NASA TECHNICAL NOTE

NASA TN D-2830

NASA TN D-2830

AN APPLICATION OF GENERALIZED MATRIX INVERSION TO SEQUENTIAL LEAST SQUARES PARAMETER ESTIMATION

by Henry P. Decell, Jr.

Manned Spacecraft Center
Houston, Texas

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MAY 1965

NASA TN D-2830

AN APPLICATION OF GENERALIZED MATRIX INVERSION TO SEQUENTIAL LEAST SQUARES PARAMETER ESTIMATION

By Henry P. Decell, Jr.

Manned Spacecraft Center
Houston, Texas

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – Price $1.00

AN APPLICATION OF GENERALIZED MATRIX INVERSION TO SEQUENTIAL LEAST SQUARES PARAMETER ESTIMATION

By Henry P. Decell, Jr.
Manned Spacecraft Center

SUMMARY

The theory of generalized matrix inversion is utilized in formulating a recursive algorithm for least squares parameter estimation. This algorithm allows the parameter estimation to begin after the first "observation" has been made and affords a means of computing the nth parameter state from the (n-1)st "parameter state" and the nth "observation." The problems associated with singular matrices encountered in iterative least squares procedures do not affect the algorithm.

INTRODUCTION

In the theory of linear least squares parameter estimation, the matrix equation $Ax = b$ is encountered, where A is an $n \times k$ matrix, x is a $k \times 1$ "parameter" state vector, and b is a $n \times 1$ "observation" vector. There is usually no vector x that will satisfy this matrix equation so that in some sense a "best solution" is to be found. In the least squares theory the "best solution" is defined to be that $k \times 1$ vector $\hat{x}$ such that $f(\hat{x}) = (A\hat{x} - b)^T(A\hat{x} - b)$ is minimum. It is well known that if $\hat{x}$ minimizes $f(x)$, then $\hat{x}$ must satisfy the normal equations $A^TA\hat{x} = A^Tb$. If A^TA is nonsingular and $f(\mathbf{x})$ has a minimum then, indeed, that minimum is attained at $\hat{x} = \left(A^TA\right)^{-1}A^Tb$. It will later be shown that, in any case, the matrix equation $A^TAx = A^Tb$ always has some solution and, in fact, may have infinitely many solutions. In case A^TA is singular, a particular solution that has physical meaning can be chosen.

Nonlinear parameter estimation problems are usually handled by linear approximations of the actual parameter state in a neighborhood of a nominal parameter state. The resulting equations are of the same general form $Ax = b$; however, in this case x denotes the deviation from the nominal state, and b denotes the deviation in "observed" and "computed" values. The solution to this equation requires iterative procedures and involves problems of singular matrices. The recursive algorithm to be developed will not be affected by these singularity problems.

The recursive algorithm differs from a similar algorithm developed by Gainer (ref. 1) in that no matrix inversion is required as a result of waiting for a sufficient number of observations to accumulate. The estimation procedure may begin after the first observation is made. The parameter state can be estimated at a fixed epoch in time (in dynamic systems) and, hence, is useful in trajectory calculations as well as in guidance procedures. There is no need for matrix inversion (and associated storage requirements) in onboard calculations.

SYMBOLS

Capital letters	matrices
Lower case letters	vectors (unless otherwise stated)
Greek letters	scalars
A^T	transpose of A
A^{-1}	inverse of A
A^+	generalized inverse of A
diag $a_1, \ldots, a_n$	diagonal matrix
Z	zero matrix
θ	zero vector
$\|\| \ \|\|$	euclidean norm

THE GENERALIZED INVERSE

A. Bjerhammar (ref. 2), E. H. Moore (ref. 3), and R. Penrose (ref. 4) independently generalized the concept of matrix inversion. The generalized inverse of a singular and nonsquare matrix possesses properties which make it a central concept in matrix theory. Only real matrices are considered in the definitions and theorems to follow. However, for complex matrices, the definitions and theorems are identical if the word "transpose" is replaced with "conjugate transpose." The following fundamental theorem due to R. Penrose (ref. 4) is stated without proof.

