June, 1978 Report ESL-R-818

ITERATIVE DECOMPOSITION OF THE LYAPUNOV AND RICCATI EQUATIONS

by

Norman August Lehtomaki

This report is based on the unaltered thesis of Norman August Lehtomaki, submitted in partial fulfillment of the requirements for the degree of Master of Science at the Massachusetts Institute of Technology in June, 1978. The research was conducted at the Decision and Control Sciences Group of the M.I.T. Electronic Systems Laboratory with support provided by the Department of Energy, Division of Electric Energy Systems under grant ERDA-E(49-18)-2087.

Electronic Systems Laboratory
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

ITERATIVE DECOMPOSITION OF THE LYAPUNOV AND RICCATI EQUATIONS

by

Norman August Lehtomaki

B.S.E.E., Michigan State University

(1974)

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May, 1978

Signature of Author Norman A. Lehtomaki
Department of Electrical Engineering and Computer Science
May 12, 1978

Certified by Nils R. Sandell Jr.
Thesis Supervisor

Accepted by _____
Chairman, Departmental Committee on Graduate Students

ITERATIVE DECOMPOSITION OF THE LYAPUNOV AND RICCATI EQUATIONS

by

Norman August Lehtomaki

Submitted to the Department of Electrical Engineering and Computer Science on May 12, 1978 in Partial Fulfillment of the Requirements for the Degree of Master of Science.

ABSTRACT

This thesis presents decomposition methods for solving large scale system Lyapunov and Riccati equations. These methods are iterative and utilize the system characteristics of weak coupling among subsystems or widely different time constants between subsystems. The algorithms have a hierarchical structure of coordinator and subproblem. The coordinator, with a coordinating input to the subproblems formed on the basis of the previous solutions of all subproblems, guides the subproblem solutions to the global problem solution.

THESIS SUPERVISOR: Nils R. Sandell, Jr.

TITLE: Associate Professor of Electrical Engineering and Computer Science

ACKNOWLEDGEMENT

I would like to thank Professor Nils R. Sandell, Jr. for his very real guidance and encouragement during the preparation of this thesis. His assistance has been invaluable. I would also like to express my appreciation to Douglas Looze, Demosthenis Teneketzis and Dr. Alan Laub for their helpful and stimulating discussions. Many thanks go to Ms. Debbie Merritt, who typed this thesis. Finally, I thank my wife for her constant and abiding love.

TABLE OF CONTENTS

TABLE OF CONTENTS (Contd.)

CHAPTER I

INTRODUCTION

1.1 The Lyapunov and Riccati Equations in Control Theory

The application of modern control theory to real problems involves solving the Lyapunov and Riccati equations accurately and in an efficient manner. These equations arise in a number of problems such as the stability of linear systems or in the evaluation of cost functions in optimal control and covariance matrices in filtering and estimation [1], [2].

Consider the linear system

$$\dot{x}(t) = A^T x(t) + w(t) \ , \quad x(s) \varepsilon R^n$$

where $w(t)$ is formally a stationary, zero mean, white noise process with spectral density matrix $Q \geq 0$ (i.e. Q is positive semidefinite).

The steady state variance of this system is given by P where

$$A^T P + PA + Q = 0$$

for stable A^T and P is the unique positive semidefinite solution of the algebraic Lyapunov equation. The Lyapunov equation also arises in connection with the Lyapunov function

$$V(x,t) = x^T(t) P x(t)$$

for the linear system

$$\dot{x}(t) = A\,x(t)\ ,\quad x(0) = x_o \quad .$$

Differentiating $V(x,t)$ we obtain

$$\dot{V}(x,t) = x^T(t)[A^TP + PA]x(t)$$

$$= - x^T(t)Qx(t)$$

where Q is positive semidefinite since $V(x,t)$ is a Lyapunov function. If P and Q are both positive definite then the system is guaranteed to be stable and $x(t)$ approaches zero as t tends to infinity [3]. Moreover, if J is the associated quadratic cost of the state trajectory over the infinite time interval so that

$$J = \int_0^\infty x^T(t)Qx(t)dt$$

then, because the system is stable

$$J = x_o^TPx_o \quad .$$

We may also obtain from the variation of constants formula and the integral formula for J that

$$P = \int_0^\infty e^{A^Tt}Qe^{At}dt \quad .$$

Turning to the Riccati equation, consider the deterministic optimal control problem

$$\min_{u}\{J(u) = \int_{0}^{\infty} x^{T}(t)Qx(t) + u^{T}(t)Ru(t)dt\}$$

subject to $\dot{x}(t) = Ax(t) + Bu(t)$, $x(0) = x_{o}$

where $R > 0$, $Q \geq 0$ and $J(u)$ is the cost.

