DTIC FILE COPY

AD-A202 620

DTIC
ELECTE
JAN 18 1989
S D
H

NONLINEAR OPTIMIZATION INVOLVING POLYNOMIAL MATRICES AND THEIR GENERALIZED INVERSES

THESIS

Raymond R. Hill Jr
Captain, USAF

AFIT/GOR/MA/88D-4

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

89 1 17 06

(1)

AFIT/GOR/MA/88D-4

NONLINEAR OPTIMIZATION INVOLVING POLYNOMIAL MATRICES AND THEIR GENERALIZED INVERSES

THESIS

Raymond R. Hill Jr
Captain, USAF

AFIT/GOR/MA/88D-4

DTIC
ELECTE
JAN 18 1989
S D
H

Approved for public release; distribution unlimited

AFIT/GOR/MA/88D-4

NONLINEAR OPTIMIZATION INVOLVING POLYNOMIAL MATRICES AND THEIR GENERALIZED INVERSES

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Operations Research

Raymond R. Hill Jr

Captain, USAF

December 1988

Approved for public release; distribution unlimited

Preface

When I began this thesis I didn't realize how broad a range of application areas this topic encompassed. For this, I thank my advisor, Dr John Jones. Since the applications are quite broad and techniques presented are widely applicable, this thesis has taken a very focused look at multiparameter (or polynomial) matrices.

You, the reader, are encouraged to first read Appendix A, Glossary of Terms, and the List of Symbols, as this will aid in understanding this thesis. Hopefully, you will come away with the notion that the generalized inverse of a matrix is a powerful tool in optimization problems, a tool that must be understood for it's usefulness in certain situations, not as an all powerful optimization technique.

I would like to thank Dr James Chrissis for his time in reading this thesis and his insights into nonlinear optimization. A special thanks goes to Maj Joe Litko whose guidance and help throughout my time at AFIT will be one of my most satisfying memories of AFIT.

Last, but most importantly, my heart felt appreciation to my wife Christy and our children, Tina, Jason, and Tiffany. Christy's hours of patiently waiting while I toiled away at the computer did more to help me succeed than any of my own self-motivation efforts.

Raymond Hill Jr.

n For
&I
ced
ation
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1
COPY INSPECTED

Table of Contents

	Page
Preface	ii
List of Figures	v
List of Symbols	vi
Abstract	viii
I. Introduction	1
Background	1
Problem	5
Research Objective	6
Approach and Presentation	7
II. Literature Review	9
Introduction	9
History	9
Applications	11
Statistical Applications	12
Nonlinear Optimization	16
Common Solutions	22
Computation	25
Penrose's Method	25
QS Decomposition	26
Direct Computation	27
Singular Value Decomposition	28
Other Techniques	29
Conclusion	32
III. Theory and Background	33
Introduction	33
Constant Coefficient Matrices	33
Fundamental Theorem of Linear Algebra, Part I and II	37
Multiparameter Matrices	39
Sontag's Factorization	42
ST Method	43
Common Solutions of Sets of Equations	49
Applications	51
Conclusion	52
IV. Applications	53
Introduction	53
Computing the Inverse	54
Nonlinear Unconstrained Optimization	59
Implicit Function Theory	61
Nonlinear Constrained Optimization	67

Control Theory. 72
Counterexamples 74
Common Solutions. 77
Conclusion. 82

V. Conclusions and Recommendations. 83

Appendix A: Glossary of Terms. 87

Appendix B: Detailed Proof of Theorem 3.13 . . . 94

Bibliography. 104

Vita 108

List of Figures

Figure		Page
1.	Sets of Generalized Inverses.	13
2.	Evolving Modular Design Formulation	23
3.	ST Technique Schematic.	48

List of Symbols

A	the m x n matrix
A^{-1}	the unique inverse matrix of matrix A
A^{+}	the unique Moore-Penrose generalized inverse of A
A^{T}	the transpose matrix of A
$A(z_j)$	the matrix A whose elements come from the j dimensional field of variable elements, z_j
A_1	matrix inverse of A that satisfies Penrose equation (1). The subscript notation is used to identify which of the Penrose equations are satisfied by the particular inverse
A^{-}	alternative notation for the generalized inverse
$A(X,Y)$	functional notation, where A is a function of X and Y
I_r	the r x r identity matrix
$\mathbb{R}$	the real number field
$\mathbb{R}^{mxn}$	the real field of m x n matrices
$\mathbb{C}$	the complex number field
$\mathbb{C}^{mxn}$	the complex field of m x n matrices
$\mathbb{R}(A)$	range, or column space, of the matrix A
$\eta(A)$	nullspace of matrix A
∇F	gradient function, gradient of function F
$\underset{\sim}{x}$	the vector x, the underscore character signifies a vector
$\hat{\underset{\sim}{b}}$	estimator of vector b, the ^ symbol stands for estimator
det()	determinant function
min	minimize, as in finding the minimum of some functional value

max	maximize, as in finding the largest of some functional value
opt	optimize, general term meaning either minimize or maximize, whichever is proper
s.t.	subject to or such that, normally used in conjuction with optimization problem constraints that must be satisfied.
$\subseteq$	subset of
$\in$	an element of
$\leq$	less than or equal to
$\geq$	greater than or equal to
$\neq$	not equal to
$\sum_{i=1}^{m}$	sum of the m elements indexed by i and sequenced from 1 to m, inclusive
$\Longrightarrow$	used to mean is equivalent to
$\{\ \}$	contains elements of a set
$^{\perp}$	as a superscript denotes orthogonal complement
$\circ$	used to denote multiplication between two matrices

Abstract

This thesis examines the applications of the generalized inverse of a matrix. In particular, use is made of the generalized inverse of a matrix containing variable elements. Such matrices are referred to as multiparameter, polynomial, or variable element matrices. The notion of a generalized inverse in fact "generalizes" the concept of a matrix inverse. A matrix inverse exists only for square, non-singular matrices. The generalized inverse extends this notion to non-square, singular matrices. The classical matrix inverse, when it exists, is a unique element of the set of generalized inverses for the matrix.

Many modern problems involve multiparameter matrices. The ability to obtain inverses for such matrices, both singular and non-singular, is a necessity in solving these problems.

This thesis consolidates the theory of generalized inverses, including extensions to multiparameter matrices. An in depth discussion is made of the ST method for computing all generalized inverses of a matrix as well as the strong interface between the ST method and the Fundamental Theorem of Linear Algebra. Finally selected application problems are solved demonstrating the utility of the generalized inverse in such problems.

NONLINEAR OPTIMIZATION INVOLVING POLYNOMIAL MATRICES AND THEIR GENERALIZED INVERSES

I. Introduction

Background

This thesis primarily involves the solving of highly nonlinear optimization problems. The approach taken is to focus on using generalized inverses of polynomial matrices as a tool for solving these problems. Since matrices are often used to describe the nonlinear problem as well as to determine optimal solutions, they play a vital role in optimization theory. The generalized inverses of these matrices, whether the matrices have polynomial or constant elements, can provide a powerful solution technique in many cases.

Optimization theory involves the problem of finding the extremum point(s) of some objective function. This function may or may not be subject to a set of constraining functions. These constraining functions limit the feasible region of potential solutions for the objective function.

An optimization study may be concerned with many different types of problem. The study may be concerned with the system cost for a new space vehicle launch system, or the problem of modeling the heat dispersion of air flow over

a wing for some fixed-wing aircraft. The study can of course address other problems in areas such as control theory, reliability, or functional analysis problems. The point is there are many examples covering all areas of current research.

In conducting an optimization study, the analyst must draw upon knowledge from numerous fields. The mathematical disciplines often used are matrix and vector theory, calculus and differential equations, and possibly some abstract mathematical theory and finite element methods. And of course in today's complex environment a thorough knowledge of computers and the computer algorithms employed to numerically find the "best candidate" solution are valuable assets.

A working definition of optimization theory can be stated as, "optimization theory is a body of mathematical results and numerical methods for finding and identifying the best candidate from a collection of alternatives without having to explicitly enumerate and evaluate all possible alternatives" (38:1). This working definition fits nicely into the previously described framework for the optimization study.

A quick look into any text on optimization theory reinforces the statements just made. In addition to the numerous fields of knowledge and research involved, the optimization text provides an appreciation of the many aspects of optimization that must be considered, not only in

formulating the problem, but also in solving the problem.

As initially stated, the optimization problem may be constrained or unconstrained. In unconstrained optimization, the function can take on any defined numerical value. In such cases, iterative techniques are quite efficient. Some popular techniques include Golden Section and Fibonacci algorithms (49:121) for single-dimensional problems. For higher-dimensional problems, calculus techniques involving the gradient function are the usual choice.

Constrained optimization must search for the best candidate in some predetermined region of the number field. In order to be a valid candidate for the optimal solution a set of constraining functions must be satisfied by the candidate objective function solution. The two dominant techniques in the linear function arena (i.e. linear programming) are the simplex algorithm, first developed by Dantzig (8:14), and Karmarkar's algorithm (14:75). For nonlinear functions, there are gradient search techniques, penalty function techniques, and iterative linear approximation techniques, to name a few (13:vi). In addition there is the classical Lagrange multiplier method and the Kuhn-Tucker optimality conditions that are the basis for most other techniques in addition to being a solution technique in themselves (38:184-200).

Regardless of the technique, the theory of matrices is a key player. Once again, nearly all optimization texts have an appendix or chapter dedicated solely to matrix theory.

Matrix theory is vital to understanding problem formulation and then understanding the solution techniques. It is this matrix theory that is the driving force of this thesis. In particular, this thesis explores the extensions of that classical matrix theory to include the generalized inverses of all matrices.

The first attractive thing about using matrices is the very compact, easy to follow, problem formulations obtained using matrices. Systems of equations are very neatly summarized using matrices. The following example demonstrates this.

Consider the following system, which is an example from Chapter IV. Disregard for the moment that the matrix in (1.3) below contains variable elements (i.e. it is a polynomial or multiparameter matrix):

$$\begin{cases} 400X^3 - 400XY + 2X = 2 & (1.1) \\ -200X^2 + 200Y = 0 \end{cases}$$

This may be written in a much more compact notation as:

$$A(X,Y) \begin{pmatrix} X \\ Y \end{pmatrix} = B(X,Y) \qquad (1.2)$$

where

$$A(X,Y) = \begin{bmatrix} 400X^2 + 2 & -400X \\ -200X & 200 \end{bmatrix} \qquad (1.3)$$

and

$$B(X,Y) = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad (1.4)$$

Equation (1.2) is easier to understand than equation (1.1) since it is uncluttered and more compact. The problem

solution can then be derived using the same matrix notation. For example a problem of the form (1.2) is easily solved using the notion of a matrix inverse. Should this inverse, A^{-1}, exist, the unique solution is given by:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = A^{-1}(X,Y)\ B(X,Y) \qquad (1.5)$$

However, not all matrices have this "classical" matrix inverse. In these cases, the notion of a generalized inverse is used to solve the optimization problem. More details will come later, but the generalized inverse does in fact generalize the idea of a matrix inverse since the A^{-1} inverse matrix, when it exists, is identical to the Moore-Penrose generalized inverse matrix (46:138).

The history of the generalized inverse is short. Some initial work in the 1920's and 1950's was followed by a flurry of activity in the late 1960's through the early 1980's. However, this work dealt primarily with constant coefficient matrices. For most problems in areas such as statistics, control theory, and even optimization, this limited applicability was sufficient. The next chapter surveys a cross section of these application areas. But as shown in equation (1.3), matrices can contain variable coefficients. This then is an example from the newest, most rapidly expanding area of research involving generalized inverses, that of multiparameter matrices, or matrices that contain variable elements.

Problem

Much of the work done involving generalized inverses is in the applications of inverses of matrices with constant coefficients. It has not been until recently that interest has turned towards working with matrices having variable elements. The reason for this interest is due to the ever increasing complexity of modern systems as well as the ability of modern supercomputer systems to manipulate and evaluate variable element equations and matrices. Such expert systems as MACSYMA (50) enable the user to manipulate purely symbolic equations or matrices. The ability of a system to handle variable elements is necessary in modern systems theory involving multiparameters (20:253 ; 4:491), and large scale network problems (42:514), to name two examples.

Thus, work must be done to enable users to solve such multiparameter systems. This thesis brings together the theory set forth to date and demonstrates the use of the generalized inverse of a multiparameter matrix as a tool applied to selected applications such as control theory and nonlinear optimization. The ability to use this technique gives the analyst an often powerful technique for solving complex problems.

Research Objective

This thesis provides a concise, yet thorough, compilation of generalized inverse theory. The theory is then used to solve practical examples from fields such as optimization and control theory. The theory and examples

provide the reader with an appreciation of how valuable a generalized inverse can be in solving a wide range of problems.

Approach and Presentation

Chapter II reviews a cross-section of the generalized inverse field. The purpose of this review is to emphasize the range of applications and the way in which the generalized inverse does in fact provide a very general solution format. It begins with a brief history of the theory, followed by discussions of various applications. First among the applications is regression analysis, followed by nonlinear optimization techniques. Each discussion explains how the technique is implemented and how the generalized inverse plays a crucial role. The final section of this chapter looks at some of the various computational techniques available. Again the goal is to explain the techniques, not just enumerate them.

Chapter III presents a consolidation of the theory and knowledge at the basis of this thesis work. The intent is to bring together in one concise chapter, the relevant theorems presented to date, supplemented by discussions of the theorems. As a result, this chapter also highlights the trend in the theory regarding multiparameter matrices. For the most part, the theorems are presented without the proofs but contain references where the proofs can be found. Each theorem is discussed to enhance reader understanding of the generalized inverse theory being presented.

Chapter IV applies the theory from Chapter III in specific application areas. The first topic is the detailed steps to follow in computing the generalized inverse using the ST computational technique. This follows directly from the theory laid out in the previous chapter. The first three applications presented involve unconstrained optimization, implicit function theory, and constrained optimization, respectively. The chapter concludes with a robust control theory problem and an example that employs the theory regarding common solutions to sets of matrix equations.

Finally, in Chapter V, the thesis is summarized and the important points are reiterated. The final point made is recommendations for future areas of research regarding multiparameter matrices.

II. LITERATURE REVIEW

Introduction

The purpose of this chapter is to examine the current knowledge in the area of generalized inverses of matrices. This concise summary focuses on the theory and applications of generalized inverses, and is divided into three main sections.

The first section briefly discusses the history of the generalized inverse. Some specific application areas follow in section two. The areas discussed represent a small cross-section of the application areas. The emphasis is on the diversity of the field, while providing an understanding of just how the particular technique under examination applies, and exploits, the generalized inverse matrix. The final section addresses techniques developed to compute the various classes of generalized inverses.

History.

Given any square matrix, A, if the determinant of A is non-zero (i.e., Det(A) $\neq$ 0) then there exists a matrix, A^{-1}, that satisfies the property

$$A^{-1}A = A\,A^{-1} = I \qquad (2.1)$$

where I is the identity matrix. The matrix A^{-1} is called the inverse of A and A is said to be invertible or nonsingular. However, if A is non-square, or is square with a zero determinant (i.e., Det(A) = 0), then there is no matrix B such that

$$A B = B A = I \qquad (2.2)$$

and A is said to be a singular matrix. In cases where the matrix A is singular, inverses from a larger class of matrix inverses must be computed. This is the class of generalized inverses, of which A^{-1} is a unique element when it does in fact exist.

In his 1985 AFIT thesis, Murray (27:1-3) classified the history of generalized inverses by identifying five key developments. The first occurred in 1903 when Fredholm introduced the concept of the generalized inverse, calling the inverse matrix a pseudoinverse. In 1920, Moore proved algebraically the concept of a unique generalized inverse for every finite matrix. He called his matrix a general reciprocal matrix. It wasn't until 1951, in work done by Bjerhammer, that the relationship of this generalized inverse was extended to a system of linear equations.

Using Bjerhammer's results, yet apparently unaware of Moore's earlier work, Penrose showed that this generalization of the "classical" matrix inverse was unique for every matrix. Penrose defined four conditions that this inverse must meet. These Penrose conditions (32:406) are used as a basis for classifying all the generalized inverses of a matrix. The conditions, along with Penrose's original theorem, are presented and discussed in depth in Chapter III. For now, the conditions are presented without proof for discussion purposes. The four conditions that Penrose identified are:

$$
\begin{aligned}
&(1) && A\,A^{+}\,A = A && (2.3)\\
&(2) && A^{+}\,A\,A^{+} = A^{+}\\
&(3) && (A\,A^{+})^{*} = A\,A^{+}\\
&(4) && (A^{+}\,A)^{*} = A^{+}\,A
\end{aligned}
$$

where * denotes the conjugate transpose of the matrix.

The matrix that satisfies these four equations, denoted as A^{+}, is referred to as the Moore-Penrose generalized inverse in recognition of contributions from both researchers. Later work involved matrices that satisfy some, but not necessarily all, of the Penrose equations. For example, a matrix B that satisfies condition (1) may not be unique, but is sufficient for use in solving sets of linear equations (28:127). A matrix that satisfies conditions (1) and (2) is often referred to as a weak-generalized inverse, WGI, or the $A_{1,2}$ generalized inverse.

Murray identifies the fifth, and final major development, to be the Jones ST method of computing all the generalized inverses of a matrix (27:3). Since the ST method is the particular technique used in this thesis, a more in-depth discussion of the technique is provided in Chapter III.

Applications.

In 1956 Penrose showed that the unique A^{+} generalized inverse, when used to solve an equation of the form:

$$\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{y} \qquad (2.4)$$

provided the least-squares, minimum-norm solution vector $\underset{\sim}{x}$. However, there are other classes of generalized inverses for a matrix that satisfy only a portion of the four Penrose

equations. Since only the A^+ is unique for any matrix A, other inverses are members of subsets of inverses. For instance, an $A_{1,2}$ matrix is not unique for a given matrix A, but just one member of the subset of $A_{1,2}$ matrix inverses. Some ideas from set theory (47:216) show this subset relationship:

$$A^+ \subseteq (A_{1,2,3}, A_{1,2,4}) \subseteq (A_{1,2}) \subseteq (A_1, A_2) \quad (2.5)$$

This set relationship is shown pictorially in Figure 1 (27:5).

After Penrose's original work, much was done in identifying the properties of these subsets of inverses. As Penrose proved, the A^+ matrix provides the least-squares, minimum-norm solution to an equation of the form (2.4). However, a lesser inverse, the $A_{1,3}$, generates the least-squares estimator. Similarly, the $A_{1,4}$ inverse generates the minimum-norm solution (28:129-132). As expected, these generalized inverses come up quite often in statistical applications, as the next section illustrates.

Statistical Applications. A regression analysis model is of the form:

$$\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{\varepsilon} \quad (2.6)$$

where $\underset{\sim}{y}$ is the vector of dependent variables, X is the matrix of independent variables, and $\underset{\sim}{\varepsilon}$ represents a random error component with zero mean and known covariance matrix, V. The vector $\underset{\sim}{b}$ of regression parameters describes a linear relationship between the independent variables and the dependent variables. In reality, the values of $\underset{\sim}{b}$ are unknown

and must be estimated from the data. A key question is how to select the best estimators of $\underset{\sim}{b}$.

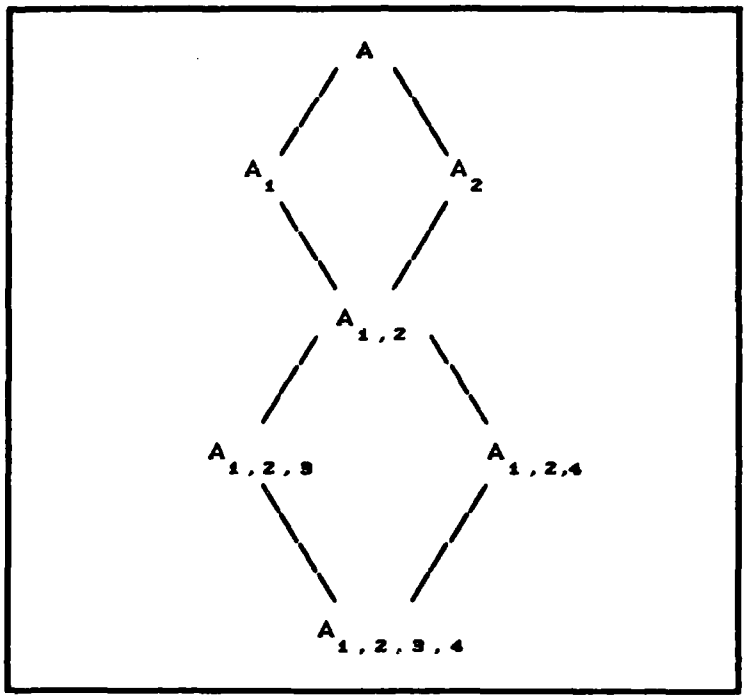

FIGURE 1: SETS OF GENERALIZED INVERSES

These estimators are generally denoted as $\hat{\underset{\sim}{b}}$ (or $\hat{\underset{\sim}{\beta}}$ in some texts). Desirable properties for the vector of estimators, $\hat{\underset{\sim}{b}}$, are minimum variance and that it be an unbiased estimator of the true parameters, $\underset{\sim}{b}$. If it satisfies these properties it is called a best linear estimator (BLE). Nelson (30:1-16) examined the use of generalized inverses in regression analysis, both in unconstrained and constrained regression analysis. In both cases, the BLE for the problem was found in terms of the Moore-Penrose generalized inverse.