THEOREM I. The four equations

$$AXA = A \tag{1}$$

$$XAX = X \tag{2}$$

$$(XA)^T = XA \tag{3}$$

$$(AX)^T = AX \tag{4}$$

have a unique solution X for each real matrix A.

The solution X in THEOREM I is denoted as $X = A^+$ and called the generalized inverse of A. It is easy to see that the defining equations for A^+ imply that AA^+ and A^+A are, respectively, orthogonal projection operators on the range spaces of A and A^+. For the sake of completeness the next theorems will give some well known properties of the generalized inverse.

THEOREM II. Let A be an arbitrary real matrix. Then, for scaler $\lambda \neq 0$ and unitary U and V

$$A^+\left(A^+\right)^T A^T = A^+ = A^T\left(A^+\right)^T A^+ \tag{5}$$

$$A^+AA^T = A^T = A^TAA^+ \tag{6}$$

$$\left(A^+\right)^+ = A \tag{7}$$

$$\left(A^T\right)^+ = \left(A^+\right)^T \tag{8}$$

$$A^+ = A^{-1} \quad \text{for nonsingular } A \tag{9}$$

$$(\lambda A)^+ = \frac{1}{\lambda}A^+ \tag{10}$$

$$\left(A^TA\right)^+ = A^+\left(A^+\right)^T \tag{11}$$

$$(UAV)^+ = V^{-1}A^+U^{-1} \tag{12}$$

$$\left.\begin{array}{l} A = \sum A_i \quad \text{and} \quad A_i^TA_j = Z \\ A_j^TA_i = Z \quad \text{for} \quad i \neq j \\ \text{imply} \\ \qquad A^+ = \sum A_i^+ \end{array}\right\} \tag{13}$$

If A is normal (i.e. $A^TA = AA^T$)
then, $A^+A = AA^+$ and $(A^n)^+ = (A^+)^n$ (14)

A, A^TA, A^+ and A^+A all have rank equal to trace A^+A (15)

$$A^+ = \left(A^TA\right)^+A^T \tag{16}$$

Note that equation (16) reduces the problem of computing A^+ to that of computing the generalized inverse of a symmetric matrix A^TA. Moreover, such a matrix can always be diagonalized by a unitary transformation, that is,

$$D = U\left(A^TA\right)V = \mathrm{diag}\left(a_1, \ldots, a_n\right)$$

Now equation (12) implies that

$$\left(A^TA\right)^+ = VD^+U = V\ \mathrm{diag}\left(\frac{1}{a_1}, \ldots, \frac{1}{a_n}\right)U$$

It is tacitly assumed that if $a_i = 0$, the corresponding term in $\mathrm{diag}\left(\frac{1}{a_1}, \ldots, \frac{1}{a_n}\right)$ is zero. It is not usually an easy task to determine the unitary transformations U and V. Methods for computing the generalized inverse have been given by various authors (refs. 2, 5, 6, 7, and 8).

The following is a theorem of major importance, characterizing all solutions of the matrix equations $AXB = C$ which have some solution X.

THEOREM III. For the matrix equation $AXB = C$ to have a solution, a necessary and sufficient condition is

$$AA^+CB^+B = C$$

in which case, the general solution is

$$X = A^+CB^+ + Y - A^+AYBB^+$$

where Y is arbitrary to within the limits of being consistent with dimension in the indicated multiplications (ref. 4).

Proof: Suppose X satisfies $AXB = C$. Then,

$$C = AXB = AA^+AXBB^+B = AA^+CB^+B$$

Conversely, if $C = AA^+CB^+B$, then A^+CB^+ is a particular solution. Clearly, for the general solution $AXB = 0$ must be solved. Any expression of the form

$$X = Y - A^+AYBB^+$$

is such a solution. Moreover, if $AXB = 0$, then,

$$X = X - A^+AXBB^+$$

The only properties required of A^+ and B^+ in the theorem are $AA^+A = A$ and $BB^+B = B$.

Corollary A. The general solution to the vector equation

$$Px = c$$

is

$$x = P^+c + \left(I - P^+P\right)y$$

where y is arbitrary, provided a solution exists.

