable, a tabulation is made of the frequency and relative frequency over assigned class intervals. Some other basic statistics on the variable are also calculated.

Two-variable tabulation: Given a series of observations, a two-way classification is found for any two of the variables. Frequency and relative frequency over given class intervals are found, as well as other basic statistics.

ELEMENTARY STATISTICS

Moments: For grouped data on equal class intervals, the mean and several higher central moments are found.

t-Tests: Several t-statistics on the means of populations under various hypotheses are computed.

CORRELATION AND REGRESSION ANALYSIS

Correlation: Means, standard deviations, sums of cross products of deviations and product-moment correlation coefficients are computed.

Regression: Multiple linear regression analysis is performed for a set of independent variables and a dependent variable. Selection of various sets of independent variables and of a dependent variable can be made as desired.

Stepwise multiple regression: Stepwise multiple regression is performed for a set of independent variables and a dependent variable. Selection of various sets of independent variables and designation of a dependent variable can be made as many times as desired.

Canonical correlation: An analysis of the interrelations between two sets of variables measured on the same subjects is performed. The canonical correlation, which gives the maximum correlation between linear functions of the two sets of variables, is calculated.

ANALYSIS OF VARIANCE (FACTORIAL DESIGN)

Analysis of variance: An analysis of variance is performed for a factorial design by use of three special operators suggested by H. O. Hartley. The analysis of many other designs can be derived by first reducing them to factorial designs and then pooling certain components of the analysis of variance table.

DISCRIMINANT ANALYSIS

Discriminant analysis: A set of linear functions is calculated from data on many groups for the purpose of classifying new individuals into one of the groups. The classification of an individual into a group is performed by evaluating each of the calculated linear functions first and then finding the group for which the associated probability value is the largest.

PRINCIPAL COMPONENT ANALYSIS

Principal component analysis: A principal component solution and the varimax rotation of the factor matrix are performed. Principal component analysis is used to determine the minimum number of independent dimensions needed to account for most of the variance in the original set of variables.

NONPARAMETRIC STATISTICS

The following statistical computations are included: chi square, Mann-Whitney U-test, Friedman two-way analysis of variance statistic, Cochran Q-test, Spearman rank correlation coefficient, Kendall rank correlation coefficient, Kendall's coefficient of concordance, Kruskal-Wallis H-test, Kolmogorov-Smirnov one and two sample tests and limiting distribution values.

DISTRIBUTION FUNCTIONS

Normal distribution function: $y=P(x) = \text{Prob}(X \leq x)$, where X is a random variable distributed normally with mean zero and variance one, is computed.

Beta distribution function: $P=I_x(m, n) = \text{Prob}(X \leq x)$, where X is a random variable following the beta distribution with degrees of freedom (continuous parameters) m and n, is computed.

Chi-square distribution function: $P=P(x) = \text{Prob}(X \leq x)$, where X is a random variable following the chi-square distribution with continuous parameter m, is computed.

Inverse of normal distribution function: $x=P^{-1}(y)$ such that $y=P(x) = \text{Prob}(X \leq x)$, where X is a random variable distributed normally with mean zero and variance one, is computed.

SUBROUTINE USAGE

All subroutines in the Scientific Subroutine Package (PL/I) are invoked by means of the standard PL/I procedure CALL statement. These subroutines are mainly computational in nature and do not contain any references to input/output devices. The user must therefore furnish, as part of his program, whatever input/output and other operations are necessary for the total solution of his problem.

REQUIRED SYSTEMS

Programming Systems

The subroutines are written in the PL/I language, using the facilities provided by the PL/I (F) compiler, which functions under Operating System/360.

Machine Configuration

A minimum requirement is a System/360 suitable for the OS/360 PL/I (F) compiler. The machine configuration required for any given problem depends upon the number of subroutines used, the size of the compiled subroutines, the size of the compiled main program, the size of the control program, and the data storage requirements.

PRECISION

The accuracy of the computations in many of the SSP subroutines is highly dependent upon the number of significant digits available for arithmetic operations. Matrix inversion, integration, and many of the statistical subroutines fall into this category. The user may, therefore, wish to use double-precision versions of these subroutines. Most of the SSP/360 (PL/I) subroutines provide a double-precision option.

REFERENCE MATERIAL

System/360 Scientific Subroutine Package (360A-CM-03X) Version III Programmer's Manual (H20-0205)
IBM System/360 Operating System PL/I (F) Reference Manual (C28-8201)
IBM System/360 Operating System PL/I (F) Programmer's Guide (C28-6594)

Printed in U.S.A. H20-0544-0

IBM

International Business Machines Corporation
Data Processing Division
112 East Post Road, White Plains, N.Y. 10601
(USA Only)

IBM World Trade Corporation
821 United Nations Plaza, New York, New York 10017
(International)