The first case discussed is that of the unconstrained case (30:2). In this instance the BLE of (2.6) is given by:

$$\hat{\underset{\sim}{b}} = (X^T V^{-1} X)^+ X^T V^{-1} \underset{\sim}{y} \qquad (2.7)$$

and the covariance of $\hat{\underset{\sim}{b}}$ is defined as:

$$\text{cov}(\hat{\underset{\sim}{b}}) = (X^T V^{-1} X)^+ \qquad (2.8)$$

Nelson considered equality constraints (30:2-7) on $\hat{\underset{\sim}{b}}$ of the form:

$$A \underset{\sim}{b} = \underset{\sim}{t} \qquad (2.9)$$

$$(\text{inequality} \quad A \underset{\sim}{b} \leq \underset{\sim}{t})$$

In this case, his best restricted linear estimator (BRLE) was defined as:

$$\underset{\sim}{b} = A^+ \underset{\sim}{t} + C^+ X V^{-1} (\underset{\sim}{y} - X A^+ \underset{\sim}{t}) \qquad (2.10)$$

$$\underset{\sim}{C} = (I - A^+ A) X^T V^{-1} X (I - A^+ A)$$

and

$$\text{cov}(\hat{\underset{\sim}{b}}) = C^+ \qquad (2.11)$$

Nelson's methodology is more involved when the problem involves inequality constraints (30:7-10) of the form in (2.9). He first determines whether the unconstrained optimum is feasible, with respect to the constraints. If not, then he uses the fact that the optimum must lie on a constraint boundary. This implies that a certain subset of constraints intersect to form the constraint boundary where the optimal point resides. Since the points lies on the boundary, the constraints must be satisfied as equality constraints.

Using this, Nelson essentially conducts a tree search of each basis or subset of constraints satisfied as equalities (other constraints must still be satisfied). In

this manner, the optimum solution is found in a finite number of steps. However, this procedure is combinatorially inefficient for problems involving larger constraint sets.

Stewart (45:634-662) did work somewhat similar to Nelson. His work differed in that he was primarily interested in how perturbations in A, due to uncertainty about A, effect the generalized inverse A^+, the minimum-norm solution, $A^+\underset{\sim}{b}$, and the least-squares problem.

Another area is the study of Markov chains, either discrete or continuous, examined by Hunter in 1982. In his work, Hunter (16) characterized all the generalized inverses of the matrix (I-P). For discrete Markov chains, P is the one-step transition matrix. In the continuous case, P is the infinitesimal generator. Hunter found the necessary stationary and first passage time distributions for the problem in terms of these generalized inverses. In proving his results, Hunter refuted previous work claiming generalized inverses were not applicable to the study of Markov chains (16:196-197).

The generalized inverse arises in many other statistical settings involving singular matrices. For example, Rao (36:201-203), Albert (28:225), and Henk Don (12:225-240) examined Maximum Likelihood Estimation involving singular information matrices. Rao and Yanai (37) looked at the Gauss-Markov model and in particular the models involving generalized inverses of partitioned matrices. Hsuan (15) examined some specific uses of the A_2

generalized inverse matrix. Among the areas he examined were the least-squares problem and the conditions under which the quadratic form of a multivariate normal random variable will follow a chi-square distribution (15:245-247). A classic text by Rao and Mitra provides chapters dedicated to just statistical applications of the generalized inverse matrix (36:136-168).

Nonlinear Optimization. As previously discussed in Chapter I, matrices play a key role in optimization studies. This section looks at some optimization techniques involving the generalized inverse.

A quadratic programming problem is a particular type of nonlinear programming problem (NLP) involving a quadratic, convex, objective function, subject to a set of linear constraints (49:14). Nelson (29:1-21) discusses a quadratic programming problem of the form:

$$\max f(\underset{\sim}{x}) \qquad (2.12)$$

$$\text{s. t.} \quad g_i(\underset{\sim}{x}) \leq 0 \text{ for } i=1,\ldots,I$$

For any feasible solution to the problem, only a portion of the I linear constraints are binding. For any feasible solution, $\underset{\sim}{x}$, a constraint can be binding or non-binding. If the functional value at $\underset{\sim}{x}$ lies on the boundary of the feasible region defined by a constraint, then that constraint is a binding constraint. If the functional value is not on the feasible region boundary defined by the constraint then the constraint is considered a non-binding constraint. Nelson examines possible

combinations of constraints as if that combination of constraints were binding (treated as strict equality constraints) and determines the optimal solution given the particular combination of constraints. If the resulting solution satisfies all the constraints the solution is a potential optimal solution. Once all possible combinations of binding constraint sets are examined, the best potential solution obtained is the optimal solution (29:19-20).

Nelson uses the generalized inverse to handle the non-square, singular, matrices that result from partitioning the constraint set. Thus, a more general technique is obtained than if the classical A^{-1} inverse were used. Since this is the same technique Nelson employs for inequality constrained least-squares problems, the technique suffers from the same combinatorial inefficiencies as before. For large problems involving many constraints, this technique would be very cumbersome and impractical.

Shankland avoided the combinatorial complexity of Nelson's algorithm in his quadratic programming technique. The formulation he used was (40:2):

$$\max \quad S = \underset{\sim}{a}^T \underset{\sim}{x} - 1/2 \; \underset{\sim}{x}^T B \; \underset{\sim}{x} \qquad (2.13)$$

$$\text{s.t.} \quad C \; \underset{\sim}{x} - \underset{\sim}{d} \leq 0$$

where B is a positive definite, symmetric matrix.

Shankland first decomposes B into the product of a lower triangular matrix, L, and it's transpose, L^T, and performs the three following transformations on the formulation of (2.13):

$$B = L L^T \qquad (2.14)$$
$$\underset{\sim}{x}' = L^T \underset{\sim}{x}$$
$$\underset{\sim}{a}' = L^{-1} \underset{\sim}{a} \quad \text{and} \quad C' = L^{-1} C$$

Shankland's final transformation shifts the origin to the unconstrained maxima of (2.13), a point he calls $\underset{\sim}{a}'$. Using the transformations:

$$\underset{\sim}{x}'' = \underset{\sim}{x}' - \underset{\sim}{a}' \qquad (2.15)$$
$$\underset{\sim}{d}' = \underset{\sim}{d} - C^T \underset{\sim}{a}'$$

produces the final formulation (40:3):

$$\max \; S = 1/2 \, (\underset{\sim}{a}'^T \underset{\sim}{a}' - \underset{\sim}{x}''^T \underset{\sim}{x}'') \qquad (2.16)$$
$$\text{s.t.} \quad C'^T \underset{\sim}{x}'' - \underset{\sim}{d}' \leq 0$$

If the feasible region defined for (2.16) contains the origin (i.e., shifted origin), then the origin is the solution. If not, then the point on the surface of the feasible region closest to the new origin is the solution point for the problem. The task is to find this point closest to the origin.

Although, the origin, $\underset{\sim}{x}_o = 0$, is not a feasible point, Shankland treats it as such for the moment. Using $\underset{\sim}{x}_o$, a subset of constraints, V, is formed from the violated constraints. This subset of constraints is solved as equality constraints. If the feasible region defined by V is non-empty, a Lagrange multiplier technique yields the solution point. If the feasible region defined by V is empty, a generalized inverse obtains a solution in terms of least-squares. The least-squares function, $(e^v)^T e^v$, arises from:

$$C^T\underset{\sim}{x} - \underset{\sim}{d} = \underset{\sim}{e}^v \neq 0 \qquad (2.17)$$

not equal to zero since the region is inconsistent in terms of the intersection of the constraints.

This least-squares solution might be improved using an iterative refining technique (40:5-7). In this refining process, violated constraints are retained, over-satisfied constraints are removed, and a feasible solution obtained for the resulting constraint subset. If the solution violates some other constraint(s), the process is repeated. If the refining process finds no feasible solution, the initial least-squares solution is retained as the best solution to the problem. The primary feature of using a generalized inverse is that the least-squares solution is obtained even when a unique solution is not available due to inconsistency in the constraint set. When this inconsistency occurs, Shankland states that, "the constraints are mutually incompatible" (40:7).

A penalty function technique is an optimization technique that transforms a constrained optimization problem into an unconstrained optimization problem. The basic concept is to force convergence to the optimal solution by applying increasing penalties for not satisfying the constraints imbedded in the unconstrained function. Thus, there is a trade off between satisfying the constraints and minimizing the objective function (13:299-300).

Fletcher (28:223-224) uses the generalized inverse to generate a penalty function from the original constrained

problem. Fletcher's technique employs the gradient of both the objective function ($\nabla f(x) = F$) and the constraints ($\nabla g(x) = N$), as well as the Hessian matrix of the objective function ($H(x)$). Then to ensure proper behavior of the function, a large positive definite matrix, Q, is added to the function giving an unconstrained function:

$$\phi = F - NN^{+}F + g^{T}Qg \qquad (2.18)$$

This unconstrained function can now be solved using any appropriate unconstrained optimization technique. Fletcher points out that this function:

(a) is suitable for a variety of problems,

(b) is well conditioned, and

(c) strongly interfaces to the classical Lagrange method of multipliers, as well as other penalty function techniques.

In a 1965 article, Charnes and Kirby used the generalized inverse, in particular the A_1 inverse, to show "that the modular design problem is simply a special case of a large class of engineering design problems" presented elsewhere in the literature (6:843). This special case problem is the separable convex function subject to linear equality constraints. A separable, convex function can be approximated by a series of linear functions (i.e., linear approximation). Such approximations enable the use of the more efficient linear programming packages to solve the problems. The algorithms specifically designed for the modular design problem are often complex and inefficient so

the increased efficiency gained from the approximations offset the effort required to reformulate the problem.

The modular design problem presented by Charnes and Kirby is of the following form (6:836):

$$\min \left[\sum_{i=1}^{m} e_i E_i\right] \left[\sum_{j=1}^{n} d_j D_j\right] \qquad (2.19)$$

$$\text{s.t.} \quad E_i D_j \geq R_{ij} \quad (i = 1,\ldots,m)$$

$$E_i, D_j \geq 0 \quad (j = 1,\ldots,n)$$

where e_i, d_j, and R_{ij} are positive constants. Charnes and Kirby make the statement that the algorithm typically used to solve these types of problems is very slow in converging to the optimum (6:837).

Charnes and Kirby use a series of transformations to produce an equivalent formulation of (2.19) whose properties of convexity and separability enable use of the linear programming packages. The key transformation involves the generalized inverse which is coupled with the concept of slack variables from linear programming. A key aspect of linear programming involving linear inequality constraints is that slack variables, ω, enable the following transformation:

$$A \underset{\sim}{x} \geq \underset{\sim}{b} \longrightarrow A \underset{\sim}{x} = \underset{\sim}{b} + \omega \qquad (2.20)$$

From the generalized inverse theory, a consistency condition for the equation $A \underset{\sim}{x} = \underset{\sim}{b}$ to have a solution is that $A A^{+} \underset{\sim}{b} = \underset{\sim}{b}$ (6:838). This concept can be combined with (2.20) above to produce the following equivalence relationship:

$$\begin{array}{lll} \text{opt } F(\underset{\sim}{x}) & & \text{opt } F(\underset{\sim}{x}) \\ \text{s.t. } A\,\underset{\sim}{x} \geq \underset{\sim}{b} & \longrightarrow & \text{s.t. } A\,\underset{\sim}{x} = \underset{\sim}{b} + \omega \\ & & A\,A^{+}(\underset{\sim}{b} + \omega) = (\underset{\sim}{b} + \omega) \\ & & \omega \geq 0 \end{array} \qquad (2.21)$$

Working from the modular design formulation (2.19), the following three transformations are applied to the initial problem formulation:

(1) $y_i = e_i E_i$

$z_j = d_j D_j$

$r_{ij} = e_i d_j R_{ij}$

(2) $y_i = e^{u_i}$

$z_j = e^{v_j}$

$c_{ij} = \ln(r_{ij})$

after which the equivalence relation of (2.21) is used as a third transformation.

These transformations change the problem formulation according to the sequence in Figures 2.a through 2.d. Although the Figure 2.d formulation may be solved for the ω values, the formulation is extended to the final form, Figure 2.e, using the transformations:

$$T = (I - A\,A^{+}) \qquad (2.22)$$

$$AA^{+}(\underset{\sim}{b} + \omega) = \underset{\sim}{b} + \omega \longrightarrow (I - A\,A^{+})\omega = -(I - AA^{+})\underset{\sim}{b}$$

The final form of the modular design problem, given in Figure 2.e, is the desired separable convex function subject to linear equality constraints. Though use of the appropriate transformations, Charnes and Kirby demonstrate the similarity between the modular design problem and linear programming problems making particular use of the

$$\min \left[\sum_{i=1}^{m} e_i E_i\right] \left[\sum_{j=1}^{n} d_j D_j\right]$$

(A) $$\text{s.t.} \quad E_i D_j \geq 0$$
$$E_i, D_j \geq 0$$

$$\min \sum_{i=1}^{m} y_i \sum_{j=1}^{n} z_j$$

(B) $$\text{s.t.} \quad y_i z_j \geq R_{ij}$$

$$\min \sum_{i=1}^{m} e^{ui} \sum_{j=1}^{n} e^{vj} \longrightarrow \sum_{i=1}^{m} \sum_{j=1}^{n} e^{ui + vj}$$

(C) $$\text{s.t.} \quad u_i + v_j \geq c_{ij}$$

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} e^{(c_{ij} + \omega_{ij})}$$

(D) $$\text{s.t.} \quad A A^{+}(\underset{\sim}{c} + \underset{\sim}{\omega}) = (\underset{\sim}{c} + \underset{\sim}{\omega})$$
$$\underset{\sim}{\omega} \geq 0$$

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} e^{(c_{ij} + \omega_{ij})} = \sum_{i=1}^{m} \sum_{j=1}^{n} r_{ij} e^{\omega_{ij}}$$

(E) $$\text{s.t.} \quad \omega_{11} - \omega_{1,l} + \omega_{k+1,1} + \omega_{k+1,l} = -c_{11} + c_{1,l} + c_{k+1,1} - c_{k+1,l}$$
$$k = 1, \ldots, m-1$$
$$l = 2, \ldots, n$$
$$\omega_{ij} \geq 0$$

FIGURE 2. EVOLVING MODULAR DESIGN FORMULATION

generalized inverse matrix.

Common Solutions. The last application area isn't so much an application as it is a demonstration of the trend regarding applications of generalized inverses. This trend is the move from matrices with constant coefficients to matrices involving polynomial elements. These matrices with polynomial elements are called multiparameter matrices. An understanding of this trend demonstrates future applications of, and research into, generalized matrix inverses.

In Penrose's original work, he presented necessary and sufficient conditions for solutions to exist to a problem or set of equations (32:409). In 1972, Mitra discussed the simultaneous solution of two matrix equations (24), unaware that a more generalized discussion was presented by Morris and Odell in 1968 (26). In their earlier article, Morris and Odell proved conditions for a common solution to n matrix equations. Also in 1972, Shurbet (41) defined the necessary and sufficient conditions for the consistency of a system of linear matrix equations. This work built on the work of Morris and Odell and advanced the theory for the constant coefficient matrices.

With developments in multiparameter, multidimensional systems, there arose the need to do similar work with the generalized inverses of polynomial matrices. The first attempt to study these multiparameter matrices was in 1978 by Bose and Mitra (4:49). Their article provided necessary and sufficient conditions regarding the algebraic structure

of the multiparameter matrix. In an extension of the work of Bose and Mitra, Sontag gave a complete characterization of the weak-generalized inverse ($A_{1,2}$) for matrices involving several polynomial elements (42).

Later work by Jones in 1983 (18) and 1985 (20) extended the theory, providing necessary and sufficient conditions for the solution of several types of matrix equations. Jones' work examined much of the work already done for constant coefficient matrices and extended the theory to the case of multiparameter matrices. Finally, to complete the cycle of research, Jones extended the work of Morris and Odell regarding common solutions to sets of matrix equations to the area of multiparameter matrices (17). In his work, Jones provided the necessary and sufficient conditions for the existence of common solutions of n multiparameter matrix equations.

Computation

The last topic in this section concerns methods for computing the generalized inverse. As previously stated, Murray cites Jones' ST method for computing all generalized inverses as the last significant contribution to the theory. The purpose of this section is to present some of the other computation techniques. The ST method is presented in detail in the next chapter.

Penrose's Method (36:208-209). The essence of this technique relies on the ability to partition a matrix A, having rank of r, into the following form:

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \qquad (2.23)$$

where A_1 is of dimensions rxr, $A_4 = A_3 A^{-1} A_2$ and A_2 and A_3 are of suitable order such that the matrix A remains an mxn dimensional matrix. From this partition, define P as:

$$P = (A_1 A_1^T + A_2 A_2^T)^{-1} A_1 (A_1^T A_1 + A_3^T A_3)^{-1} \qquad (2.24)$$

and then

$$A^+ = \begin{bmatrix} A_1^T P A_1^T & A_1^T P A_3 \\ A_2^T P A_1^T & A_2^T P A_3^T \end{bmatrix} \qquad (2.25)$$

Though straightforward, this method relies upon knowledge of the rank of the matrix A and the computation of classical matrix inverses.

QS Decomposition. This method depends upon decomposing the matrix A into the product of two matrices, Q and S so that:

$$A = Q S \qquad (2.26)$$

where Q has orthogonal columns and S is an upper triangular matrix (28:285). This decomposition is then used to obtain:

$$A^+ = S^T (S S^T)^{-1} Q^T \qquad (2.27)$$

The recommended numerical technique for accomplishing the computations in (2.27) is to solve:

$$(S S^T) X = Q^T \qquad (2.28)$$

and form the product:

$$A^+ = S^T X \qquad (2.29)$$

which gives the desired A^+ generalized inverse (28:286).

A somewhat related technique uses a different decomposition, namely:

$$A = L\,U \tag{2.30}$$

where L and U are lower and upper diagonal matrices, respectively. Though this decomposition technique is numerically cheaper to perform than the decomposition of (2.26), computing the A^+ from the L and U matrices of (2.30) involves more operations than using operations (2.27) through (2.29). Details can be found comparing each technique in an article by Noble in Nashed's volume on Generalized Inverses (28:285-288). Numerical techniques for each decomposition fall under the headings of QR decomposition (31:315-323) for (2.26) and LU factorization for (2.30) (5:342-350).

Direct Computation. Given an mxk matrix A, of known rank r, the most straightforward technique for computing A^+ is by the following formula:

$$A^+ = (A^TA)^{-1}\,A^T \tag{2.31}$$

which is the familiar least-squares solution of linear equations and linear regression. The actual computation of (2.31) can be accomplished by numerically solving the following for the matrix X (28:278):

$$(A^TA)\,X = A^T \tag{2.32}$$

The problem with (2.31) and (2.32) arises when the matrix A contains linearly dependent rows (or columns). This

causes a singular matrix (A^TA), which causes (2.31) to fail, and worsens the numerical computation of (2.32) (28:279).

For example, compute A^+, using equation (2.31), for the matrix A defined as:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad (2.33)$$

The product A^TA is the following:

$$A^TA = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \qquad (2.34)$$

which is singular implying $(A^TA)^{-1}$ doesn't exist. The previously discussed decomposition methods improve the conditions for computing the generalized inverse. Thus the decomposition methods are recommended over this particular direct method (28:284).