Corollary B. A necessary and sufficient condition for the equations

$$AX = C$$

and

$$XB = D$$

to have a common solution is that each have a solution and $AD = CB$ (ref. 4).

Proof: If $AX = C$ and $XB = D$ have a common solution, then clearly each has a solution, and

$$AXB = CB$$

$$AXB = AD$$

so that

$$CB = AD$$

In order to obtain the sufficiency, it is assumed that

$$X = A^+C + DB^+ - A^+ADB^+$$

which is a solution if $AD = CB$, $AA^+C = C$, and $DB^+B = D$.

THEOREM IV. The terms

$$A^+A,\ AA^+,\ I-A^+A, \text{ and } I - AA^+ \text{ are symmetric idempotents} \tag{17}$$

and

$$H \text{ is a symmetric idempotent which implies that } H^+ = H \tag{18}$$

THE RECURSIVE ALGORITHM

An easy consequence of equation (6) is that the matrix equation $A^TAx = A^Tb$ always has a solution. Indeed, $x = A^+b$ is a solution. In fact, if A^TA is nonsingular, this solution is unique and is the usual least squares solution. Moreover, according to corollary A, $x = A^+b + \left(I - A^+A\right)y$, for arbitrary y, gives all solutions. The triangle inequality also implies that $x = A^+b$ is that solution with minimum norm. When x is a parameter state deviation vector (i.e. the nonlinear estimation case), $x = A^+b$ yields the estimate of the parameter state deviation that has minimum norm. Physically speaking, this means that among all possible estimates of the actual parameter state, the estimate of the actual parameter state obtained by correcting the nominal parameter state by x is as close as possible (in norm) to nominal parameter state.

The possibility of $A^TAx = A^Tb$ having infinitely many solutions is clearly equivalent to the singularity of the matrix A^TA. In this case an iterative procedure would halt. However, the adroit choice of the physically meaningful minimal norm solution would allow the iteration to continue.

From a computational viewpoint, it is desirable to have a recursive algorithm for the state estimation. The following theorem will give rise to the recursive computation of the parameter state deviation.

THEOREM V. Let $a = \left(a_1, \ldots, a_n\right) \neq \theta$ be any row vector

$$a^+ = \frac{1}{\|a\|^2}\, a^T = \left(\sum_{i=1}^{n} a_i^2\right)^{-1} a^T \tag{19}$$

and

$$aa^+ = (1) \tag{20}$$

Proof: It will be shown that $a^+ = \frac{1}{||a||^2} a^T$ satisfies equations (1) to (4).

$$aa^+a = a \frac{a^T}{||a||^2} a = \frac{||a||^2}{||a||^2} a = a$$

$$a^+aa^+ = \frac{a^T}{||a||^2} a \frac{a^T}{||a||^2} = \frac{||a||^2}{||a||^4} a^T = a^+$$

$$\left(a^+a\right)^T = \left(\frac{a^Ta}{||a||^2}\right)^T = \frac{a^T}{||a||^2} a = a^+a$$

$$\left(aa^+\right)^T = \left(\frac{aa^T}{||a||^2}\right)^T = \frac{aa^T}{||a||^2} = aa^+$$

Greville (ref. 9) noted that a^+ is a constant multiple of a^T but does not give this explicit form.

As a simple example, consider the vector equation

$$(1,1) \begin{pmatrix} a \\ b \end{pmatrix} = (2)$$

This equation has a solution $a = 1$, $b = 1$. In fact, the equation has infinitely many solutions so that according to Corollary A the general solution is given by

$$\begin{pmatrix} a \\ b \end{pmatrix} = (1,1)^+ \ (2) + \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - (1,1)^+(1,1)\right]y$$

where y is an arbitrary 2×1 vector. Let c and d denote the arbitrary components of y and note that $(1,1)^+ = \left(\frac{1}{2}, \frac{1}{2}\right)^T$ to obtain

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{bmatrix} \frac{1}{2}(c-d) \\ -\frac{1}{2}(c-d) \end{bmatrix}$$

Since c and d are arbitrary, $c-d = f$ is arbitrary, and the following equation results:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 + \frac{1}{2}f \\ 1 - \frac{1}{2}f \end{pmatrix}$$

These coordinates (see fig. 1) describe a straight line whose equation is $b = 2 - a$. Hence, any vector whose coordinates satisfy $b = 2 - a$ (e.g., R, whose end point lies on the line) is a solution of the given vector equation. Note that the solution with minimum norm is $R_0 = (1,1)^T$. In a practical sense, the origin might be considered to be a nominal set of initial conditions, a given nominal parameter state, and so forth. Figure 1 gives the geometrical significance of the minimum norm solution.