Recent advances in computer algebra enable researchers to expand into symbolic computations. Computer-based expert systems such as REDUCE, MAPLE, SMP, muMATH, and in particular MACSYMA (50:3) provide such symbolic computational environments. Sample manipulations are limits and integrals. Frawley (11:419) uses a limit form of (2.31) to compute generalized inverses in MACSYMA. The form used is:

$$A^+ = \lim_{X \to 0} ((A^TA + X^2I)^{-1}A^T) \qquad (2.35)$$

Singular Value Decomposition. The singular value decomposition method (31:323-330) is somewhat more

complicated. The method makes use of the eigenvalues and eigenvectors of a matrix. A key idea is that a matrix of orthonormal eigenvectors of a matrix A can be used to decompose A into a diagonal form where the diagonal elements are the eigenvalues of A. If M is this matrix of eigenvectors, this decomposition is given in the following equation:

$$M^{-1}AM = \Sigma \qquad (2.36)$$

where Σ is the diagonal matrix of eigenvalues.

The technique for finding A^+ (31:326) involves the orthonormal eigenvectors of the matrices AA^T and A^TA, as well as their eigenvalues, which are equal. The matrix V contains the eigenvectors of A^TA and the matrix U contains the eigenvectors of AA^T. A diagonal matrix, Σ, is defined as before in (2.36). The equation for A^+ is then (31:337):

$$A^+ = V \Sigma^{-1} U^T \qquad (2.37)$$

Although very straight forward computationally, the problem with the method involves finding the eigenvalues and eigenvectors of the matrices. Numerical computations to find the eigenvalues, and the corresponding eigenvectors, of a matrix can introduce error into the computations as well as require significant computer resources.

Other Techniques. As previously mentioned, the A^+ matrix is actually a unique member of a class of generalized inverses. Further, mention was made of the fact that a lesser inverse, such as the A_1 or $A_{1,2}$ inverse may suffice

in some applications. Thus, there are techniques for computing these subsets of matrices.

One technique to compute the A_1, or the A_2 matrix is similar to Penrose's technique (31:208). To compute the A_1, partition the matrix A such that $A = (B_1 \mid B_2)$, where the dimensions of the submatrix B_1 are determined by the rank of A, and the dimensions of B_2 are appropriate for the matrix. Using these submatrices, the generalized inverse, A_1, is computed as:

$$A_1 = \begin{bmatrix} (B_1^T B_1)^{-1} B_1^T \\ 0 \end{bmatrix} \qquad (2.38)$$

The A_2 generalized inverse can be computed in a similar fashion. The matrix A is partitioned so that

$$A = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \qquad (2.39)$$

where as before the dimensions of the matrix C is determined by the rank of the matrix A. The dimensions of C_2 are again appropriate for the matrix. The formula for the A_2 generalized inverse is:

$$A_2 = \begin{bmatrix} (C_1^T C_1)^{-1} C_1^T & 0 \end{bmatrix} \qquad (2.40)$$

The observant reader will note the similarity of the previous formulas with equation (2.31) and the least-squares estimator. For full-rank matrices, this technique is a special case of the direct computation method. The drawback with the technique is the need to predetermine the rank of

the matrix A. There is also the need to permute the matrix A to obtain the proper partitioning.

One final technique is worth noting. This technique, due to Urquhart (47), starts with any technique to compute the A_1 matrix. Using this technique, the following matrices are computed for the matrix A:

$$B_1 = (AA^T)_1 \qquad C_1 = (A^TA)_1 \tag{2.41}$$

These additional matrices, the original matrix A, along with A^T, are then used to obtain representatives from each set of generalized inverses. The following formulas compute these generalized inverses:

$$A_{1,2} = A_1\, AA_1 \tag{2.42}$$

$$A_{1,2,3} = A^TB_1 \tag{2.43}$$

$$A_{1,2,4} = C_1A^T \tag{2.44}$$

$$A_{1,2,3,4} = A^+ = A^TB_1\, AC_1\, A^T \tag{2.45}$$

Each of the techniques discussed make explicit use of the predetermined rank of A and use some form of matrix decomposition. Iterative methods, that converge to the Moore-Penrose generalized inverse (A^+), were not addressed in this review. Since an iterative method converges to A^+, ideally A^+ must be known to say with certainty that $A^\phi \rightarrow A^+$, where A^ϕ is defined as the intermediate values of the inverse.

Standard computer software packages, such as IMSL, EISPACK, and LINPACK contain routines for the generalized

inverse (28:298), but none of these use Gauss-Elimination as the basis of the computations. These routines also cannot handle matrices with polynomial elements, but then were not really designed to do so. The technique used in this thesis, Jones' ST method, uses Gauss-Elimination, determines the rank of the matrix A, generates all generalized inverses of A, and is very applicable to multiparameter matrices. The algorithm is easy to understand and is numerically and computationally efficient (27:87). The theory behind the technique as well as a description of the algorithm are discussed at length in the next chapter.

Conclusion

This chapter examined the published knowledge regarding the theory, application, and computation of generalized inverses. The applications section highlighted the diversity of the field in applying generalized inverses. The minimum-norm, least-squares property of A^+ make the inverse valuable in statistics as well as in optimization theory. A big benefit is that, since a generalized inverse exists for all matrices, mathematical problem formulation is not limited to just using square, non-singular, matrices.

III. Theory and Background

Introduction

As chapter two highlighted, the theory of generalized inverses has touched a wide range of disciplines. This chapter presents the theoretical groundwork of this thesis effort, generalized inverses of multiparameter matrices. This is a new area of research, sparked by the growing complexity of modern systems. Theory regarding constant coefficient matrices falls short in solving current problems. Computationally, expert systems such as MACSYMA enable efficient manipulation of variable element matrices and provide exact answers to complex problems. Thus, with theory and computational tools available, the application of multiparameter generalized inverses can progress.

The theory presented here has emerged from the constant coefficient matrix theory. Most of the theorems have been proved elsewhere, so the source of the proof is provided as a reference. The purpose of this chapter is to provide an understanding of the theory behind generalized inverses. This insight comes from the contents of the theorem, not necessarily from the proof of that theorem. The intent then is to consolidate that theory at the very core of this research effort.

Constant Coefficient Matrices

The basic conditions a matrix must satisfy for the matrix to be a generalized inverse were set forth by Penrose

in 1954. Penrose's purpose in considering these inverses was to solve inconsistent linear equations, those involving singular and rectangular matrices, cases where classical matrix theory fell short. These conditions, now referred to as the Penrose conditions, come from the following theorem:

Theorem 3.1: (32:406) The four equations:

$$(1)\quad A\,X\,A = A \tag{3.1}$$

$$(2)\quad X\,A\,X = X \tag{3.2}$$

$$(3)\quad (A\,X)^{*} = A\,X \tag{3.3}$$

$$(4)\quad (X\,A)^{*} = X\,A \tag{3.4}$$

have a unique solution for any matrix A.

Proof: See cited reference. Numbering of the equations added for future reference.

In the course of his proof, Penrose showed that the matrix A did not necessarily have to be a square matrix. Since the classical inverse from matrix theory covers only non-singular, square matrices, Penrose said his inverse was a generalization of the notion of a matrix inverse. He called his inverse a pseudoinverse and designated it as A^{+}. Along with the proof, Penrose provided two key lemmas. These are:

Lemma 1.1: $A^{+^{+}} = A$

Lemma 1.2: If A is a non-singular matrix, then $A^{+} = A^{-1}$ (32:408).

With Lemma 1.2, Penrose tied together the notion of a generalized inverse with the classical inverse theory he sought to generalize. Thus, working with the A^{+} inverse did

in fact provide for a more general methodology than using just the A^{-1} inverse.

The next theorem, and it's associated corollaries, gave Penrose the ability to solve all types of linear equations, both consistent and inconsistent.

Theorem 3.2. (32:409) A necessary and sufficient condition for the equation

$$A \underset{\sim}{X} B = C \tag{3.5}$$

to have a solution is

$$A A^{+} C B^{+} B = C \tag{3.6}$$

in which case the general solution is

$$\underset{\sim}{X} = A^{+} C B^{+} + \underset{\sim}{Y} - A^{+} A \underset{\sim}{Y} B B^{+} \tag{3.7}$$

where Y is arbitrary.

Proof: See cited reference.

Corollary 3.2.1. The general solution of the vector equation:

$$P \underset{\sim}{X} = C \tag{3.8}$$

is

$$\underset{\sim}{X} = P^{+} C + (I - P^{+} P) \underset{\sim}{Y} \tag{3.9}$$

where $\underset{\sim}{Y}$ is arbitrary, provided that the equation has a solution.

Corollary 3.2.2. A necessary and sufficient condition for the equations:

$$A \underset{\sim}{X} = C \tag{3.10}$$

$$\underset{\sim}{X} B = D$$

to have a common solution is that each equation should individually have a solution and that

$$A\,D = C\,B \qquad (3.11)$$

Proof: See cited reference.

These then are the pertinent results from the classical work of Penrose. Each forms the basis for future work, as shown throughout the remainder of this chapter. It should be noted that soon after the publication of Penrose's work, it became evident that the four Penrose conditions were equivalent to earlier work done by Moore. In his earlier work, Moore defined the generalized inverse, G, of a matrix A as satisfying:

$$AG = P_A \qquad (3.12)$$

$$GA = P_G$$

where P_X is defined as the orthogonal projection onto the column space of the matrix X = G or A (28:xi-xii). Thus, the A^+ inverse is generally referred to as the Moore-Penrose generalized inverse (28:111).

The work of Penrose laid the groundwork for later advances in generalized inverse theory. However, to properly understand some of the later work, particularly the work of Jones and others with the ST computational method, as well as the connection between the early work of Moore and Penrose, one must first understand some fundamental ideas from linear algebra. In particular there is Strang's discussion of the four fundamental subspaces associated with a matrix and extensions of this work into the multiparameter matrix area. Strang develops these concepts in the form of two theorems, which he labels as Fundamental Theorems of

Linear Algebra. From Strang (46:75):

Fundamental Theorem of Linear Algebra, Parts I and II

1. $\mathcal{R}(A^T)$ = row space of matrix A; dimension r
2. $\eta(A)$ = nullspace of matrix A; dimension n-r
3. $\mathcal{R}(A)$ = column space of matrix A; dimension r
4. $\eta(A^T)$ = left nullspace of matrix A; dimension m-r
5. $\eta(A) = (\mathcal{R}(A^T))^{\perp}$
6. $\mathcal{R}(A^T) = \eta(A)^{\perp}$
7. $\eta(A^T) = (\mathcal{R}(A))^{\perp}$
8. $\mathcal{R}(A) = (\eta(A^T))^{\perp}$

The above theorem says that associated with any matrix A there are four fundamental subspaces, and these subspaces are related according to the orthogonal complement relationships depicted in 5 through 8. Although Strang aimed his theorem at matrices with constant coefficients, the results are just as valid for matrices defined over the polynomial field.

For the equation $A\underset{\sim}{x} = \underset{\sim}{b}$, the idea of subspaces is critical. For a solution to exist, the vector $\underset{\sim}{b}$ must lie within the column space of A. In other words, $\underset{\sim}{x}$ is a linear combination of A. Those linear combinations of $\underset{\sim}{x}$ that satisfy the homogeneous equation $A\underset{\sim}{x} = \underset{\sim}{0}$ are members of the nullspace of the matrix A. This nullspace is also referred to as the nullity or the kernel of the linear transformation provided by the matrix A (1:216).

The idea of rank of a matrix must also be understood. The rank of A is the number of linearly independent vectors that span the column (and row) space of the matrix. A more traditional definition is that the rank of the matrix A is

the dimension of the smallest nonsingular submatrix of A. If A is nxn, and the rank, r, equals n, then A has full rank. If A is mxn and m≠n, then A is not full rank. Only full rank matrices have a classical inverse, A^{-1}, and provide a unique solution vector, $\underset{\sim}{x}$, to the equation, $A\underset{\sim}{x} = \underset{\sim}{b}$.

For full rank matrices, the homogeneous equation $A\underset{\sim}{x}=\underset{\sim}{0}$ is satisfied only by the trivial solution, $\underset{\sim}{x}=0$. The nullspace consists of this single point. In singular matrices, the nullspace is of dimension n-r, or m-r, depending upon the rank of the matrix. The solutions of $A\underset{\sim}{x}=\underset{\sim}{0}$ are non-trivial and form a non-trivial subspace, the nullspace, that is orthogonal to the column space, which is of dimension r, the rank of A.

Looking at $A\underset{\sim}{x}=\underset{\sim}{b}$, a solution, $\underset{\sim}{x}$, can be found if $\underset{\sim}{b}$ is orthogonal to the nullspace and is a member of the column space. But if the equation is inconsistent then the A^{-1} inverse does not exist and the solution is no longer unique. The best solution to the problem must then be selected from among the possibly infinite number of solutions. This turns out to be the point, call it $\underset{\sim}{t}$, that is in the column space of A and is closest to the point $\underset{\sim}{b}$. This solution is deemed "best" since it is the closest solution among all possible solutions. Typically least-squares or minimum-norm criteria are used to determine the closest solution.

This is where the generalized inverse comes into play, as alluded to in Chapter II by Penrose's observation that A^{+} provides the least-squares, minimum-norm solution. The A^{+} is

a projection of $\underset{\sim}{b}$ onto the column space of A. In terms of least-squares, $\underset{\sim}{x} = (A^TA)^{-1}A\underset{\sim}{b}$, is the projection of $\underset{\sim}{b}$ onto the $\underset{\sim}{x}$ in the column space. This projection is accomplished by the $(A^TA)^{-1}A$ term, which as shown in Chapter II, can sometimes be used to find the A^+ inverse. Thus, the reasoning behind the claim that A^+ provides the minimum-norm or least-squares (closest) solution to the inconsistent problem.

Since the row space and nullspace are orthogonal complements, any solution vector consists of two portions. One portion is a projection onto the row space, the other is the projection onto the nullspace. This solution vector can thus be written as $\underset{\sim}{x} = (\underset{\sim}{x}_r + \underset{\sim}{w})$. Here $\underset{\sim}{x}_r$ is the row space component and $\underset{\sim}{w}$ is the nullspace component. Strang (46:138) points out that any solutions to $A\underset{\sim}{x}=\underset{\sim}{b}$ will share a common $\underset{\sim}{x}_r$ and differ only in the nullspace component, which is the solution to the homogeneous equation $A\underset{\sim}{x}=\underset{\sim}{0}$. This homogeneous portion can also be expressed in a general form as $\underset{\sim}{x} = (I - A^+A)\underset{\sim}{z}$, for arbitrary $\underset{\sim}{z}$ (36:23-26 ; 34:35).

Taking these ideas into account, Strang states the following conclusion, which is found embodied in the results of Corollary 3.2.1 (46:138):

> The general solution is the sum of one particular solution (in this case $\underset{\sim}{x}_r$) and an arbitrary solution $\underset{\sim}{z}$ of the homogeneous equation

Multiparameter Matrices

The first attempt to study generalized inverses of

multiparameter matrices, was by Bose and Mitra (4:491). Their motivation for delving into this new area was the study of multi-input/ multi-output control systems. These systems often require the use of matrices having elements that are not constant, but variable. Thus, Bose and Mitra sought to extend the extensive work already done for constant matrices into the multiparameter matrices defined over rings of polynomials of a single variable (42:514).

Theorem 3.3. (4:491) Any (m x n) integer matrix having rank r will have an integer matrix for its generalized inverse if and only if A can be expressed in the Smith canonical form (see Appendix A for a definition of Smith form):

$$A = M D N \tag{3.13}$$

where M and N are integer matrices with determinant equal ±1 and D is of the form:

$$D = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \tag{3.14}$$

I_r, being the identity matrix of order r.

Proof: See cited reference.

Bose and Mitra use this theorem to extend the notions to multiparameter matrices with the following theorem:

Theorem 3.4. (4:492) Any (m x n) polynomial matrix A(z), of rank r with coefficients in a number field, will have a polynomial matrix (with coefficients in the same field) for its generalized inverse if and only if A(z) can be expressed in the Smith canonical form

$$A(z) = M(z)\, D\, N(z) \qquad (3.15)$$

where $M(z)$ and $N(z)$ are polynomial matrices with determinant equal ± 1 and D is of the form:

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} \qquad (3.16)$$

D_1 being an $(r \times r)$ diagonal matrix of constants belonging to the chosen number field.

Proof: See cited reference.

Bose and Mitra use reduction to Smith normal form to characterize the "weak generalized inverse", which are those inverses satisfying Penrose conditions (1) and (2). They also addressed just the single variable, polynomial matrix case. Extended results were obtained by Sontag in June 1980 with the following theorem:

Theorem 3.5. (42:514) The following statements are equivalent for a matrix $A = A(z_1, z_2, \ldots, z_n)$ over $\mathbb{R} \in \mathbb{C}(z_1, z_2, \ldots, z_n)$:

a) A has a weak generalized inverse (WGI)

b) There exist square, unimodular (i.e. nonzero scalar determinant) matrices P and Q defined over $\mathbb{R}$ such that $A = P\, A_o\, Q$ with

$$A_o = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \qquad (3.17)$$

where I_r is the identity matrix of order $r = \text{rank}(A)$

c) As a function of the complex variables

$(z_1, z_2, \ldots, z_n)$, the rank of $A(z_1, z_2, \ldots, z_n)$ is constant.

Proof: See cited reference.

Theorem 3.6 (42:516) The following statements are equivalent for any matrix A over $\mathbb{R}$

a) A has a generalized inverse

b) A has a weak generalized inverse

c) A has constant rank over all $(z_1, z_2, \ldots, z_n)$ in $\mathbb{R}$

d) A can be written as $P\ A_o\ Q$ with P and Q unimodular $\mathbb{R}$ matrices (meaning having determinant not equal zero for all $(z_1, z_2, \ldots, z_n)$ in $\mathbb{R}$ and

$$A_o = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \qquad (3.18)$$

with I_r being the identity matrix of order r = rank(A)

Proof: See cited reference.

Sontag made two significant advances. First, the results were now extended to matrices defined over $\mathbb{R}^n$. Secondly, he showed a generalized inverse in fact forces the existence of a Smith form for the original matrix. Recall the work of Bose and Mitra where the Smith form implied existence of the generalized inverse. The question Sontag faced was how to determine the P and Q matrices that perform the necessary transformations.

Sontag's Factorization

Sontag used a full-rank factorization of the matrix A

(42:516) to derive a formula for A^+. Recall a factorization of A requires two matrices, B and C, such that $A = B C^T$ for $B \in \mathbb{C}^{mxr}$, $C \in \mathbb{C}^{nxr}$ and rank(A) = r (36:5). But clearly B and C function in the same role as do the P and Q matrices referenced by Sontag. There is still the need to determine P and Q. Another method (31:326), namely the singular valued decomposition, employed the matrix M of eigenvectors of $A A^T$ and $A^T A$. The Moore-Penrose inverse, A^+, is then computed by the formula:

$$A^+ = M \Sigma^{-1} M^{-1} \qquad (3.19)$$

where Σ^{-1} is the inverse of the matrix Σ whose diagonal elements are the eigenvalues of $A A^T$, or $A^T A$. The eigenvalues of $A A^T$ and $A^T A$ are equivalent.

ST Method

The easiest method of computation involves use of the ST canonical form, along with extensions of work from Sontag and classical linear algebra, to determine the P and Q matrices while reducing the A matrix to it's Smith form. This method is the ST method of Jones (7:3-4 ; 27:vi). Extensive detail of how the ST method is implemented can be found in a recent AFIT thesis by Murray. Although Murray considered only the constant coefficient case, the technique remains the same for reducing multiparameter matrices. A brief explanation of the technique is followed by the underlying theorems.

Consider $A \in \mathbb{C}^{mxn}$. Augment A with identity matrices below and to the right to obtain the following form:

$$\begin{bmatrix} A & I_m \\ I_n & 0 \end{bmatrix} \tag{3.20}$$

where $I_n \in \mathbb{C}^{nxn}$, and $I_m \in \mathbb{C}^{mxm}$. This is referred to as the initial ST canonical form.

Reduce the A matrix to the identity matrix, I_r, where the dimensions of the identity matrix, rxr, are equal to the rank of the original A matrix. Any row operations performed in the reduction are carried out on the augmented matrix to the left. Similarly, any column operations are carried out on the augmented matrix directly below the original matrix. Once the A matrix has been reduced to it's identity form, the augmented form is now in the final ST canonical form:

$$\left[\begin{array}{c|c||c} I_r & 0 & T \\ \hline 0 & 0 & M \\ \hline S & N & \end{array}\right] \tag{3.21}$$

All that is required to accomplish this initial reduction are elementary transformations, commonly known as elementary row and column operations. Whether the matrices involved are constant coefficient or multiparameter matrices, the elementary transformations remain the same. Appendix A contains the definition of these operations.

If A is full row rank, the M submatrix will not exist. If A is full column rank, the N submatrix will not exist. If A is both full row and column rank, then A has a A^{-1} and a trivial nullspace, the point zero. In this case, neither the M nor the N submatrices will exist in the final ST canonical

form. The rank of A, rank(A) = r, is determined as a result of the elementary transformations used to reduce the augmented form (3.20) to the reduced canonical form (3.21).

From the form of (3.21), the P and Q matrices required by Sontag can be read off directly. These matrices are:

$$P = \left[\begin{array}{c} T \\ \hline M \end{array}\right] \qquad Q = \left[\begin{array}{c|c} S & N \end{array}\right] \qquad (3.22)$$

A quick check of a reduced matrix will verify that the product $P\,A\,Q = I_r$ does indeed hold.

From the form of (3.21), all the generalized inverses of the matrix A may be generated. The next set of theorems prove this in addition to proving the validity of the above reduction technique.

Theorem 3.7. (9:23 ; 27:16) For any given matrix $A \in \mathbb{C}^{m \times n}$, there exist two nonsingular matrices $P \in \mathbb{C}^{m \times m}$ and $Q \in \mathbb{C}^{n \times n}$ such that

$$\left[\begin{array}{cc} A & I_m \\ I_n & 0 \end{array}\right] \quad \text{and} \quad \left[\begin{array}{c|c|c} I_r & 0 & T \\ \hline 0 & 0 & M \\ \hline S & N & \end{array}\right] \qquad (3.23)$$

are equivalent.

Proof: See cited reference.

Theorem 3.7 takes the results of Sontag's work (Theorem 3.6) and incorporates it into a computational technique. The P and Q matrices of the form given in (3.22) are the matrices required by Sontag and provide the weak generalized inverse of the original matrix A. However, the real strength of the ST method is embodied in extending the P and Q

submatrices in accordance with the following theorems.

Theorem 3.8. (9:24) From the matrices defined in theorem 3.7, an $A_{1,2}$ matrix (weak generalized inverse) is determined by the product of the submatrices S and T.

Proof: See cited reference.

This means that the WGI of a matrix A is attainable simply through elementary row and column operations performed on the augmented form given by (3.20). Higher generalized inverses are obtained using properties of orthogonality. In particular, the rows (columns) of M (N) are made orthogonal to the rows (columns) of T (S).

Theorem 3.9. (9:25-27 ; 27:33-35) From the matrices defined in theorem 3.7, if the condition $T M^T = 0$ holds then an $A_{1,2,3}$ matrix generalized inverse is defined by the product S T.

Proof: See cited reference.

Theorem 3.10. (9:28; 27:33-36) From the matrices defined in theorem 3.7, if the condition $N^T S = 0$ holds then an $A_{1,2,4}$ matrix generalized inverse is defined by the product S T.

Proof: See cited reference.

The final theorem in this set of theorems comes from the work of Doma and Murray, who combine the previous two theorems to provide the conditions under which to produce the unique Moore-Penrose, generalized inverse, the $A_{1,2,3,4}$ or simply the A^+.

Theorem 3.11. (9:29 ; 27:33-41) From the matrices

defined in Theorem 3.7, if the conditions $T M^T = 0$ and $N^T S = 0$ hold, then the product ST defines the A^+ matrix generalized inverse.

Proof: See cited reference.

The ST technique can be summarized by the schematic in Figure 3. The matrix A gives rise to the initial canonical form by augmenting A with identity matrices below and to the right. Through elementary transformations and orthogonalizations, the initial canonical form is transformed into the final canonical form. During the transformation process, each of the $A_{1,2}$, $A_{1,2,3}$, $A_{1,2,4}$, and the $A_{1,2,3,4}$ generalized inverses can be computed.

In Corollary 3.2.1, Penrose gave a general form for the solution of the $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$ equation. The geometry of this solution form was then briefly discussed, based in large part upon Strang's work. Presented here as a corollary is a result of the subspace concept and the ST reduction process.

Corollary 3.11.1. (9:39 ; 27:25 ; 7:40 ; 19:463) The equation $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$ has a solution $\underset{\sim}{x}$ if and only if $M\underset{\sim}{b} = \underset{\sim}{0}$. In this case the general solution is given by:

$$\underset{\sim}{x} = (ST)\underset{\sim}{b} + N\underset{\sim}{z} \qquad (3.24)$$

where the matrix $\underset{\sim}{z}$ is arbitrary, and the S, T, M, and N matrices come from the final ST canonical form as shown in (3.23).

A final point regarding the general power of the ST computational technique is the strong interface it has with the Fundamental Theorem of Linear Algebra previously

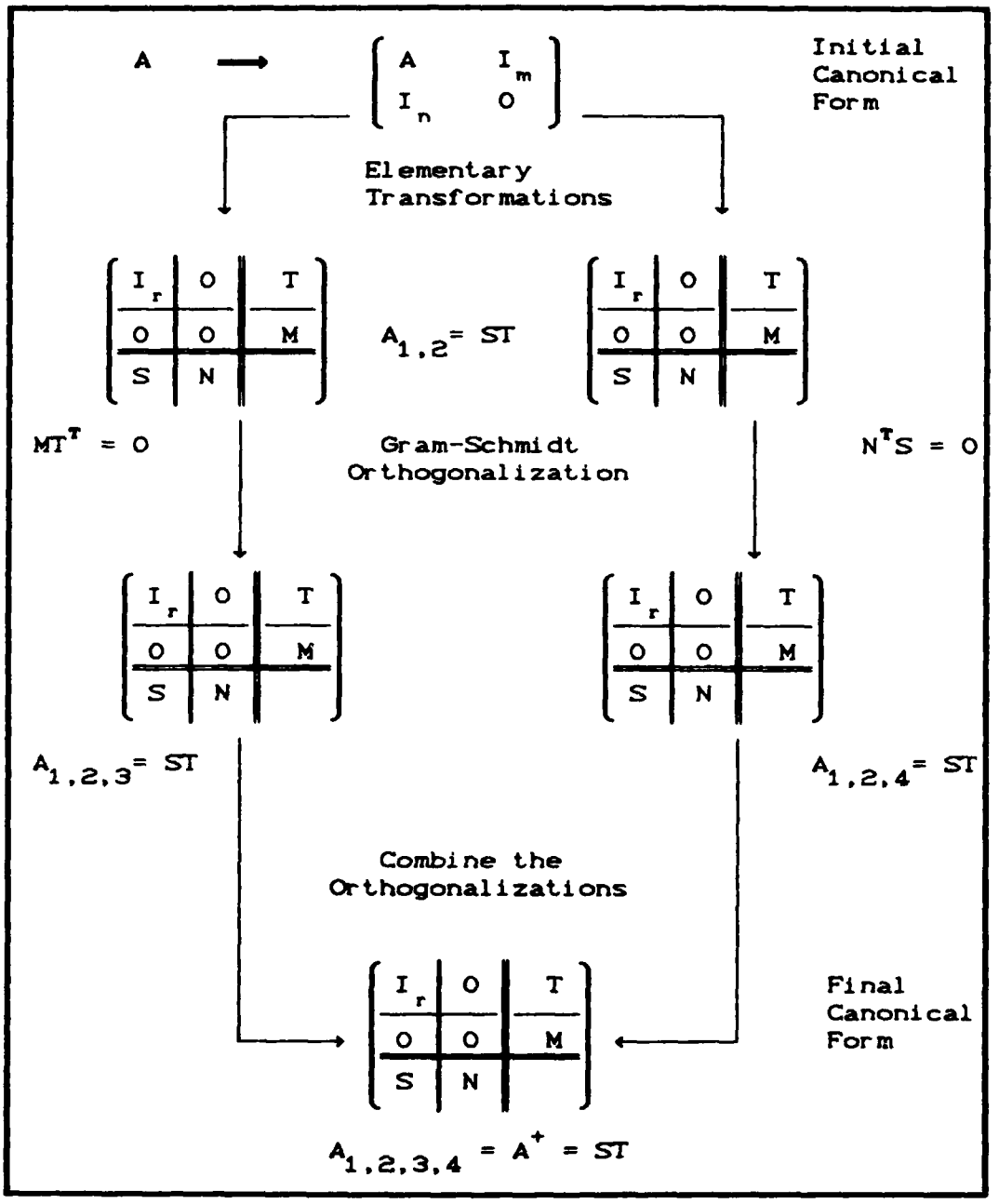

FIGURE 3. ST TECHNIQUE SCHEMATIC

discussed. Once fully reduced, the final ST canonical form provides:

- the rank of the matrix A (dimension of I_r)
- a basis for the column space, R(A), of A given in

the S submatrix

- a basis for the row space, $\mathcal{R}(A^T)$, of A given in the T submatrix
- a basis for the nullspace, $\eta(A)$, of A in submatrix N
- a basis for the left nullspace, $\eta(A^T)$, of A in the submatrix M

Each of the above are byproducts of the elementary transformations and orthogonalizations performed to determine the generalized inverses of a given matrix A.

Common Solutions of Sets of Equations

The final topic addressed in this chapter involves necessary and sufficient conditions for common solutions of matrix equations. These sets of equations arise in many applications, for instance, network design problems or critical path systems. In the case of constant coefficients and multiparameter systems, parallel processing techniques can be exploited to determine solutions to systems more efficiently. Current hardware and software technology limit the parallel processing applications for the multiparameter case, but this section shows that the theory is in place.

Common solutions to sets of multiparameter matrix equations have been extended from the work done on constant coefficient matrix equations. However, the details of the theorem presented are provided for the first time. Previous work by Morris and Odell (26) and then by Jones (17) left these details out.

Mitra followed Penrose's original work with an

extension to the common solution of two matrix equations (24:213). In particular, Mitra used corollary 3.2.2 to prove the following:

Theorem 3.12. (24:214) Let A_1, A_2, B_1, and B_2 be non-negative definite matrices. A necessary and sufficient condition for the consistent equations:

$$\begin{aligned} A_1 X B_1 &= C_1 \\ A_2 X B_2 &= C_2 \end{aligned} \tag{3.25}$$

to have a common solution is

$$A_1(A_1+A_2)^- C_2(B_1+B_2)^- = A_2(A_1+A_2)^- C_1(B_1+B_2)^- B_2 \tag{3.26}$$

in which case the general solution is

$$\begin{aligned} X = (A_1+A_2)^- (C_1 + Y + Z + C_2)(B_1 + B_2)^- + U - \\ (A_1 + A_2)^- (A_1 + A_2) U (B_1 + B_2)(B_1 + B_2)^- \end{aligned} \tag{3.27}$$

where U is arbitrary, Y and Z are arbitrary matrices satisfying respectively the equations

$$\begin{aligned} A_2(A_1 + A_2)^- Y &= A_1(A_1 + A_2)^- C_2 \\ Y(B_1 + B_2)^- B_1 &= C_1(B_1 + B_2)^- B_2 \end{aligned} \tag{3.28}$$

and

$$\begin{aligned} A_1(A_1 + A_2)^- Z &= A_2(A_1 + A_2)^- C_1 \\ Z(B_1 + B_2)^- B_2 &= C_2(B_1 + B_2)^- B_1 \end{aligned} \tag{3.29}$$

where the $()^-$ notation denotes an A_1 matrix generalized inverse.

Proof: See cited reference.

This proof provides an expression for the general common solution. However, Mitra's work was limited to the case of n=2 matrix equations. A generalization to n constant coefficient matrix equations comes from the work of Morris

and Odell in 1968 (26). This was then extended to multiparameter matrices, defined over the ring of polynomial elements, by Jones (17) in 1987. Since both theorems are similar in content, only the latter is presented.

Theorem 3.13. (17:768) Let $A_i \in \mathbb{C}^{pxr}$ and $B_i \in \mathbb{C}^{pxr}$ for $i = 1,\ldots,m$. Define the following relationships:

$$C_1 = A_1 \qquad D_1 = B_1 \qquad (3.30)$$

$$E_1 = A_1^- B_1 \qquad F_1 = I - A_1^- A_1$$

and

$$C_k = A_k F_{k-1} \qquad D_k = B_k - A_k E_{k-1}$$

$$E_k = E_{k-1} + F_{k-1} C_k^- D_k \qquad F_k = F_{k-1}(I - C_k^- C_k)$$

Then $A_i \underset{\sim}{x} = B_i$, for $i=1,\ldots,m$, has a common solution if and only if $C_i C_i^- D_i = D_i$ for $i=1,\ldots,m$. In this case the general common solution is given by

$$\underset{\sim}{x} = E_n + F_n \underset{\sim}{z} \qquad (3.31)$$

where $\underset{\sim}{z}$ is arbitrary.

Proof: See cited references for general proof. Since neither reference provides the explicit proof, this detailed proof is presented in Appendix B. The proof is a double induction proof in that both the conditions for existence of common solutions and the defining relationships for those solutions are proved using inductive methods.

Applications

These last two theorems, 3.12 and 3.13, provide some powerful applications. For instance, examine the following problem:

$$A \underset{\sim}{x} = B \qquad (3.32)$$

for $A \in \mathbb{C}^{pxq}$

$B \in \mathbb{C}^{pxr}$

If each row of the matrix equation is treated as a separate matrix equation, then the results of Theorem 3.13 applies to the resulting system of p matrix equations (26:273). This system may then be solved on a parallel processing implementation greatly reducing the processing.

Conclusion

This chapter has consolidated much of the theoretical knowledge in the area of generalized inverses of matrices. The intent has been to provide a readable, yet thorough, presentation of the underlying foundations for this thesis effort. The trend towards multiparameter matrices has been clearly defined and explained. The extensions of constant coefficient matrix theory to sets of equations provides a promising area of research involving parallel processing. However, it is the multiparameter trend that is the focus of this thesis. Thus, in the next chapter, some particular examples are selected and solved using the generalized inverses of multiparameter matrices.

IV. Applications

Introduction

While the previous chapter laid the theoretical groundwork for this thesis, this chapter addresses how to use the generalized inverse of a multiparameter matrix as a tool to solve problems arising in optimization. The history of the generalized inverse supports the trend towards more work involving multiparameter matrices as a necessity to keep pace with the ever increasing complexity of today's problems. The increasing capabilities of modern computer systems allow researchers to investigate, and solve, problems that previously took months to solve by hand (50:2). Before any computer solution can be implemented however, the theory and technique must be thoroughly established.

Since the generalized inverse plays a key role in a diverse range of disciplines, a small cross-section has been selected. However, the techniques employed are generally applicable to many other areas. In addition, a couple of "counter-examples" are provided in examples 8 and 9. These are labeled counter-examples since the generalized inverse technique does not provide a value for the optimal solution. However, though the optimal solution is not found, valuable information regarding the function can be obtained from the general form of the solution. The two examples, Kantorovich's function and a Lagrange multiplier problem, demonstrate this aspect of the generalized inverse technique

in nonlinear optimization.

Computing the Inverse

The previous chapter discussed the ST technique for computing generalized inverses of matrices. Before demonstrating specific optimization examples, it is best to detail the workings of the technique. The purpose is to demonstrate the applicability of the technique to multiparameter matrices, and to demonstrate the steps in the algorithm. Later examples leave out much of the computational detail to conserve space and enhance readability of this report.

Example 1. (21) Compute the generalized inverses of the following multiparameter matrix:

$$A = \begin{bmatrix} x & x^2+1 \\ x^2y & x^3y + xy \end{bmatrix} \qquad (4.1)$$

Augment this matrix with 2x2 identity matrices below and to the right to obtain the initial ST canonical form:

$$\begin{bmatrix} A & I_2 \\ I_2 & \end{bmatrix} \longrightarrow \left[\begin{array}{cc|cc} x & x^2+1 & 1 & 0 \\ x^2y & x^3y + xy & 0 & 1 \\ \hline 1 & 0 & & \\ 0 & 1 & & \end{array}\right] \qquad (4.2)$$

The A matrix portion of equation (4.2) must now be reduced to an I_r identity matrix. The ST technique requires the use of elementary transformations to accomplish this task. As yet r, the rank of the matrix A, is undetermined, but is computed through the reduction process carried out on

the A matrix. Recall from the discussion last chapter that any row operations are carried over to the matrix augmented on the right. In a similar fashion, any column operations are carried out on the matrix augmented below the matrix A.

The reduction starts by multiplying the first row by the polynomial (-xy), and adding the resulting row to the second row of the matrix. This causes equation (4.2) to transform to:

$$\longrightarrow \left[\begin{array}{cc|cc} x & x^2+1 & 1 & 0 \\ 0 & 0 & -xy & 1 \\ \hline 1 & 0 & & \\ 0 & 1 & & \end{array}\right] \qquad (4.3)$$

The next step is to multiply column one by the polynomial (-x), and add the resulting column to column two. This operation results in the matrix:

$$\longrightarrow \left[\begin{array}{cc|cc} x & 1 & 1 & 0 \\ 0 & 0 & -xy & 1 \\ \hline 1 & -x & & \\ 0 & 1 & & \end{array}\right] \qquad (4.4)$$

The upper left element of (4.4) must equal 1. The easiest transformation is to simply interchange columns one and two. Once interchanged, the final transformation, which completes the matrix reduction, is to multiply the new column one by the polynomial (-x) and add the result to column two. These last two operations produce this final

matrix:

$$\longrightarrow \left[\begin{array}{c|c||cc} 1 & 0 & 1 & 0 \\ \hline 0 & 0 & -xy & 1 \\ \hline -x & 1+x^2 & & \\ 1 & -x & & \end{array}\right] \begin{array}{l} \\ \\ \\ \\ \end{array} \begin{array}{l} \\ \rangle T \\ \rangle M \\ \\ \end{array} \qquad (4.5)$$

$$\begin{array}{cc} S & N \end{array}$$

The final result of the elementary transformations producing (4.5) is displayed with the S, N, T, and M submatrices appropriately labeled. Note the rank of the matrix A is one since I_r is of dimension one. From this form, the computations of Theorem 3.8 from the previous chapter produce the following:

$$A_1 = A_2 = A_{1,2} = ST = \begin{bmatrix} -x \\ 1 \end{bmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{bmatrix} -x & 0 \\ 1 & 0 \end{bmatrix} \qquad (4.6)$$

This may be verified by using Penrose conditions (1) and (2) of Theorem 3.1. To obtain the $A_{1,2,3}$ inverse, Theorem 3.9 must be used. This theorem requires the orthogonality of the vectors comprising the T and M matrices labeled in (4.5). This may be accomplished using a modified Gram-Schmidt process. This process produces an updated canonical form matrix, shown here:

(4.7)

$$\longrightarrow \left[\begin{array}{c|c||cc} 1 & 0 & \dfrac{1}{x^2y^2+1} & \dfrac{xy}{x^2y^2+1} \\ \hline 0 & 0 & -xy & 1 \\ \hline -x & 1+x^2 & & \\ 1 & -x & & \end{array}\right] \begin{array}{l} \\ \rangle T \\ \rangle M \\ \\ \end{array}$$

$$\begin{array}{cc} S & N \end{array}$$

From the final ST canonical form provided in (4.7), $A_{1,2,3}$, the product of the S and T submatrices, is found as a result of the following computations:

$$A_{1,2,3} = ST = \begin{bmatrix} -x \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{x^2y^2+1} & \frac{xy}{x^2y^2+1} \end{bmatrix} = \tag{4.8}$$

$$= \frac{1}{(x^2y^2+1)} \begin{bmatrix} -x & -x^2y \\ 1 & xy \end{bmatrix}$$

To obtain the $A_{1,2,4}$ generalized inverse, Theorem 3.10 requires orthogonality between submatrices N and S. Once again the Gram-Schmidt process is used to produce the matrix given:

$$\longrightarrow \left[\begin{array}{c|c|cc} 1 & 0 & 1 & 0 \\ \hline 0 & 0 & -xy & 1 \\ \hline \frac{x}{(x^2+1)^2+x^2} & \frac{x^5+3x^3+2x}{(x^2+1)^2+x^2} & & \\ \frac{x^2+1}{(x^2+1)^2+x^2} & \frac{-4x^4-2x^2}{(x^2+1)^2+x^2} & & \end{array}\right] \tag{4.9}$$

From the submatrices in (4.9), the $A_{1,2,4}$ generalized inverse is found to be:

$$A_{1,2,4} = ST = \begin{bmatrix} \frac{x}{(x^2+1)^2+x^2} \\ \frac{x^2+1}{(x^2+1)^2+x^2} \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{4.10}$$

$$= \begin{bmatrix} \frac{x}{(x^2+1)^2+x^2} & 0 \\ \frac{x^2+1}{(x^2+1)^2+x^2} & 0 \end{bmatrix}$$

Finally, the results of making M orthogonal to T and S orthogonal to N are combined in order to produce the Moore-Penrose, the $A_{1,2,3,4}$, or simply the A^+ generalized inverse. The final canonical form used is:

$$\left[\begin{array}{c|c||cc} 1 & 0 & \dfrac{1}{x^2y^2+1} & \dfrac{xy}{x^2y^2+1} \\ \hline 0 & 0 & -xy & 1 \\ \hline\hline \dfrac{x}{(x^2+1)^2+x^2} & \dfrac{x^5+3x^3+2x}{(x^2+1)^2+x^2} & & \\ \dfrac{x^2+1}{(x^2+1)^2+x^2} & \dfrac{-4x^4-2x^2}{(x^2+1)^2+x^2} & & \end{array}\right] \quad (4.11)$$

Computing the product of the above S and T submatrices will produce the A^+ generalized inverse. This A^+ inverse is:

$$\left[\begin{array}{cc} \dfrac{x}{(x^2y^2+1)[(x^2+1)^2+x^2]} & \dfrac{x^2y}{(x^2y^2+1)[(x^2+1)^2+x^2]} \\ \dfrac{x^2+1}{(x^2y^2+1)[(x^2+1)^2+x^2]} & \dfrac{xy(x^2+1)}{(x^2y^2+1)[(x^2+1)^2+x^2]} \end{array}\right] \quad (4.12)$$

This then is a detailed example of how the ST method can be used to sequentially compute representatives of all the generalized inverses of a matrix as well as the unique Moore-Penrose generalized inverse. The explicit detail provided in example 1 is excluded from the remaining examples. The reason for excluding most of the computational details from next nine examples is that the examples come from various areas of optimization theory and the focus of this work is on finding the general solution using the generalized inverse, not on the mechanics of computing the

inverses.