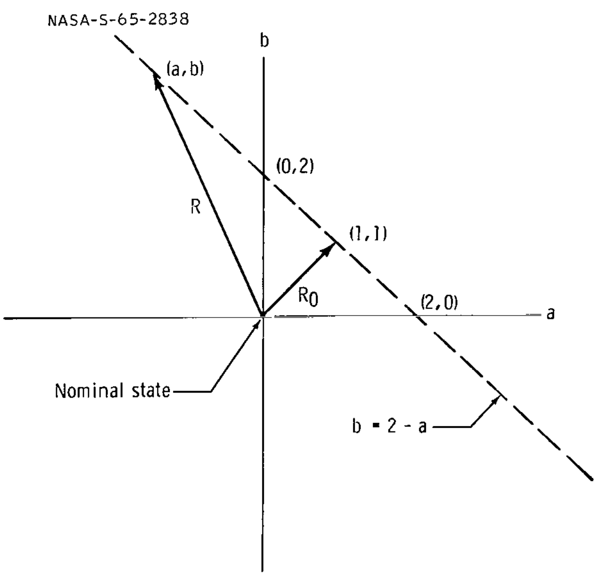

Figure 1. Graphical solution

THEOREM VI. The following theorem is due to Greville (ref. 9). Let F be any $m \times p$ matrix, f_p denote the pth column of F, and F_{p-1} represent the submatrix consisting of the first $p - 1$ columns of F, that is,

$$F = \left(F_{p-1} \vdots f_p\right)$$

then, letting

$$\left(I - F_{p-1}F_{p-1}^+\right)f_p = s_p$$

the following equation results:

$$F^+ = \begin{pmatrix} F_{p-1}^+ - c_p d_p \\ \cdots\cdots\cdots \\ d_p \end{pmatrix}$$

where

$$c_p = F_{p-1}^+ f_p$$

$$d_p = \begin{cases} s_p^+ & \left(s_p \neq \theta\right) \\ \left(1 + c_p^T c_p\right)^{-1} c_p^T F_{p-1}^+ & \left(s_p = \theta\right) \end{cases}$$

THEOREM VII. Let A be any $n \times k$ matrix, a_n denote the nth row of A, and A_{n-1} represent the submatrix consisting of the first $n-1$ rows of A, that is,

$$A = \begin{pmatrix} A_{n-1} \\ \cdots \\ a_n \end{pmatrix}$$

Then, letting

$$a_n\left(I - A^+_{n-1}A_{n-1}\right) = h_n$$

the following equation results

$$A^+ = \left(A^+_{n-1} - p_n q_n \ \vdots \ p_n\right)$$

where

$$q_n = a_n A^+_{n-1}$$

$$p_n = \begin{cases} h^+_n & \left(h_n \neq \theta\right) \\ \left(1 + q_n q^T_n\right)^{-1} A^+_{n-1} q^T_n & \left(h_n = \theta\right) \end{cases}$$

Proof: If

$$A = \begin{pmatrix} A_{n-1} \\ \cdots \\ a_n \end{pmatrix}$$

then according to equation (8) of THEOREM II

$$A^+ = \left[\left(A^T_{n-1} \ \vdots \ a^T_n\right)^+\right]^T$$

Clearly $\left(A^T_{n-1} \ \vdots \ a^T_n\right)$ has the form described in THEOREM VI with $n = p$, $m = k$, $F_{n-1} = A^T_{n-1}$, $f_n = a^T_n$, and $s_n = \left(I - A^T_{n-1}A^{T+}_{n-1}\right)a^T_n = \left(I - A^+_{n-1}A_{n-1}\right)a^T_n$.