Nonlinear Unconstrained Optimization

Example 2: Rosenbrock's function. (21 ; 22).

Rosenbrock's function, also referred to as Rosenbrock's banana-valley function (44:41), was specifically devised as a challenge to gradient-based optimization methods and has become a test function for testing computer-based algorithms. As such it is often used in comparison studies of optimization techniques (44 ; 38:120-126). The function possesses a steep sided valley, nearly parabolic in shape, and is defined as:

$$f(x,y) = 100(y - x^2)^2 + (1-x)^2 \qquad (4.13)$$

The maximum or minimum point(s) of this function are those points for which the partial derivative of the function evaluates to zero. Whether or not the function has a minimum or a maximum depends upon the value of the Hessian matrix at a particular stationary point. Expanding the function (4.13) produces:

$$f(x,y) = 100y^2 - 200x^2y + 100x^4 + 1 - 2x + x^2 \qquad (4.14)$$

and determining the partial derivative, $\partial f/\partial x$ and $\partial f/\partial y$ produces the system of equations:

$$\begin{cases} 400x^3 - 400xy + 2x = 2 \\ -200x^2 + 200y = 0 \end{cases} \qquad (4.15)$$

which may then be written in the matrix form:

$$A(x,y) \begin{bmatrix} x \\ y \end{bmatrix} = B(x,y) \qquad (4.16)$$

where

$$A(x,y) = \begin{bmatrix} 400x^2 + 2 & -400x \\ -200x & 200 \end{bmatrix} \qquad (4.17)$$

and

$$B(x,y) = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \qquad (4.18)$$

It isn't known whether or not the A matrix is singular, but using the generalized inverse technique does not restrict the computations. Forming the initial ST canonical form as shown in example 1 and performing elementary row and column operations produces the following:

(4.19)

$$\left[\begin{array}{cc|cc} 400x^2+2 & -400x & 1 & 0 \\ -200x & 200 & 0 & 1 \\ \hline 1 & 0 & & \\ 0 & 1 & & \end{array}\right] \rightarrow \left[\begin{array}{cc|cc} 2 & -400x & 1 & 0 \\ 0 & 200 & 0 & 1 \\ \hline 1 & 0 & & \\ x & 1 & & \end{array}\right]$$

$$\left[\begin{array}{cc|cc} 0 & 2 & 1 & 0 \\ 200 & 0 & 0 & 1 \\ \hline 200x & 1 & & \\ 1+200x^2 & x & & \end{array}\right] \rightarrow \left[\begin{array}{cc|cc} 1 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 1/200 \\ \hline 1 & 200x & & \\ x & 1+200x^2 & & \end{array}\right]$$

The operations performed, for the interested reader to verify, were:

(1) multiply column two by (x) and add to column one

(2) multiply column one by (200x) and add to column two

(3) interchange columns one and two, divide row one by the scalar 2, and divide row two by 200.

From this final form, the A^+ inverse can be computed. Note that since the rank of A is 2, A has full rank meaning the A^+ inverse is also the A^{-1} inverse. This inverse is:

$$A^+ = ST = \begin{bmatrix} 1 & 200x \\ x & 1+200x^2 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1/200 \end{bmatrix} = \begin{bmatrix} 1/2 & x \\ x/2 & 1/200 + x^2 \end{bmatrix} \quad (4.20)$$

The solution of the system of equations given in (4.15) is therefore given by:

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^+B(x,y) = \begin{bmatrix} 1/2 & x \\ x/2 & 1/200 + x^2 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ x \end{bmatrix} \quad (4.21)$$

which in turn implies that the solution vector (x,y) equals (1,1). The Hessian, evaluated at the point (1,1) is:

$$\begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} \quad (4.22)$$

which is positive definite, indicating that the stationary point (1,1) is in fact a minimum of the Rosenbrock function (44:41).

Implicit Function Theory

An equation of the form F(x,y) = 0 represents a relationship between the variables x and y. The set of points that satisfy the relationship is called the locus of the equation. The behavior of the function in the vicinity of some of these points, it's local behavior, is of concern in analytical geometry and optimization. The study of this local behavior is often conducted as a result of first finding the variable y as a function of x, or say x as a function of y. Through the use of an implicit function, a complex functional system of the form:

$$\begin{cases} F(x_1, x_2, x_3, x_4) = 0 \\ G(x_1, x_2, x_3, x_4) = 0 \end{cases} \tag{4.23}$$

can be simplified into the form of:

$$\begin{cases} F(x_1, x_2, f(x_1, x_2), g(x_1, x_2)) = 0 \\ G(x_1, x_2, f(x_1, x_2), g(x_1, x_2)) = 0 \end{cases} \tag{4.24}$$

where $x_3 = f(x_1, x_2)$ and $x_4 = g(x_1, x_2)$ are the implicit functions obtained from (4.23) (33:479-510).

In optimization, implicit function theory can be used to eliminate variables from systems, thereby reducing the dimension of the problem. Implicit function theory is also found in implicit differentiation. In this technique, an implicit function of one variable is found in terms of it's partial derivative. Once found, the resulting system of equations can be solved simultaneously for the maximum or minimum of the system (3:172-173).

Example 3 (33:141-149). Suppose the common solution to the following system of equations is sought:

$$\begin{cases} F(x,y,u,v) = x^2 + 2xy - 3xu + 4yv = 0 \\ G(x,y,u,v) = 4xy + x^3u - 8yv + 2 = 0 \end{cases} \tag{4.25}$$

The above system may be written in the now familiar matrix form of $A\underset{\sim}{x} = \underset{\sim}{b}$ as the following:

$$\begin{bmatrix} x & 2x & -3x & 4y \\ 0 & 4x & x^3 & -8y \end{bmatrix} \begin{bmatrix} x \\ y \\ u \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix} \tag{4.26}$$

where this particular representation for A matrix is not necessarily unique.

This system may be solved for the general solution using the generalized inverse of the A matrix.

$$\left[\begin{array}{cccc|cc} x & 2x & -3x & 4y & 1 & 0 \\ 0 & 4x & x^3 & -8y & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & & \\ 0 & 1 & 0 & 0 & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \end{array}\right] \rightarrow \left[\begin{array}{cccc|cc} x & 2x & -3x & 4y & 1 & 0 \\ 0 & 4x & x^3 & -8y & 0 & 1 \\ \hline 1 & -2 & 3 & 0 & & \\ 0 & 1 & 0 & 0 & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \end{array}\right]$$

$$\left[\begin{array}{cccc|cc} x & 0 & 0 & 4y & 1 & 0 \\ 0 & 4x & 0 & -8y & 0 & 1 \\ \hline 1 & -2 & 3+x^2/2 & 0 & & \\ 0 & 1 & -x^2/4 & 0 & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \end{array}\right] \rightarrow \left[\begin{array}{cccc|cc} x & 0 & 0 & 4y & 1 & 0 \\ 2x & 4x & 0 & 0 & 2 & 1 \\ \hline 1 & -2 & 3+x^2/2 & 0 & & \\ 0 & 1 & -x^2/4 & 0 & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \end{array}\right]$$

$$\left[\begin{array}{cccc|cc} x & 0 & 0 & 4y & 1 & 0 \\ 0 & 4x & 0 & 0 & 2 & 1 \\ \hline 2 & -2 & 3+x^2/2 & 0 & & \\ -1/2 & 1 & -x^2/4 & 0 & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \end{array}\right] \rightarrow \left[\begin{array}{cc|cc|cc} 1 & 0 & 0 & 4y/x & 1/x & 0 \\ 0 & 1 & 0 & 0 & 2/4x & 1/4x \\ \hline 2 & -2 & 3+x^2/2 & 0 & & \\ -1/2 & 1 & -x^2/4 & 0 & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \end{array}\right]$$

$$\left[\begin{array}{cc|cc|cc} 1 & 0 & 0 & 0 & 1/x & 0 \\ 0 & 1 & 0 & 0 & 2/4x & 1/4x \\ \hline 2 & -2 & 3+x^2/2 & -8y/x & & \\ -1/2 & 1 & -x^2/4 & 2y/x & & \\ 0 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 1 & & \end{array}\right] = \left[\begin{array}{cc|c} I_r & 0 & T \\ \hline 0 & 0 & \\ \hline S & N & \end{array}\right] \qquad (4.27)$$

For the interested readers, the elementary transformations used to reduce the initial ST canonical form to the final ST canonical form in the above sequence of matrices were:

(1) multiply column 1 by (-2) and (3) and add to

columns 2 and 3 respectively

(2) multiply column 2 by $(-x^2/4)$ and add to column 3

(3) multiply row 1 by (2) and add to row 3

(4) multiply column 2 by $(-1/2)$ and add to column 1

(5) divide row 1 by (x) and row 2 by $(4x)$

(6) multiply column 1 by $(-4y/x)$ and add to column 4

In this case the $A_{1,2}$ is obtained by the product of the S and T submatrices above. The conditions, (ST)A(ST) = (ST) and A(ST)A = A may be verified as holding true. In this case the general solution of the equation (4.26) is given by the following form:

$$\underset{\sim}{x} = A_{1,2}\underset{\sim}{b} + N\underset{\sim}{z} \qquad (4.28)$$

where $\underset{\sim}{z}$ is an arbitrary vector. This general form generates all solutions to equation (4.26) by appropriate choice of values for $\underset{\sim}{z}$. This solution is carried out in the following equations:

$$\underset{\sim}{x} = \begin{bmatrix} 1/x & -1/2x \\ 0 & 1/4x \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix} + \begin{bmatrix} 3+x^2/2 & -8y/x \\ -x^2/4 & 2y/x \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \qquad (4.29)$$

for arbitrary z_1 and z_2, elements of $\underset{\sim}{z}$, of (4.28) above. So the final general solution is:

$$\underset{\sim}{x} = \begin{bmatrix} 1/x \\ -1/2x \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} (3+x^2/2)z_1 - (8y/x)z_2 \\ (-x^2/4)z_1 + (2y/x)z_2 \\ z_1 \\ z_2 \end{bmatrix} \qquad (4.30)$$

Clearly, x cannot assume a zero value as the solutions

are undefined for that value. Since $\underset{\sim}{z}$ is an arbitrary vector, selecting $\underset{\sim}{z}$= (0,0) yields a specific solution to the system, namely, x=1, y= -1/2, u=0, and v=0.

In more classical implicit function settings, F(x,y,u,v) would have been solved for u, and the resulting expression substituted into G(x,y,u,v). The resulting equation would be solved for v to obtain v=g(x,y) as discussed above. This implicit function would then be used to obtain u=f(x,y) again as discussed above. For this particular problem, this process yields (33:490):

$$\begin{cases} u = f(x,y) = \dfrac{(x^2+2xy)(6y-x^2) + (x^4+2x^3y + 12xy +6)y}{3x(6y - x^2)} \\ v = g(x,y) = \dfrac{x^4 + 2x^3y + 12xy + 6}{24y - 4x^2y} \end{cases} \qquad (4.31)$$

These solutions from implicit function theory agree with the results obtained using the generalized inverse of the multiparameter system of equations. Choose x=1 and y= -1/2 and equation (4.31) yields values of u = v = 0.

Example 4. Given the system of homogeneous equations:

$$\begin{cases} x + y + z = 0 \\ x^2 + y^2 + z^2 + 2xz - 1 = 0 \end{cases} \qquad (4.32)$$

show whether x and y can be considered as functions of the variable z. One particular method of solving this problem is to use Jacobian Determinants. As an alternative approach, consider the use of the generalized inverse using a formulation of the form $A\underset{\sim}{x} = \underset{\sim}{b}$ as in the following:

$$\begin{bmatrix} 1 & 1 & 1 \\ x + 2z & y & z \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad (4.33)$$

From equation (4.33), the A matrix can be placed into the initial ST canonical form. This initial form can then be reduced, using only a series of elementary transformations, to the following final ST canonical form.

(4.34)

$$\left[\begin{array}{cc|c|cc} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -(x+2z)/(y-x-2z) & 1/(y-x-2z) \\ \hline 1 & -1 & (z-y)/(y-x-2z) & & \\ 0 & 1 & (x+z)/(y-x-2z) & & \\ 0 & 0 & 1 & & \end{array}\right]$$

From this final form, the product of the S and T submatrices produce the $A_{1,2}$ generalized inverse. This inverse:

$$ST = A_{1,2} = \begin{bmatrix} y/(y-x-2z) & -1/(y-x-2z \\ -(x+2z)/(y-x-2z) & 1/(y-x-2z) \\ 0 & 0 \end{bmatrix} \qquad (4.35)$$

satisfies the consistency condition for the existence of a solution to (4.33), namely $A\ A_{1,2}\underset{\sim}{b} = \underset{\sim}{b}$, and can therefore be used to obtain the general solution as given by the equation, $\underset{\sim}{x} = A_{1,2}\underset{\sim}{b} + [I - A_{1,2}A]\underset{\sim}{w}$, where $\underset{\sim}{w}$ is an arbitrary vector. The final form of the general solution, after computing the products and simplifying the expressions is:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} (-1+yw_3 + zw_3) / (y-x-2z) \\ (1 + xw_3 + zw_3)/ (y-x-2z) \\ w_3 \end{bmatrix} \qquad (4.36)$$

Now equation (4.36) is the general form of the solution to equation (4.33). This implies that all solutions of (4.33) can be generated by (4.36), through appropriate choice of $\underset{\sim}{w}$, in particular just the w_3 component of $\underset{\sim}{w}$. To find a solution to this problem, simply choose $w_3 = z$ and the general solution becomes:

$$\begin{cases} x = (-1 + yz + z^2) / (y - x - 2z) \\ y = (1 + xz + z^2) / (y - x - 2z) \end{cases} \qquad (4.37)$$

with z free to take on all values except zero. Thus, x and y can be expressed as functions of z.

Nonlinear Constrained Optimization

In the unconstrained section, use was made of the generalized inverse to solve the system of equations arising when the partial derivatives of the function were set equal to zero (i.e. to find the stationary points). In this section, the problem is that of constrained optimization. The type of problems addressed involve objective functions of higher order than quadratic, and concave constraints. The particular technique used is a generalization of the quadratic programming technique of Nelson discussed in chapter II.

Example 5. Consider the following problem (38:320):

$$\begin{aligned} \max \quad & x_1^4 + x_2 \qquad (4.38) \\ \text{s.t.} \quad & 2x_1^2 + 3x_2 \leq 9 \\ & x_1 , x_2 \geq 0 \end{aligned}$$

Since the unconstrained maximum of the objective function is unbounded (i.e. infinite), the constraint in the

problem must be a binding constraint. The constraint can therefore be rewritten as an equality constraint, and in the $A\underset{\sim}{x} = \underset{\sim}{b}$ format as:

$$\begin{bmatrix} 2x_1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 9 \end{bmatrix} \qquad (4.39)$$

For the A matrix in (4.39), a representative of the $A_{1,2}$ class of generalized inverses can be obtained after producing and reducing the ST canonical form in the following manner:

$$\left[\begin{array}{cc|c} 2x_1 & 3 & 1 \\ \hline 1 & 0 & \\ 0 & 1 & \end{array}\right] \rightarrow \left[\begin{array}{cc|c} 1 & 0 & 1/2x_1 \\ \hline 1 & -3/2x_1 & \\ 0 & 1 & \end{array}\right] \qquad (4.40)$$

from which $(A_{1,2})^T = (ST)^T = (1/(2x_1 , 0)$. The solution to the problem, in the general form, is found using the equation $\underset{\sim}{x} = A_{1,2}\underset{\sim}{b} + (I - A_{1,2}A)\underset{\sim}{z}$, where $\underset{\sim}{z}$ is arbitrary. This solution is:

$$\underset{\sim}{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (9 - 3z_2) / (2x_1) \\ z_2 \end{bmatrix} \qquad (4.41)$$

where $z_2 \in \underset{\sim}{z}$. A particular solution, obtained by choosing $z_2 = 0$, is $\underset{\sim}{x} = (9/(2x_1) , 0)$. Using this particular solution in the constraint gives the boundary point as $\underset{\sim}{x} = (\sqrt{4.5}, 0)$, and an objective function value of $f(\underset{\sim}{x}) = 20.25$ (38:320).

The question that must be addressed is why choose z_2 as zero. Since z_2, and hence x_2, is free to vary it is sensible to select the value to get the most gain in the objective function. The x_1 variable is raised to the fourth power, while x_2 is only linear. The most gain per unit increase

will come from x_1 as compared to x_2. It is therefore advantageous to allow x_1 as large a value as possible. Due to the constraints, this requires x_2 to go as low as possible, or zero. Thus, the insight gained from the generalized inverse solution is augmented by knowledge of the problem and combined with an understanding of the fundamental subspaces of a matrix to determine the optimal solution to the problem.

Example 6 The previous example only had a single constraint, and may therefore have seemed trivial. However, it served to demonstrate the technique. This next example provides a more detailed insight. Consider the following problem (38:333-335):

$$\begin{aligned} \min \quad & -x_1 - x_2 \qquad (4.42) \\ \text{s.t.} \quad & 2x_1 - x_2^2 \geq 1 \\ & -.8x_1^2 - 2x_2 \geq -9 \\ & x_1 , x_2 \geq 0 \end{aligned}$$

In the first stage, each constraint is individually treated as an equality constraint. The second stage will consider both constraints simultaneously as equality constraints. Had the problem been large (i.e. more constraints), the combinatorial considerations of the subsets of equality constraints in each stage would have been significantly *more* involved.

The first step, of the first stage, is to consider the constraint $2x_1 - x_2^2 = 1$, ignoring the second constraint for the moment. Placing the constraint into the ST format and

reducing the augmented form to the final ST canonical form produces the following:

$$\left[\begin{array}{cc|c} 2 & -x_2 & 1 \\ \hline 1 & 0 & \\ 0 & 1 & \end{array}\right] \rightarrow \left[\begin{array}{c|c|c} 1 & 0 & 1/2 \\ \hline 1 & x_2/2 & \\ 0 & 1 & \end{array}\right] \quad (4.43)$$

Using the $A_{1,2}^T = S^T T = (1,0)\ (1/2) = (1/2\ ,\ 0)$ and the same formula for the general solution as was used in example 5 yields the following expressions:

$$\underset{\sim}{x} = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}\begin{bmatrix} 1 \end{bmatrix} + \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}\begin{bmatrix} 2 & -x_2 \end{bmatrix} \right)\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 + x_2 z_2/2 \\ z_2 \end{bmatrix} = \begin{bmatrix} (z^2+1)/2 \\ z_2 \end{bmatrix} \quad (4.44)$$

When this expression is evaluated in the objective function, the resulting function, $f(\underset{\sim}{x}) = (z_2+1)^2/2$, is minimized when the z_2 variable takes on a value of -1. Selecting $z_2 = -1$ produces the particular solution of $\underset{\sim}{x} = (0,-1)$ and $f(\underset{\sim}{x}) = 1$. However, this solution violates the non-negativity constraint for the problem and is thus an infeasible solution.

In a similar manner, the second constraint is now considered as an equality constraint, $(-.8x_1^2 - 2x_2 = 9)$, ignoring for the moment the first constraint. This particular constraint produces an $A_{1,2}^T = (-1.25x_1, 0)$ which is used in the formula for the general solution to produce the following expression:

$$\underset{\sim}{x} = \begin{bmatrix} (90 - 20z_2)/(8x_1) \\ z_2 \end{bmatrix} \qquad (4.45)$$

This form, when used in the objective function, produces a potential solution of $\underset{\sim}{x} = (0, 4.5)$ yielding an objective function value of $f(\underset{\sim}{x}) = -4.5$. However, this solution is not feasible for the first constraint.

This completes consideration of single constraints. In the second stage, constraints are considered as equality constraints in pairs of equations. Since this results in considering all the constraints, this is the last step for this particular problem. The solution obtained will be either feasible and optimal or it will be infeasible. If infeasible, the problem is inconsistent and the solution obtained will be the best in a least-squares sense.

Considering both constraints, yields the following transformation to the final ST canonical form:

(4.46)

$$\left[\begin{array}{cc|cc} 2 & -x_2 & 1 & 0 \\ -.8x_1 & -2 & 0 & 1 \\ \hline 1 & 0 & & \\ 0 & 1 & & \end{array}\right] \rightarrow \left[\begin{array}{cc|cc} 1 & 0 & 1/2 & 0 \\ 0 & 1 & \frac{-x_1}{5+x_1x_2} & \frac{-5}{2(5+x_1x_2)} \\ \hline 1 & x_2/2 & & \\ 0 & 1 & & \end{array}\right]$$

Since the matrix involving both constraints has full rank, the product of the S and T submatrices not only produces the $A_{1,2}$ inverse but the A^+ and A^{-1} as well, since all are the same matrix. This means the general solution given by $\underset{\sim}{x} = A^+\underset{\sim}{b} + (I - A^+A)\underset{\sim}{z}$ reduces to simply $\underset{\sim}{x} = A^+\underset{\sim}{b}$ and

is given by:

$$\underset{\sim}{x} = \begin{bmatrix} \frac{5(2 + 9x_2)}{4(5 + x_1x_2)} \\ \frac{45 - 2x_1}{2(5 + x_1x_2)} \end{bmatrix} \qquad (4.47)$$

This is the expression for all solutions to the pair of equality constraints. Using these expressions in the objective function yields the functional form of:

$$f(x_1,x_2) = \frac{45x_2 + 4x_1 - 80}{4(5 + x_1x_2)} \qquad (4.48)$$

The minimum points of (4.48) can be found by computing the two partial derivatives, $\partial f/\partial x_1$ and $\partial f/\partial x_2$, setting the equations equal to zero, and solving the resulting set of equations. The partial derivatives are:

$$\begin{cases} \frac{\partial f}{\partial x_1} = \frac{-20(9x_2 - 2)(x_2 - 2)}{(20 + 4x_1x_2)^2} \\ \frac{\partial f}{\partial x_2} = \frac{-4(2x_1 + 5)(2x_1 - 45)}{(20 + 4x_1x_2)^2} \end{cases} \qquad (4.49)$$

giving possible, feasible, solutions of (22.5,6.6), (.525,.222), and (2.5,2) of which f(2.5,2) = -4.5 is the best solution and ultimately the optimal solution of the problem (38:335).

Control theory

Example 7. In a 1988 article, Šebek (39) considered a robust control theory problem of the form:

$$A\underset{\sim}{x} + B\underset{\sim}{y} = I \qquad (4.50)$$

where A and B are matrices defined over polynomials of n

unknowns and I is a vector of constants. The example used is the following:

$$(1 - vw)\,x + v^2 y = 1 \qquad (4.51)$$

and solutions to x and y are sought. The $A_{1,2}$ generalized inverse can be applied to this problem, after first expressing (4.51) in the $A\underset{\sim}{x} = \underset{\sim}{b}$ format of:

$$\begin{bmatrix} 1 - vw & v^2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \qquad (4.52)$$

and using the A matrix in the ST canonical form to obtain the following transformation:

$$\left[\begin{array}{cc|c} 1-vw & v^2 & 1 \\ \hline 1 & 0 & \\ 0 & 1 & \end{array}\right] \rightarrow \left[\begin{array}{c|c|c} 1 & 0 & 1 \\ \hline 1 & -v^2 & \\ w/v & 1-vw & \end{array}\right] \qquad (4.53)$$

The $A_{1,2}$ generalized inverse, found by computing the product of S and T, is $(\,1\,,\ w/v\,)^T$. Since the consistency condition, $AA_{1,2}\underset{\sim}{b} = \underset{\sim}{b}$ holds, the general solution is given by $\underset{\sim}{x} = A_{1,2}\underset{\sim}{b} + (I - A_{1,2}A)\underset{\sim}{z}$, for arbitrary $\underset{\sim}{z}$. This general form works out to the following:

$$\underset{\sim}{x} = \begin{bmatrix} 1 \\ w/v \end{bmatrix} + \begin{bmatrix} vw & -v^2 \\ w^2 - w/v & 1-vw \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \qquad (4.54)$$

$$= \begin{bmatrix} 1 + z_1(vw) - z_2(v^2) \\ w/v + z\,(w^2 - w/v) + z\,(1-wv) \end{bmatrix}$$

Now the expression in (4.54) is an expression for the general solution, meaning that all particular solutions to the problem may be generated by choice of $\underset{\sim}{z}$. These

particular solutions differ only in the homogeneous portion of the solution. For instance letting $\underset{\sim}{z} = (1,0)$ gives rise to the particular solution of $\underset{\sim}{x}^T = (1+vw\ ,\ w^2)$, which is the only solution obtained by Šebek. The advantage gained by the generalized inverse technique is the more powerful general form of the solution obtained.

Counter examples

The examples presented thus far have shown how the generalized inverse technique can lead to expressions from which to determine optimal solutions. However, this is not always the case. In the next two examples an explicit solution to the problem is not found, but at the same time these two examples show that the general form of the solution can still provide useful information.

Example 8. Another function specifically designed to test the gradient based optimization techniques is a test function due to Kantorovich (44:42-43). This function:

$$f(x,y) = (3x^2y + y^2 - 1)^2 + (x^4 + xy^3 - 1)^2 \qquad (4.55)$$

can be minimized by solving the system of equations:

$$\begin{cases} 3x^2 + y^2 = 1 \\ x^4 + xy^3 = 1 \end{cases} \qquad (4.56)$$

Setting the system of equations in (4.56) into the $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$ format gives:

$$\begin{bmatrix} 3xy & y \\ x^3 & xy^2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad (4.57)$$

The A matrix is placed into the ST canonical form and

reduced using a series of elementary transformations to the final ST canonical form. This produces the following transition, from initial ST canonical form:

$$\left[\begin{array}{cc||cc} 3xy & y & 1 & 0 \\ x^3 & xy^2 & 0 & 1 \\ \hline 1 & 0 & & \\ 0 & 1 & & \end{array}\right] \tag{4.58}$$

to the final ST canonical form:

$$\left[\begin{array}{cc||cc} 1 & 0 & \frac{1}{y} & 0 \\ 0 & 1 & \frac{-xy}{x^3-3x^2y^2} & \frac{1}{x^3-3x^2y^2} \\ \hline 0 & 1 & & \\ 1 & -3x & & \end{array}\right] \tag{4.59}$$

Just as in example 2, Rosenbrock's function, A is a full rank matrix meaning $A^+ = A^{-1}$ and the solution to the problem is given by $\underset{\sim}{x} = A^+\underset{\sim}{b}$, which produces the expression:

$$\underset{\sim}{x} = (ST)\underset{\sim}{b} = \begin{bmatrix} \frac{1 - xy}{x^3-3x^2y^2} \\ \frac{x^3 - 3xy}{y(x^3-3x^2y^2)} \end{bmatrix} \tag{4.60}$$

The expression for $\underset{\sim}{x}$ is quite complicated. It shows the dependent relationship between x and y and just how delicate the process of finding the optimal solution can be. Although the expression for the general solution obtained using the generalized inverse proves that a unique solution to the problem does in fact exist, the problem is that the above

expression for $\underset{\sim}{x}$ does not help find that optimal solution. In his work on gradient based optimization techniques, Stein found the best solution to date for the function (4.55) as $\underset{\sim}{x} = (.99278 , .30644)$ and $f(\underset{\sim}{x}) = .28173 \times 10^{-11}$ (44:43).

Example 9. This example is a particular nonlinear, constrained optimization problem for which the method of Lagrange multipliers fails (48:85):

$$\begin{aligned} \min \quad & (x^2 + y^2)^{1/2} \\ \text{s.t.} \quad & y^2 - (x - 1)^3 = 0 \end{aligned} \tag{4.61}$$

The generalized inverse technique is generally applicable to Lagrange multiplier optimization (see Appendix A for definition). The Lagrangian function, when differentiated, produces a system of partial derivatives, which when set equal to zero, provide the optimal solutions to the problem.

If there are say m variables, or unknowns, in the problem, and n constraints, the system of partial derivatives produces m+n equations and m+n unknowns. Generally such systems produce a unique solution. The generalized inverse of the (m+n)x(m+n) matrix is used to determine the optimal values of the undetermined multipliers and the variables of the problem.

For this problem, the Lagrangian function is:

$$L(x,y,\lambda) = (x^2 + y^2)^{1/2} + \lambda(y^2 - x^3 + 3x^2 - 3x + 1) \tag{4.62}$$

Taking partial derivatives, $\partial L/\partial x$, $\partial L/\partial y$, and $\partial L/\partial \lambda$, and rewriting the resulting system of equations into the

$\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$ form produces:

$$\begin{bmatrix} 1+6y^2\lambda-3x\lambda & -3x^2y\lambda-3y\lambda & 6x^3-3x^4 \\ 2xy\lambda & 1 & 2y^3 \\ -x^2+3x-3 & y & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \qquad (4.63)$$

The A matrix is full rank so again $A^+ = A^{-1}$. The MACSYMA expert system contains an inverse function, called INVERT(matrix), so MACSYMA was used to obtain the A^{-1} matrix for (4.63) (50:6-55). Since MACSYMA does not compute any generalized inverses, had A been singular, the ST method would have been used to compute the A^+. The A^{-1} matrix provides the solution to the problem by the formula $\underset{\sim}{x} = A^{-1}\underset{\sim}{b}$. The final expression is then:

$$\underset{\sim}{x} = \begin{bmatrix} 3((2x^2+2)y^4\lambda - x^4 + 2x^3) \\ 2y((6y^4 - 3xy^2 + 3x^5 - 6x^4)\lambda + y^2) \\ -((6x^3 + 6x)y^2\lambda^2 + (6y^2 - 3x)\lambda + 1) \end{bmatrix} W \qquad (4.64)$$

and

$$W = \frac{1}{((12y^6+(-6x^4+18x^3-24x^2+12x-8)y^4+(6x^5-12x^4)y^2)\lambda+ 12y^4+3x^6-15x^5+27x^4-18x^3)} \qquad (4.65)$$

Just as in example 8, this final expression does not yield an optimal solution. However, using the generalized inverse to obtain the expression for the general solution demonstrates the non-existence of a finite λ, Lagrange multiplier, for the problem.

Common Solutions

The last theorems in Chapter III were presented to demonstrate the trend towards multiparameter matrices.

Another purpose was to lay the theoretical groundwork for future applications of parallel processing capabilities. The next, and final, example demonstrates the applicability of this theory.

Example 10. Consider the following system of the form of $A\underset{\sim}{x} = \underset{\sim}{b}$ (21):

$$\begin{bmatrix} x & x^2+1 \\ x^2y & x^3y+xy \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x+x^3 \\ 2x^2y+x^4y \end{bmatrix} \quad (4.66)$$

In this particular example, first the solution to the equation is found and then the equation is decomposed into a system of n=2 matrix equations. This system of matrix equations is then solved individually and also using the results of Theorem 3.13.

First the $A_{1,2}$ generalized inverse of A is computed. The transformation from the initial ST canonical form to the final ST canonical form is given by:

$$\left[\begin{array}{cc|cc} x & x^2+1 & 1 & 0 \\ x^2y & x^3y+xy & 0 & 1 \\ \hline 1 & 0 & & \\ 0 & 1 & & \end{array}\right] \rightarrow \left[\begin{array}{c|c|cc} 1 & 0 & 1 & 0 \\ \hline 0 & 0 & -xy & 1 \\ \hline -x & 1+x^2 & & \\ 1 & -x & & \end{array}\right] \quad (4.67)$$

From this the $A_{1,2}$ generalized inverse is found to be:

$$A_{1,2} = ST = \begin{bmatrix} -x \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} -x & 0 \\ 1 & 0 \end{bmatrix} \quad (4.68)$$

which when used in the general solution formula used in other examples, $\underset{\sim}{x} = A_{1,2}B + (I - A_{1,2}A)\underset{\sim}{z}$, yields:

$$\underset{\sim}{x} = \begin{bmatrix} -x & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2x+x^3 \\ 2x^2y+x^4y \end{bmatrix} + \left[I - \begin{bmatrix} -x & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x & x^2+1 \\ x^2y & x^3y+xy \end{bmatrix}\right]\underset{\sim}{z}$$

$$= \begin{bmatrix} -2x^2 - x^4 \\ 2x + x^3 \end{bmatrix} + \begin{bmatrix} 1+x^2 & x^3+x \\ -x & -x^2 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \qquad (4.69)$$

Since $\underset{\sim}{z}^T = (z_1, z_2)$ is an arbitrary vector, choosing a value of $\underset{\sim}{z}^T = (1\ ,\ x)$ will give a particular solution to the problem. When used in (4.69), this solution is $\underset{\sim}{x}^T = (1, x)$.

Next, reconsider (4.66) as a system of matrix equations of the form:

$$\begin{cases} A_1\underset{\sim}{x} = B_1 \\ A_2 x = B_2 \end{cases} \qquad (4.70)$$

where

$$\begin{cases} A_1 = (\ x \quad x^2+1) & B_1 = (2x + x^3) \\ A_2 = (x^2y \quad x^3y+xy) & B_2 = (2x^2 + x^4y) \end{cases} \qquad (4.71)$$

The intent is to find a common solution to the set of matrix equations. In a manner similar to that of (4.67), each of A_1 and A_2 can be reduced, within the ST canonical form, to find an $A_{1,2}$ generalized inverse of the matrix. To avoid confusion of notation with the subscripts used in (4.70) and (4.71), these $A_{1,2}$ generalized inverses are denoted as A_1^- and A_2^- respectively through the remainder of this example.

From the ST computations, representatives of the A_1^- and A_2^- inverses are found. These inverses are:

$$A_1^- = \begin{bmatrix} -x \\ 1 \end{bmatrix} \qquad A_2^- = \begin{bmatrix} 1/x^2y \\ 0 \end{bmatrix} \qquad (4.72)$$

A check of Penrose conditions (1) and (2) will verify that these are in fact $A_{1,2}$ inverses of A_1 and A_2 respectively.

Theorem 3.13 requires that each of the conditions $A_1A_1^-B_1=B_1$ and $A_2A_2^-B_2=B_2$ hold for each of the equations of (4.70) to have a solution. To demonstrate $A_1A_1^-B_1=B_1$:

$$A_1A_1^-B_1=B_1=\begin{bmatrix}x & x^2+1\end{bmatrix}\begin{bmatrix}-x\\ 1\end{bmatrix}\begin{bmatrix}2x+x^3\end{bmatrix}=\begin{bmatrix}1\end{bmatrix}\begin{bmatrix}2x+x^3\end{bmatrix}=B_1 \qquad (4.73)$$

and to demonstrate $A_2A_2^-B_2=B_2$:

$$=\begin{bmatrix}x^2y & x^3y+xy\end{bmatrix}\begin{bmatrix}1/x^2y\\ 0\end{bmatrix}\begin{bmatrix}2x^2+x^4y\end{bmatrix}=\begin{bmatrix}1\end{bmatrix}\begin{bmatrix}2x^2+x^4y\end{bmatrix}=B_2 \qquad (4.74)$$

Since (4.73) is true, the general solution to $A_1\underset{\sim}{x}=B_1$ is given by $\underset{\sim}{x}=A_1^-B_1+(I-A_1A)\underset{\sim}{z}$ for arbitrary $\underset{\sim}{z}$. This is expressed as:

$$\underset{\sim}{x}=\begin{bmatrix}-x\\ 1\end{bmatrix}\begin{bmatrix}2x+x^3\end{bmatrix}+\left[\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}-\begin{bmatrix}-x\\ 1\end{bmatrix}\begin{bmatrix}x & x^2+1\end{bmatrix}\right]\begin{bmatrix}z_1\\ z_2\end{bmatrix} \qquad (4.75)$$

$$=\begin{bmatrix}-2x^2-x^4\\ 2x+x^3\end{bmatrix}+\begin{bmatrix}1+x^2 & x^3+x\\ -x & -x^2\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix}$$

Letting $\underset{\sim}{z}^T=(1\ ,\ x)$ in the above gives the particular solution to the problem, $\underset{\sim}{x}^T=(1\ ,\ x)$. In a similar manner, the condition verified in equation (4.74) means the general solution to $A_2\underset{\sim}{x}=B_2$ is given by $\underset{\sim}{x}=A_2^-B_2+(I-A_2A)\underset{\sim}{z}$ for $\underset{\sim}{z}$ an arbitrary vector. This expression turns out, after simplification, to be:

$$\underset{\sim}{x}=\begin{bmatrix}(2+x^2y)/y\\ 0\end{bmatrix}+\begin{bmatrix}0 & (-1-x^2)/x\\ 0 & 1\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix} \qquad (4.76)$$

Letting $\underset{\sim}{z}^T = (1 \ , \ x)$ as before yields a particular solution $\underset{\sim}{x}^T = \left[(2-y)/y \ , \ x\right]$. Though this appears somewhat different from the previous result in equation (4.75), it really is not. If $\underset{\sim}{x} = (1 \ , x)$, then y=1. If y=1 for this current particular solution, then $\underset{\sim}{x} = (1 \ , \ x)$. So actually this particular solution is equivalent to the previous particular solution.

Thus far this example has found a solution to (4.66) and (4.70). This demonstrates that a common solution to an equation must satisfy the individual elements of that equation.

The next step is to actually demonstrate the validity of the recursive formulas given in Theorem 3.13. Using (4.70), Theorem 3.13 gives the following expression for the common solution to the set of matrix equations:

$$\underset{\sim}{x} = E_2 + F_2\underset{\sim}{z} \qquad (4.77)$$

where $\underset{\sim}{z}$ is an arbitrary vector. Using the recursive relationships of Theorem 3.13, namely:

$$E_2 = E_1 + F_1C_2^-D_2 \qquad F_2 = F_1(I - C_2^-C_2) \qquad C_2 = A_2F_1 \qquad (4.78)$$
$$D_2 = B_2 - A_2E_1 \qquad E_1 = A_1^-B_1 \qquad F_1 = I - A_1^-A_1$$

results in the following expanded form of (4.77). This form is the formula that is ultimately evaluated to provide the common solution to the set of equations:

$$\underset{\sim}{x} = \left[A_1^- B_1 + \left[I - A_1^- A_1\right]\left[A_2 - A_2 A_1^- A_1\right]^- \left[B_2 - A_2 A_1^- B_1\right]\right] + \left[\left[I - A_1^- A_1\right] - \left[I - A_1^- A_1\right]\left[A_2 - A_2 A_1^- A_1\right]^- \left[A_2 - A_2 A_1^- A_1\right]\right]\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{4.79}$$

These computations finally produce the following expression for the general solution to the system of equations in (4.70):

$$\underset{\sim}{x} = \begin{bmatrix} -2x^2 - x^4 \\ 2x + x^3 \end{bmatrix} + \begin{bmatrix} 1+x^2 & x^3+x \\ -x & -x^2 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \tag{4.80}$$

and letting $\underset{\sim}{z}^T = (1 \ , \ x)$ yields $\underset{\sim}{x} = (1 \ , \ x)$, the common solution to the system of equations (4.70) and the solution to the problem as originally stated (4.66).