Hence,

$$A^+ = \begin{pmatrix} A^{+\ T}_{n-1} - c_n d_n \\ \cdots\cdots\cdots \\ d_n \end{pmatrix}^T$$

where

$$c_n = \left(A_{n-1}^T\right)^+ a_n^T$$

$$d_n = \begin{cases} s_n^+ & \left(s_n \neq \theta\right) \\ \left(1 + c_n^T c_n\right)^{-1} c_n^T A_{n-1}^{T+} & \left(s_n = \theta\right) \end{cases}$$

$$A^+ = \left(A_{n-1}^+ - d_n^T c_n^T \; \vdots \; d_n^T\right)$$

Noting that $\left(s_n^+\right)^T = h_n^+$ and $s_n = \theta$ if and only if $h_n = \theta$, the following equations result:

$$c_n^T = a_n A_{n-1}^+$$

$$d_n^T = \begin{cases} h_n^+ & \left(h_n \neq \theta\right) \\ \left(1 + c_n^T c_n\right)^{-1} A_{n-1}^+ c_n & \left(h_n = \theta\right) \end{cases}$$

so that defining $p_n = d_n^T$ and $q_n = c_n^T$ completes the proof of the theorem.

Using the results and notation of THEOREM VII, the least squares solution of $Ax = b$ can be realized as a recursive process. To this end it need only be noted that the least squares solution after all n observations are made is given by

$$\hat{x}_n = A^+ b$$

so that writing $A^+ b$ in partitioned form the following equation is obtained:

$$\hat{x}_n = \left(A_{n-1}^+ - p_n q_n \; \vdots \; p_n\right) \begin{pmatrix} b_{n-1} \\ \cdots \\ b_n \end{pmatrix}$$

where b_{n-1} is a column vector of the first $n-1$ observations and b_n is a 1×1 vector corresponding to the nth observation. Multiplication yields

$$\hat{x}_n = A^+_{n-1}b_{n-1} - p_n q_n b_{n-1} + p_n b_n$$

$$= \hat{x}_{n-1} - p_n q_n b_{n-1} + p_n b_n$$

Now since $q_n = a_n A^+_{n-1}$, the last expression may be written as

$$\hat{x}_n = \hat{x}_{n-1} - p_n a_n \hat{x}_{n-1} + p_n b_n$$

$$= \left(I - p_n a_n\right)\hat{x}_{n-1} + p_n b_n$$

Note that this recursive algorithm involves at most the generalized inversion of the row vector $a_n\left(I - A^+_{n-1}A_{n-1}\right)$. All other quantities are either known from the (n-1)st state, or they are simple functions of the nth observation. The generalized inversion of the row vector $a_n\left(I - A^+_{n-1}A_{n-1}\right)$ is trivial in light of THEOREM V.

THE COVARIANCE OF THE ESTIMATE

Consider the vector equation $Ax - b = e$ where e is an $n \times 1$ error vector. In the absence of weighting it is usually assumed that $E(e) = \theta$ and $E\left(ee^T\right) = I$, where E denotes the expected value operator. The covariance matrix, $C\left(\hat{x}_n, \hat{x}_n\right)$, of the nth estimate is given by

$$C\left(\hat{x}_n, \hat{x}_n\right) = E\left[\left(\hat{x}_n - x\right)\left(\hat{x}_n - x\right)^T\right]$$

The equation $Ax - b = e$ may have infinitely many solutions; however, the minimal norm solution $x = A^+(b + e)$ is chosen. With this solution and the least squares estimate $\hat{x}_n = A^+b$, the following equation results:

$$C\left(\hat{x}_n, \hat{x}_n\right) = E\left[\left(A^+b - x\right)\left(A^+b - x\right)^T\right]$$

$$= E\left[\left(A^+e\right)\left(A^+e\right)^T\right]$$

$$= E\left(A^+ee^TA^{+T}\right)$$

$$= A^+E\left(ee^T\right)A^{+T}$$

$$= A^+IA^{+T} = A^+A^{+T}$$

The recursive computation of $C(\hat{x}_n,\hat{x}_n)$ from $C(\hat{x}_{n-1},\hat{x}_{n-1})$ and the nth observation is achieved in the following way:

$$C(\hat{x}_n,\hat{x}_n) = A^+A^{+T} = \left(A^+_{n-1} - p_nq_n : p_n\right)\left(A^+_{n-1} - p_nq_n : p_n\right)^T$$

$$= \left(A^+_{n-1} - p_nq_n\right)\left(A^+_{n-1} - p_nq_n\right)^T + p_np_n^T$$

$$= A^+_{n-1}A^{+T}_{n-1} - A^+_{n-1}\, q_n^Tp_n^T - p_nq_nA^{+T}_{n-1} + p_nq_nq_n^Tp_n^T$$