Conclusion

This chapter focused on using the generalized inverse of a multiparameter matrix as a technique to help solve various optimization problems. The technique is useful in solving systems of equations and sets of equations. For quadratic programming types of problems, the generalized inverse provides for an easy to understand methodology to solve the problem. Finally, current theory regarding simultaneous solutions of sets of matrix equations was demonstrated for the case of n=2 matrix equations.

V. Conclusions and Recommendations

This thesis has examined the application of generalized inverses of multiparameter matrices. Such matrices arise in many modern applications, such as multi-input/multi-output systems, highly nonlinear functions in optimization, and in modern control theory problems. The use of these generalized inverses provides, quite often, a powerful analysis tool.

A large part of this thesis dealt with consolidating the vast amount of theory regarding generalized inverses. The theorems presented in Chapter III, though not proved in this work (except for Theorem 3.13), form the basis for the current research into multiparameter matrix theory. The format of the chapter was intended to provide an understanding of the theory as well as an appreciation of how the research has evolved into the multiparameter arena.

For the first time, the ST technique was tied to the Fundamental Theorem of Linear Algebra. This key concept of linear algebra was discussed at some length in Chapter III because there is a crucial link between the concept and the idea of a generalized inverse. The ST technique bridges whatever gap may have existed and provides a representative basis from each of the four fundamental subspaces associated with any given matrix. In addition, the ST technique can produce representatives of various generalized inverses rather than being limited to just computing the unique Moore-Penrose generalized inverse.

Also for the first time, an explicit proof was provided for Theorem 3.13 (Appendix B) which deals with the common solutions of a system of matrix equations. The recursive formulas as well as the identities used within the formulas were proved. This theorem, an extension of work done by previous researchers (26 ; 17), sets the stage for possible future applications in parallel processing of systems of matrix equations, with either constant coefficients or polynomial coefficients in the matrices.

The main thrust of this research effort was to explore the applications of the generalized inverse in non-linear optimization involving multiparameter matrices. This was the focus of Chapter IV in which various types of nonlinear optimization problems were solved using the generalized inverse technique as a basis.

There is a very strong interface between the generalized inverse of multiparameter matrices, and the solution to a system of equations. This is easily extended to the solution of Lagrangian optimization problems since the partial derivatives of the Lagrangian function yield a homogeneous system of equations.

The generalized inverse of multiparameter matrices was shown as providing the capability of extending Nelson's optimization algorithm to problems of higher dimension than quadratic. Though the generalized algorithm introduced can be combinatorially inefficient for large problems, the algorithm is easy to implement and is sufficient for many

common optimization problems.

Finally an explicit example of solving sets of matrix equations for a common solution was shown. This example demonstrated the applicability of Theorem 3.13 and highlights a potential area of future research, parallel processing of matrix equations.

There are some areas of future research that can be undertaken. First is in the area of the ST computational algorithm. This technique was first programmed, in FORTRAN, for purely numerical matrices in 1985 by Murray (27). His program can be improved upon in the sense of computer storage requirements and in numerical accuracy.

The ST technique is applicable to finding the generalized inverses of multiparameter matrices. The ST technique uses elementary transformations to compute the inverses. These transformations are easy to implement in a computer algorithm. Expert systems such as MACSYMA enable users to work with variable element matrices in symbolic form. Such systems also provide some kind of capability for developing macros (sequences of system commands or procedures) to perform certain functions thereby increasing the power of the system. There is already at least one macro in MACSYMA for the generalized inverse (11), but it requires the use of limits and for large matrices is not as efficient as the ST technique. One very important area of research would be to develop the ST algorithm for the MACSYMA environment. Ideally, this program would be written in LISP

since MACSYMA is also written in the LISP language.

Another potential area of research is in the common solutions to sets of matrix equations. In particular, the application of this theory to a parallel processing environment. Initially such work would be limited to the application of constant coefficient matrices. However, the algorithms developed for constant matrices should easily extend to multiparameter matrices. Of course computer capabilities must provide a parallel processing environment for systems such as MACSYMA in order to extend any parallel processing algorithms to multiparameter matrices.

Control theory was just briefly touched upon in this thesis, but a new area of research is in parameterized families of systems (i.e. multiparameter matrices) involving the design of parameterized controllers. Sontag (43) discusses such a problem in his 1985 tutorial article. In particular would be to design these controllers, "in the form of a parameterized controller which regulates once its parameters are properly tuned" (43:370). It appears this may be a fertile area of future research.

Appendix A: Glossary of Terms

Determinant. Defined as the sum of all signed elementary products from a square, nxn, matrix A. An elementary product is any product of n entries of matrix A, no two of which shall come from the same row or column. The elementary product is termed a "signed elementary product" when multiplied by ±1, dependent upon whether the product is an even or an odd permutation (1:59-70)

Diagonal Matrix. Matrix with all zero entries except along the diagonal, which contains any non-zero entries in the matrix. In more formal terms, the elements $a_{ij} = 0$ if $i \neq j$. The diagonal elements are the non-zero elements a_{ij} where $i = j$. Also defined as an upper and lower triangular matrix.

Diagonable. If a matrix A is similar to a diagonal matrix, then A is a diagonable matrix. See definition of similar matrices (2:157).

Eigenvalues. The eigenvalues of a matrix A are the scalar values, λ, for which $A\underset{\sim}{x} = \lambda\underset{\sim}{x}$ has non-zero solutions (31:264).

Eigenvectors. The non-zero solutions of $A\underset{\sim}{x} = \lambda\underset{\sim}{x}$, where the λ scalars are the eigenvalues of the matrix A (31:264).

Elementary transformations. Very commonly referred to as elementary row and column operations. These are operations performed on a matrix that preserve the matrices order and rank. These transformations are:

(1) interchange any rows (columns)

(2) multiplication of each element in a row (column) by a non-zero scalar

(3) the addition of a scalar multiple of one row (column) to another row (column)

Changing the term scalar in (2) and (3) to polynomial defines the elementary transformations on polynomial (multiparameter) matrices (2:39,188).

Expert System. Computer system (primarily software) that is programmed to exhibit capabilities normally attributed to human experts in the particular specialty area. In the current context the expert systems exhibit symbolic, algebraic reasoning normally associated only with human mathematical reasoning.

Field. A field is a communitive division ring. See definition of ring (10:198).

Gradient. If the function F(x) is differentiable at a point x_o, then the associated vector of partial derivatives evaluated at x_o is called the gradient vector. The gradient vector provides information regarding the direction of steepest ascent (descent) along a function from a particular point. Thus, it is the key aspect of iterative optimization techniques.

Hessian. Matrix of second partial derivatives. Valuable in numerical analysis and optimization theory as the Hessian of a function provides information about the stationary points of a function.

Homogeneous set of equations. A homogeneous set of equations

is of the form:

$$A\underset{\sim}{x} = \underset{\sim}{0}$$

Thus every solution to the system is a member of the nullspace of the system. If there are more unknowns in the system than there are equations, then the system has non-trivial solutions. This means there is a solution to the system other than simply $\underset{\sim}{x} = \underset{\sim}{0}$ (46:58)

Identity matrix. Square, diagonal matrix with all diagonal elements equal to 1. Also the multiplicative identity of the algebraic field of matrices.

Inner product. (dot product) Sum of the element-wise product of two vectors. For example,

$$X \circ Y = \sum_{i=1}^{n} (x_i y_i)$$

Lagrange multipliers. Optimization method involving the derivatives of the objective function and the equality constraints of the problem. A maximum or minimum point occurs when the derivative of the objective function equals zero. The constraint derivative already equals zero (since the derivative of a constant right hand side is zero). Rather than solve for each variable in the constraint and back solve the system of equations, the constraint is multiplied by an undetermined value, λ_i, and added to the objective function. The partial derivatives of the resulting equation produces n equations in n unknowns which can be solved uniquely. Usually, the system is solved for the λ_i values and the optimal x_j values can then be determined (3:174).

Least-squares. Method of estimating parameters of an equation where the function minimized is the squared difference of the actual data value and the value predicted by the current set of estimated parameters.

Markov chains. The most common model for which the input and output relationship is random is the Markov process. When this process is discussed with respect to discrete time, discrete range, the process is termed a Markov chain. The Markov chain, with n states in the state space, is typically characterized by an nxn matrix of probabilities of transition from state to state (35:992).

Minimum-norm. The minimal value of the function $\|A\underset{\sim}{x}-\underset{\sim}{b}\|$, where $\underset{\sim}{x}$ is a vector of estimated parameters.

Minimum-variance. A minimum-variance estimator of a parameter is a random variable with the property of having the smallest variance among any other estimators of that parameter (25:15).

Maximum Likelihood Estimation (MLE). A parameter estimation technique that maximizes the probability (or likelihood) of the observed sample (23:372).

Norm of a vector. Also referred to as the length of a vector. Denoted $\| x \|$, the norm is defined as the square root of the inner product of a vector with itself. For example: $\| x \| = \sqrt{x \cdot x}$ (1:95).

Nullspace. The set of solutions to $A\underset{\sim}{x} = \underset{\sim}{0}$, forms a vector space called the null space. In more abstract terms, the nullspace is the set of points that the transformation

matrix, A, maps into the point zero. The null space is the kernel of the transformation given by the matrix A. The dimension of the nullspace is defined as the nullity.

Orthogonal. Two vectors are said to be orthogonal if their inner product (dot product) is equal to zero. Orthogonal vectors intersect at 90° angles (i.e. they are perpendicular). A square matrix is orthogonal if $A^{-1}=A^{T}$, and the equality $A^{T}A = AA^{T} = I$ holds (2:103).

Orthonormal. In general terms, if a set of vectors are mutually orthogonal, and each has a norm of one, the set is an orthonormal set of vectors (2:105)

Parallel processing. Computer processing of more than one task on the same computer, simultaneously.

Positive definite matrix. A matrix, A, is positive definite if it's quadratic form, $x^{T}Ax > 0$ for all $x \neq 0$ (38:768).

Rank of a matrix. The rank of a matrix A is the dimension of the row and column space of A. The dimension of the row and column space is defined as the number of vectors required to span the space (1:157).

Range. The range of a matrix A, or the transformation induced by A, is defined as all possible values of $A\underset{\sim}{x}$.

Ring. A ring is a set with the binary operations of addition and multiplication. The addition operation is communitive with additive identity 0. If the multiplication operation is communitive, the ring is termed a communitive ring. If the ring contains a multiplicative identity, such as 1, and each element has a unique multiplicative inverse

in the ring, the ring is called a division ring (10:195).

Set, Subset. A set is a well defined collection of objects. A subset is itself a set, but also entirely contained within some other set of equal or larger size (10:2).

Singular matrices. A square matrix, A, that does not possess an inverse matrix, A^{-1}. In these cases, the determinant of A, det(A), is zero. All non-square matrices are singular matrices, with an undefined determinant.

Slack variable. A slack, or surplus, variable represents the positive difference between the left and right hand side of an inequality equation. These variables are used in linear programming (LP) to transform inequality constraints into equality constraints. For example, adding the slack variable S_1 allows:

$$2x_1 + x_2 - 3x_3 \leq 25 \longrightarrow 2x_1 + x_2 - 3x_3 + S_1 = 25$$

Smith form. Diagonal form of a multiparameter matrix, $A(\lambda)$, where the diagonal polynomial elements, $f_i(\lambda)$, are monic and $f_i(\lambda)$ divides $f_{i+1}(\lambda)$, for every i. These polynomial elements are the invariant factors of $A(\lambda)$. If each $f_i(\lambda)=1$, the identity matrix, each $f_i(\lambda)$ is called a trivial invariant factor (2:188).

Stationary point. Point of the surface of the function where the function is neither increasing nor decreasing, within a sufficiently small region about the point.

Similar matrices. Two nxn, square, matrices, A and B, are similar if there exists some non-singular matrix, P, such that $B = P^{-1}A\,P$. The matrices A and B are said to be

equivalent matrices. This definition also holds if the matrices A and B are defined over the ring of polynomials (2:156).

Triangular matrix. The two triangular matrices are upper and lower triangular matrices. In an upper triangular matrix, each $a_{ij} = 0$ if $i > j$. Conversely, a lower triangular matrix requires each $a_{ij} = 0$ if $i < j$. If a matrix is both upper and lower triangular, the matrix is a diagonal matrix.

Unbiased estimator. For an estimator of a parameter to be unbiased, the long-run average, or expected value, of the estimator should be the parameter being estimated (25:14).

Appendix B: Proof of Theorem 3.13

Introduction

This theorem has a couple of important uses. First, it provides necessary and sufficient conditions for the existence of a common solution to a system of matrix equations. Secondly, it provides a recursive algorithm for computing the common solution to the system. The work in this appendix verifies, by mathematical induction, the necessary and sufficient conditions and the recursive algorithm supplied by the proof.

All matrices are assumed defined over the complex ring of polynomials, $\mathbb{C}^{mxn}$. Further, to avoid confusion regarding the subscript notation, a generalized inverse is denoted as A^-. Subscripts denote matrix numbering only.

The method of proof is to first verify the necessary and sufficient conditions for the system of equations. The next step is to demonstrate the validity of the general solution expression.

Theorem 3.13

Let $A_i \in \mathbb{C}^{pxr}$ and $B_i \in \mathbb{C}^{pxr}$ for $i = 1,\ldots,m$. Define the following recursive relationships:

$$C_1 = A_1 \qquad D_1 = B_1 \qquad \text{(B.1)}$$

$$E_1 = A_1^- B_1 \qquad F_1 = I - A_1^- A_1$$

and

$$C_k = A_k F_{k-1} \qquad D_k = B_k - A_k E_{k-1}$$

$$E_k = E_{k-1} + F_{k-1} C_k^- D_k \qquad F_k = F_{k-1}(I - C_k^- C_k)$$

Then $A_i \underset{\sim}{x} = B_i$, for $i=1,\ldots,m$, has a common solution if and only if $C_i C_i^- D_i = D_i$ for $i=1,\ldots,m$. In this case the general common solution is given by

$$\underset{\sim}{x} = E_m + F_m \underset{\sim}{z} \qquad (B.2)$$

where $\underset{\sim}{z}$ is arbitrary.

Proof

Case n=1.

Consider the following system of equations:

$$A_1 \underset{\sim}{x} = B_1 \qquad (B.3)$$

From Corollary 3.2.1, the system of equations given by (B.3) has a solution if and only if the following consistency condition is true:

$$A_1 A_1^- B_1 = B_1 \qquad (B.4)$$

The solution then is given by the equation:

$$\underset{\sim}{x} = A_1^- B_1 + (I - A_1^- A_1)\underset{\sim}{z} \quad \forall\ \underset{\sim}{z} \text{ arbitrary} \qquad (B.5)$$

Using the recursive definitions provided in the theorem, equation (B.1), and directly substituting these into equation (B.4), an equivalent form becomes:

$$C_1 C_1^- D_1 = D_1 \qquad (B.6)$$

and the general solution, (B.5) is equivalent to:

$$\underset{\sim}{x} = E_1 + F_1 \underset{\sim}{z} \quad \forall\ \underset{\sim}{z} \text{ arbitrary} \qquad (B.7)$$

Case i=2.

Consider the following system of equations:

$$A_1 \underset{\sim}{x} = B_1 \qquad (B.8)$$
$$A_2 \underset{\sim}{x} = B_2$$

The necessary and sufficient conditions are that $C_2 C_2^- D_2 = D_2$. This may be expanded out using the definitions as in the following:

$$C_2 C_2^- D_2 = D_2 \rightarrow \left[A_2(I - A_1^- A_1)\right]\left[A_2(I - A_1^- A_1)\right]^- D_2 = D_2 \qquad (B.9)$$

Consider the expression $\left[A_2(I - A_1^- A_1)\right]$. Suppose that A_1 and A_2 are matrices defined over $\mathbb{C}^{pxq}$, then $(I - A_1^- A_1)$ is qxq, and the entire expression is pxq. Since the expression, $P = \left[A_2(I - A_1^- A_1)\right]$ is a matrix, and an element of $\mathbb{C}^{pxq}$, there exists a generalized inverse, P^-, such that $PP^-P = P$. Thus equation (B.9) reduces to simply D_2. This verifies the necessary and sufficient conditions.

The general solution is given by:

$$\underset{\sim}{x} = E_2 + F_2 \underset{\sim}{z} \quad \forall \ \underset{\sim}{z} \text{ arbitrary} \qquad (B.10)$$

Again using the recursive definitions, (B.10) can be expanded out to the following form:

$$\underset{\sim}{x} = A_1^- B_1 + (I - A_1^- A_1)(A_2 - A_2 A_1^- A_1)^- (B_2 - A_2 A_1^- B_1) + (I - A_1^- A_1)\left[I - (A_2 - A_2 A_1^- A_1)^- (A_2 - A_2 A_1^- A_1)\right]\underset{\sim}{z} \qquad (B.11)$$

Now that the formula is in terms of the matrices given in (B.8), the solution is easily verified. Premultiply by A_1 and (B.11) becomes:

$$A_1 \underset{\sim}{x} = A\,A_1^{-}B_1 + (A_1 - A_1A_1^{-}A_1)(A_2 - A_2A_1^{-}A_1)^{-}(B_2 - A_2A_1^{-}B_1) + \quad \text{(B.12)}$$

$$(A_1 - A_1A_1^{-}A_1)\left[I - (A_2 - A_2A_1^{-}A_1)^{-}(A_2 - A_2A_1^{-}A_1)\right]\underset{\sim}{z}$$

Since $(A_1 - A_1A_1^{-}A_1) = (A_1 - A_1) = 0$, (B.12) reduces to:

$$A_1 \underset{\sim}{x} = A_1A_1^{-}B_1 = B_1 \qquad \text{(B.13)}$$

Premultiply (B.11) by A_2 and the result is:

$$A_2 \underset{\sim}{x} = A_2A_1^{-}B_1 + (A_2 - A_2A_1^{-}A_1)(A_2 - A_2A_1^{-}A_1)^{-}(B_2 - A_2A_1^{-}B_1) + \quad \text{(B.14)}$$

$$(A_2 - A_2A_1^{-}A_1)\left[I - (A_2 - A_2A_1^{-}A_1)^{-}(A_2 - A_2A_1^{-}A_1)\right]\underset{\sim}{z}$$

The homogeneous portion of (B.14) falls out of the equation, as demonstrated:

$$(A_2 - A_2A_1^{-}A_1)\left[I - (A_2 - A_2A_1^{-}A_1)^{-}(A_2 - A_2A_1^{-}A_1)\right] \rightarrow \qquad \text{(B.15)}$$

$$(A_2 - A_2A_1^{-}A_1) - (A_2 - A_2A_1^{-}A_1)(A_2 - A_2A_1^{-}A_1)^{-}(A_2 - A_2A_1^{-}A_1) \rightarrow$$

$$(A_2 - A_2A_1^{-}A_1) - (A_2 - A_2A_1^{-}A_1) = 0$$

Furthermore, $(A_2 - A_2A_1^{-}A_1)(A_2 - A_2A_1^{-}A_1)^{-}(B_2 - A_2A_1^{-}B_1)$ is equal to $(B_2 - A_2A_1^{-}B_1)$. Thus (B.14) becomes:

$$A_2 \underset{\sim}{x} = A_2A_1^{-}B_1 + B_2 - A_2A_1^{-}B_1 = B_2 \qquad \text{(B.16)}$$

Case n=3.

Consider the following system of equations:

$$\begin{aligned} A_1 \underset{\sim}{x} &= B_1 \\ A_2 \underset{\sim}{x} &= B_2 \\ A_3 \underset{\sim}{x} &= B_3 \end{aligned} \qquad \text{(B.17)}$$

The necessary and sufficient conditions are that

$C_3C_3^-D_3=D_3$. This may be expanded out using the definitions as in the following:

$$\left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]\cdot \tag{B.18}$$

$$\left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]^- D_3$$

where the D_3 term has not been expanded. Just as in the n=2 case considered above, each term in parenthesis is considered. It is already established that $P=(A_2-A_2A_1^-A_1)$ is pxq, provided each of the A_i matrices are pxq. In a similar manner, the $R=(A_3-A_3A_1^-A_1)$ expression can be shown to be a pxq matrix. Thus (B.18) may be rewritten as:

$$(R-RPP)(R-RPP)^-D_3 \tag{B.19}$$

The expression (B.19) is equivalent to (B.18), just easier to read. From (B.19), it is easy to see that the matrix defined by (R-RPP) and its generalized inverse $(R-RPP)^-$ cause the following to hold true:

$$C_3C_3^-D_3 = (R-RPP)(R-RPP)^-D_3 = D_3 \tag{B.20}$$

This then verifies the necessary and sufficient conditions for the case i=3. The general solution is then given by the expression, $\underset{\sim}{x} = E_3 + F_3\underset{\sim}{z}$ $\forall$ $\underset{\sim}{z}$ arbitrary, which, when expanded out using the recursive definitions, is equivalent to the following:

$$\underset{\sim}{x} = A_1^-B_1 + (I-A_1^-A_1)(A_2-A_2A_1^-A_1)^-(B_2-A_2A_1^-B_1) + \tag{B.21}$$

$$(I-A_1^-A_1)\left[I-(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]\cdot$$

$$\left[(A_3 - A_3A_1^-A_1) - (A_3 - A_3A_1^-A_1)(A_2 - A_2A_1^-A_1)^-(A_2 - A_2A_1^-A_1)\right]^- \cdot$$

$$\left[(B_3 - A_3A_1^-B_1) - (A_3 - A_3A_1^-A_1)(A_2 - A_2A_1^-A_1)^-(B_2 - A_2A_1^-B_1)\right] +$$

$$(I - A_1^-A_1)\left[I - (A_2 - A_2A_1^-A_1)^-(A_2 - A_2A_1^-A_1)\right] \cdot$$

$$\left[I - \left[(A_3 - A_3A_1^-A_1) - (A_3 - A_3A_1^-A_1)(A_2 - A_3A_1^-A_1)^-(A_2 - A_2A_1^-A_1)\right]^- \cdot\right.$$

$$\left.\left[(A_3 - A_3A_1^-A_1) - (A_3 - A_3A_1^-A_1)(A_2 - A_2A_1^-A_1)^-(A_2 - A_2A_1^-A_1)\right]\right] \underset{\sim}{z}$$

This expression is verified by premultiplying the expression by A_1, A_2, and A_3. First the A_1 solution is verified:

(B.22)

$$A_1\underset{\sim}{x} = A_1A_1^-B_1 + (A_1 - A_1A_1^-A_1)(A_2 - A_2A_1^-A_1)^-(B_2 - A_2A_1^-B_1) +$$

$$(A_1 - A_1A_1^-A_1)\left[I - (A_2 - A_2A_1^-A_1)^-(A_2 - A_2A_1^-A_1)\right] \cdot$$

$$\left[(A_3 - A_3A_1^-A_1) - (A_3 - A_3A_1^-A_1)(A_2 - A_2A_1^-A_1)^-(A_2 - A_2A_1^-A_1)\right]^- \cdot$$

$$\left[(B_3 - A_3A_1^-B_1) - (A_3 - A_3A_1^-A_1)(A_2 - A_2A_1^-A_1)^-(B_2 - A_2A_1^-B_1)\right] +$$

$$(A_1 - A_1A_1^-A_1)\left[I - (A_2 - A_2A_1^-A_1)^-(A_2 - A_2A_1^-A_1)\right] \cdot$$

$$\left[I - \left[(A_3 - A_3A_1^-A_1) - (A_3 - A_3A_1^-A_1)(A_2 - A_3A_1^-A_1)^-(A_2 - A_2A_1^-A_1)\right]^- \cdot\right.$$

$$\left.\left[(A_3 - A_3A_1^-A_1) - (A_3 - A_3A_1^-A_1)(A_2 - A_2A_1^-A_1)^-(A_2 - A_2A_1^-A_1)\right]\right] \underset{\sim}{z}$$

Since the expression $(A_1 - A_1A_1^-A_1) = A_1 - A_1 = 0$, the entire expression reduces to:

$$A_1\underset{\sim}{x} = A_1A_1^-B_1 = B_1 \qquad (B.23)$$

Premultiply (B.21) by A_2 to obtain:

(B.24)

$$A_2\underset{\sim}{x} = A_2A_1^-B_1 + (A_2-A_2A_1^-A_1)(A_2-A_2A_1^-A_1)^-(B_2-A_2A_1^-B_1) +$$

$$(A_2-A_2A_1^-A_1)\left[I - (A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right] \cdot \quad \longleftarrow$$

$$\left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]^- \cdot$$

$$\left[(B_3-A_3A_1^-B_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(B_2-A_2A_1^-B_1)\right] +$$

$$(A_2-A_2A_1^-A_1)\left[I - (A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right] \cdot \quad \longleftarrow$$

$$\left[I - \left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_3A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]^- \cdot\right.$$

$$\left.\left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]\right] \underset{\sim}{z}$$

Next consider the terms in (B.24) indicated by the arrows. Each of these simplify to:

(B.25)

$$\left[(A_2-A_2A_1^-A_1)-(A_2-A_2A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right] =$$

$$\left[(A_2-A_2A_1^-A_1) - (A_2-A_2A_1^-A_1)\right] = 0$$

This causes the entire (B.24) expression to simplify to just:

$$A_2\underset{\sim}{x} = A_2A_1^-B_1 + (B_2 - A_2A_1^-B_1) = B_2 \qquad (B.26)$$

Premultiply (B.21) by A_3 to obtain:

(B.27)

$$A_3\underset{\sim}{x} = A_3A_1^-B_1 + (A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(B_2-A_2A_1^-B_1) +$$

$$(A_3-A_3A_1^-A_1)\left[I - (A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right] \cdot \quad \longleftarrow$$

$$\left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]^- \cdot$$

$$\left[(B_3-A_3A_1^-B_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(B_2-A_2A_1^-B_1)\right] +$$

$$(A_3-A_3A_1^-A_1)\left[I - (A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right] \cdot \longleftarrow$$

$$\left[I - \left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_3A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]^- \cdot\right.$$

$$\left.\left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]\right] \underset{\sim}{z}$$

Consider the expression pointed to by the first arrow in (B.27). This multiplies out to:

(B.28)

$$\left[(A_3-A_3A_1^-A_1) -(A_3-A_3A_1^-A_1)\,(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right] \cdot$$

$$\left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right]^- \cdot$$

$$\left[(B_3-A_3A_1^-B_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(B_2-A_2A_1^-B_1)\right]$$

$$= (B_3-A_3A_1^-B_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(B_2-A_2A_1^-B_1)$$

The expression pointed to by the second arrow in (B.27) reduces to the following:

(B.29)

$$\left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right] -$$

$$\left[(A_3-A_3A_1^-A_1)-(A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(A_2-A_2A_1^-A_1)\right] = 0$$

Using (B.28) and (B.29), the expression for the general solution, (B.27), reduces to:

(B.30)

$$A_3\underset{\sim}{x} = A_3A_1^-B_1 + (A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(B_2-A_2A_1^-A_1) +$$

$$B_3 - A_3A_1^-B_1 - (A_3-A_3A_1^-A_1)(A_2-A_2A_1^-A_1)^-(B_2-A_2A_1^-A_1) = B_3$$

Suppose, for any k, $A_i\underset{\sim}{x} = B_i$ for i=1,...,k has a common solution if and only if $C_iC_iD_i = D_i$, for i=1,...,k. And suppose this solution is given by:

$$\underset{\sim}{x} = E_k + F_k \underset{\sim}{z} \quad \forall\ \underset{\sim}{z} \text{ arbitrary} \qquad (B.31)$$

Case n=k+1.

Consider the following system of equations:

$$\begin{aligned} A_1 \underset{\sim}{x} &= B_1 \\ &\vdots \\ A_{k+1} \underset{\sim}{x} &= B_{k+1} \end{aligned} \qquad (B.32)$$

If the common solution to $A_{k+1}\underset{\sim}{x}=B_{k+1}$ is the same as the common solution to the set of the first k equations, then:

$$A_{k+1}(E_k + F_k \underset{\sim}{z}) = B_{k+1} \qquad (B.33)$$

$$\Rightarrow\ A_{k+1}F_k \underset{\sim}{z} = B_{k+1} - A_{k+1}E_k$$

which, from the definitions, is equivalent to

$$C_{k+1} \underset{\sim}{z} = D_{k+1} \qquad (B.34)$$

For (B.34) to have a solution, Corollary 3.2.1 requires that $C_{k+1}C_{k+1}^{-}D_{k+1}=D_{k+1}$. In this case, the general solution is:

$$\underset{\sim}{z} = C_{k+1}^{-}D_{k+1} + (I - C_{k+1}^{-}C_{k+1})\,\underset{\sim}{z}' \qquad (B.35)$$

The expression in (B.35) can be used along with (B.31) to conclude that for (B.32) to have a common solution,

$$C_{k+1}C_{k+1}^{-}D_{k+1} = D_{k+1} \qquad (B.36)$$

and the general solution is:

$$\begin{aligned} \underset{\sim}{x} &= E_k + F_k \underset{\sim}{z} \qquad (B.37) \\ &= E_k + F_k \left[C_{k+1}^{-}D_{k+1} + (I - C_{k+1}^{-}C_{k+1})\underset{\sim}{z}'\right] \\ &= E_k + F_kC_{k+1}^{-}D_{k+1} + F_k(I - C_{k+1}^{-}C_{k+1})\underset{\sim}{z}' \end{aligned}$$

Using the definitions for the E_k and F_k terms defined in the theorem, and updating for the current situation:

$$E_{k+1} = E_k + F_k C_{k+1}^{-} D_{k+1} \qquad \text{(B.38)}$$

$$F_{k+1} = F_k (I - C_{k+1}^{-} C_{k+1})$$

and replacing in equation (B.37) causes the general solution expression to reduce to:

$$\underset{\sim}{x} = E_k + F_k C_{k+1} D_{k+1} + F_k (I - C_{k+1} C_{k+1}) \underset{\sim}{z}' \qquad \text{(B.39)}$$

$$= E_{k+1} + F_{k+1} \underset{\sim}{z}'$$

Since k is an arbitrary number of equations, this holds true for all k. This completes the proof.

Bibliography

1. Anton, Howard. Elementary Linear Algebra. New York: John Wiley and Sons, 1977.

2. Ayres, Frank. Schaum's Outline of Theory and Problems of Matrices. New York: McGraw-Hill Inc, 1962.

3. Boas, Mary L., Mathematical Methods in the Physical Sciences. New York: John Wiley and Sons, 1983.

4. Bose, N. K. and S. K. Mitra, "Generalized Inverse of Polynomial Matrices," IEEE Transactions on Automatic Control, AC-23: 491-493 (1978).

5. Burden, R.L. and J.D. Faires. Numerical Analysis. Boston: Prindle, Weber & Schmidt, 1985.

6. Charnes, A. and M. Kirby, "Modular Design, Generalized Inverses, and Convex Programming," Operations Research, 13: 836-847 (1965).

7. Colletti, Capt Bruce. Solutions of Nonlinear Matrix Equations. MS thesis, AFIT/GOR/MA/85D-1. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1985 (AD-A163 841).

8. Dantzig, George B., "Impact of Linear Programming on Computer Development," OR/MS Today, 15: 12-17 August 1988).

9. Doma, Lt Col Abel-Monem E. Generalized Inverses of Matrices and its Applications. MS thesis, AFIT/GCS/MA /83D-1. School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB OH, December 1983 (AD A138 274).

10. Fraleigh, John B. A First Course in Abstract Algebra. Reading MA: Addison-Wesley Publishing Co, 1976.

11. Frawley, W.J. "Computer Generation of Symbolic Generalized Inverses and Applications to Physics and Data Analysis," Applications of Computer Algebra, edited by R. Pavelle. Boston: Kluwer Academic Publishers, 1985.

12. Henk Don, F. J., "The Use of Generalized Inverses in Restricted Maximum Likelihood," Linear Algebra and its Application, Vol 70: 225-240, Oct 1985.

13. Himmelblau, David M. Applied Nonlinear Programming. New York: McGraw-Hill Book Company, 1972.

14. Hooker, J.N., "Karmarkar's Linear Programming Algorithm," Interfaces, 16: 75-90, July-August 1986.

15. Hsuan, Francis et. al. "The (2)-Inverse with Applications in Statistics," Linear Algebra and its Application, 70: 241-248 (Oct 1985).

16. Hunter, Jeffrey J., "Generalized Inverses and Their Application to Applied Probability Problems," Linear Algebra and its Application, 45: 157-198 (June 1982).

17. Jones, John Jr. "Common Solutions of Sets of Matrix Equations Arising in Computer Simulation of Multi-Parameter Systems," Proceedings of the 18th Annual Pittsburgh Conference on Modeling and Simulation. 767-773. Pittsburgh: Instrument Society of America, April 1987.

18. -----. "Matrix Equations over Rings," Mathematics and Computers in Simulation, XXV: 489-492 (December 1983).

19. -----. "Multi-dimensional and Multi-parameter Systems Theory and Supercomputers," International Supercomputing Institute Inc., II: 461-470 (May 1987).

20. -----. "Solutions of Matrix Equations Containing Parameters in Multidimensional Systems Simulation," IEEE Computer Society, I: 253-263 (1985).

21. ----. Lecture Notes From MATH 799. School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OHIO, August 1988.

22. Jones, John Jr and Raymond Hill, "A Nonclassical Approach to Nonlinear Optimization Theory I," Fourth Annual Conference on Applied Mathematics: 141-149, Edmond Oklahoma: Central State University, April 1988.

23. Mendenhall W. et. al. Mathematical Statistics With Applications. Boston MA: Duxbury Press, 1986.

24. Mitra, S. K. "Common Solutions to a Pair of Linear Matrix Equations $A_1XB_1 = C_1$ and $A_2XB_2 = C_2$," Proceedings of the Cambridge Philosophical Society, 74: 213-215 (September 1973).

25. Montgomery, Douglas C. Design and Analysis of Experiments. New York: John Wiley and Sons, 1984.

26. Morris, Gerald L. and Patrick L. Odell. "Common Solutions for n Matrix Equations with Applications," Journal of the Association for Computing Machinery, 15: 272-274 (April 1968).

27. Murray, Capt Craig F. A New Method for Computing the Generalized Inverses of a Matrix and it's Application to the Lyapunov Equation. MS thesis, AFIT/GOR/MA/85D-2. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1985 (AD-A163 840).

28. Nashed, M. Generalized Inverses and Applications. New York: Academic Press, 1979.

29. Nelson, David L., Truman O. Lewis, and Thomas L. Boullion, "A Quadratic Programming technique using Matrix Pseudoinverse," The Journal of the Industrial Mathematics Society, 21: 1-21 (1971).

30. ------. "A Generalized Inverse Method for Regression Analysis with Linear Restrictions," The Journal of the Industrial Mathematics Society, 22: 1-10 (1972).

31. Noble, Ben. Applied Linear Algebra. New Jersey: Prentice-Hall, 1977.

32. Penrose, R. "A Generalized Inverse for Matrices," Proceedings of the Cambridge Philosophical Society, 51: 406-413 (January 1955).

33. Protter, Murray H. and Charles B. Morrey, Jr. Modern Mathematical Analysis. Reading, Mass: Addison-Wesley Publishing Company, 1964.

34. Ragab, Ali M. Special Nonlinear Optimization Techniques. MS thesis, AFIT/GOR/MA/83D-5. School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB OH, Dec 1983, (AD-A141 097).

35. Ralston, Anthony and E.D. Reilly JR. editors. Encyclopedia of Computer Science and Engineering. New York: Van Nostrand Reinhold Co Inc., 1983.

36. Rao, C.R. and S.K. Mitra. Generalized Inverses of Matrices and it's Application. New York: John Wiley and Sons, 1971.

37. Rao, C.R. and H. Yanai. "Generalized Inverses of Partitioned Matrices Useful in Statistical Applications," Linear Algebra and its Applications, 70: 105-113 (1985).

38. Reklaitis, G. V., A. Ravindran, and K. M. Ragsdell, Engineering Optimization: Methods and Applications. New York: Wiley Interscience, 1983.

39. Sebek, Michael. "One More Counterexample in nD Systems - Unimodular Versus Elementary Operations," IEEE Transactions on Automatic Control, 33: 499-503 (May 1988)

40. Shankland, Donn G. Quadratic Programming Using Generalized Inverses. Technical Report - AFIT TR 75-2, School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, May 1975.

41. Shurbet, G.L. and others. "Consistency of the System of Linear Matrix Equations," The Journal of the Industrial Mathematics Society, 22: 55-64 (1972).

42. Sontag, Eduardo D. "On Generalized Inverses of Polynomial and Other Matrices," IEEE Transactions on Automatic Control, AC-25: 514-517 (June 1980).

43. -----. "An Introduction to the Stabilization Problem for Parameterized Families of Linear Systems," Contemporary Mathematics, 47: 369-399 (1985).

44. Stein, Junior. Conjugate Direction Algorithms in Numerical Analysis and Optimization, 1 August 1973 - 31 January 1975. Grant Number DAHC 04-74-G-0006. Toledo Ohio: The University of Toledo, June 1975 (AD A013 418)

45. Stewart, G.W. "On the Pertubation of Psuedo-Inverses, Projections and Linear Least Squares Problem," SIAM Review, 19: 634-662 (October 1977).

46. Strang, Gilbert. Linear Algebra and its Applications. New York: Academic Press INC, 1980.

47. Urquhart, N.S. "Computation of Generalized Inverse Matrices Which Satisfy Specified Conditions," SIAM Review, 10: 216-218 (April 1968).

48. Wismer, David A. and R. Chattergy, Introduction to Nonlinear Optimization: A Problem Solving Approach, New York: North Holland, 1978.

49. Zangwill, Willard I., Nonlinear Programming: A Unified Approach, New Jersey: Prentice-Hall, 1969.

50. VAX UNIX MACSYMA Reference Manual Version II, Symbolics Inc., October 1985.

VITA

Captain Raymond R. Hill Jr. [redacted] entered the active duty Air Force in January 1977. Upon separation from active duty in 1981, he entered Eastern Connecticut State University. He graduated summa cum laude in 1983 with a Bachelor of Science in Mathematics. Upon graduation he was employed as a programmer analyst with MacSea Ltd until he reentered the Air Force through the Officer Training School (OTS) program. Upon graduation from OTS in May 1984, he attended Communications Officer technical training at Keesler AFB, Mississippi followed by an assignment to the Standard Systems Center, Gunter AFS, Alabama. He served there as Chief, Automated Systems Division until entering the School of Engineering, Air Force Institute of Technology, in June 1987. [redacted]

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AD-A202 620

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GOR/MA/88D-4	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION School of Engineering 6b. OFFICE SYMBOL (If applicable) AFIT/ENS	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Air Force Institute of Technology Wright-Patterson AFB, Ohio 45433-6583	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO / WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) See section 19 (unclassified)	
12. PERSONAL AUTHOR(S) Raymond R. Hill Jr., Capt, USAF	
13a. TYPE OF REPORT MS Thesis 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) 1988 December 15. PAGE COUNT 117
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
12	01	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Optimization, Nonlinear programming, linear algebra, polynomial matrices, matrix theory, generalized inverse

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Title: NONLINEAR OPTIMIZATION INVOLVING POLYNOMIAL MATRICES AND THEIR GENERALIZED INVERSES (unclassified)

Thesis Committee Chairman: Dr. J. Jones Jr.
Professor

10 Jan 89

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. J. Jones Jr., Professor	22b. TELEPHONE (Include Area Code) 513 - 255- 3000 22c. OFFICE SYMBOL AFIT/ENC

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

This thesis examines the applications of the generalized inverse of a matrix. In particular, use is made of the generalized inverse of a matrix containing variable elements. Such matrices are referred to as multiparameter, polynomial, or variable element matrices. The notion of a generalized inverse in fact "generalizes" the concept of a matrix inverse. A matrix inverse exists only for square, non-singular matrices. The generalized inverse extends this notion to non-square, singular matrices. The classical matrix inverse, when it exists, is a unique element of the set of generalized inverses for the matrix.

Many modern problems involve multiparameter matrices. The ability to obtain inverses for such matrices, both singular and non-singular, is a necessity in solving these problems.

This thesis consolidates the theory of generalized inverses, including extensions to multiparameter matrices. An in depth discussion is made of the ST method for computing all generalized inverses of a matrix as well as the strong interface between the ST method and the Fundamental Theorem of Linear Algebra. Finally selected application problems are solved demonstrating the utility of the generalized inverse in such problems.