Since

$$q_n = a_nA^+_{n-1}$$

then,

$$C(\hat{x}_n,\hat{x}_n) = C(\hat{x}_{n-1},\hat{x}_{n-1}) - C(\hat{x}_{n-1},\hat{x}_{n-1})a_n^Tp_n$$

$$- p_na_nC(\hat{x}_{n-1},\hat{x}_{n-1}) + p_na_nC(\hat{x}_{n-1},\hat{x}_{n-1})a_n^Tp_n^T$$

WEIGHTED OBSERVATIONS

To minimize $(Ax - b)^TW^{-1}(Ax - b)$, that is, weighted least squares, note that W is usually a positive definite symmetric covariance matrix and hence there exists a matrix Q such that $Q^TQ = W^{-1}$. For $\tilde{A} = QA$ and $\tilde{b} = Qb$ the theoretical results are the same.

CONCLUDING REMARKS

The recursive least squares equation, by its very nature, does not involve singularity problems. In nonlinear parameter estimation procedures the minimal norm solution of the parameter state deviation is always given by A^+b. The minimal norm solution is the only solution when A^TA is nonsingular. However, in iterative procedures the singularity of A^TA would not bring the iteration to a halt. In any case, the state estimation may begin after the first observation is made. An additional computational feature of the algorithm is a small machine storage requirement. Note that the computation of $\hat{x}_n$ only requires storage (from computation of $\hat{x}_{n-1}$) of the $k \times k$ matrices $A^+_{n-1}A_{n-1}$ and $A^+_{n-1}A^{+T}_{n-1}$ and the estimate $\hat{x}_{n-1}$. Such a feature would be important, for instance, in onboard spacecraft computations.

REFERENCES

1. Gainer, Patrick A.: A Method for Computing the Effect of an Additional Observation on a Previous Least-Squares Estimate. NASA TN D-1599, 1963.

2. Bjerhammar, A.: Application of Calculus of Matrices to the Method of Least Squares With Special Reference to Geodetic Calculations. Trans. Roy. Inst. Tech. Stockholm, vol. 49. 1951, pp. 1-86.

3. Moore, E. H.: Bull. Amer. Math. Soc., vol. 26, 1920, pp. 394-395.

4. Penrose, R.: A Generalized Inverse for Matrices. Proc. Camb. Philos. Soc., vol. 51, 1955, pp. 406-413.

5. Cline, Randall: Note on the Generalized Inversion of the Product of Matrices. SIAM Review, vol. 6, 1964, pp. 57-58

6. Decell, Henry: An Alternate Form of the Generalized Inverse of an Arbitrary Complex Matrix. SIAM Review, 1965.

7. Decell, Henry: A Special Form of the Generalized Inverse of an Arbitrary Complex Matrix. NASA TN D-2748, 1965.

8. Decell, Henry: An Application of the Cayley-Hamilton Theorem to Generalized Matrix Inversion. SIAM Review, 1965.

9. Greville, T. N. E.: The Pseudoinverse of a Rectangular or Singular Matrix and the Application to the Solution of Systems of Linear Equations. SIAM Review, vol. 1, 1959, pp. 38-43.

Manned Spacecraft Center
National Aeronautics and Space Administration
Houston, Texas, March 18, 1965

2/22/85
GY

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Technical information generated in connection with a NASA contract or grant and released under NASA auspices.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

TECHNICAL REPRINTS: Information derived from NASA activities and initially published in the form of journal articles.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities but not necessarily reporting the results of individual NASA-programmed scientific efforts. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546