Provided $[A,B]$ is controllable and $[A,\sqrt{Q}]$ is observable, the minimizing control and minimum cost are given by

$$u(t) = - R^{-1}B^{T}Kx(t)$$

$$J = x_{o}^{T}Kx_{o}$$

where K is the unique positive definite solution of

$$A^{T}K + KA + Q - KBR^{-1}B^{T}K = 0$$

the algebraic Riccati equation [28].

1.2 Decomposition and Decentralization in Large Scale Systems

As the dimension of the state space increases and becomes very large, as may happen with interconnected systems, the task of computing solutions to linear-quadratic control problems using general purpose algorithms becomes exceedingly difficult. These difficulties arise because the number of operations specified by an algorithm and the associated storage requirements exceed reasonable time and size limits dictated by the computing facility. Even if these limits are not exceeded, the numerical accuracy achieved may not be sufficient, due to round-off errors.

However, if the dimension of the problem is very large, there is usually a special sparse structure that can be taken advantage of to reduce these computational difficulties. There are two approaches to do this, which are somewhat related; they are decomposition and decentralization. Briefly decomposition is concerned with the reduction of off-line computation required to obtain a control law. Decomposition is attractive, since it breaks the original problem into several subproblems that are simpler and easier to solve and then reconstructs the centralized solution. Decentralization on the other hand, is concerned with the reduction of on-line communication and computation implicit in a defined control law. Formulation of decentralized control laws leads to a number of smaller subproblems or infimals, which reduce the

on-line computation and communication. This however, may increase the off-line computation to obtain the control law since the infimals may not be simpler and easier to solve than the original problem [4]. The approach considered here is similar in some respects to that of Laub and Bailey [5] and is one of decomposition. However, since subproblems are defined in decomposition approaches, there is a certain off-line decentralization of information about system process parameters. Laub and Bailey point this out in defining a decomposition for the centralized decision problem leading to a decentralized solution. In this decomposition, each controller or infimal decision maker needs only limited information about the system process parameters and only a part of the state. A supremal decision maker called the coordinator directs the decisions of each infimal decision maker or controller by supplying a correction term to each controller's decision. To obtain this type of decomposition and decentralization, they pose the problem as a classical least squares problem in Hilbert space. They then obtain conditions under which this decentralized decision making with coordination can achieve the optimal centralized decision process. For the linear regulator problem over the infinite interval, this results in at best, a partially closed loop solution.

In the approach taken here, the only decentralized aspect

will be that of the decentralization of information about the system parameters. The solution obtained will be in closed loop form, but will not be implementable in a decentralized fashion, since it is merely the centralized solution in partitioned form.

To contrast the decomposition approach of Laub and Bailey to the one taken here, it can be noted for the linear regulator problem that at each iteration, they solve the minimization of the cost function exactly while approximating successively the linear constraint of the differential system equation. In this thesis, the necessary conditions for optimality are decomposed instead of the original problem. However, this can be interpreted as satisfying both the state equation constraint and the minimization of the cost function approximately and then obtaining convergence conditions under which the successive constraints and cost functions converge to the true state equation and true cost function.

1.3 Summary of Thesis

The Lyapunov and Riccati equations of control theory are presented and analyzed in the case where special structure may be taken advantage of. In Chapter II, the relationship of the Lyapunov equation to linear vector equations and techniques for solving them is explored. Chapter III discusses an iterative method to solving the Lyapunov and Riccati equations motivated by singular perturbation theory which suggests the particular decomposition. This decomposition is identified with the separation of time scales between subsystems of a large scale system. Chapter IV presents a decomposition algorithm motivated by the idea of weakly coupled subsystems of a large scale system being approximated by independent subsystems. This method is applied to a power system model to obtain some numerical results. Chapter V presents some conclusions and suggestions for further research along with the appendix which contains a somewhat lengthy algebraic derivation of the two-time scale decomposition algorithm for the Riccati equation.

The main contributions of this thesis are:

1) The presentation of a new two time scale, iterative decomposition algorithm for the Lyapunov and Riccati equations.

2) The extension of the weak coupling iterative decomposition algorithm by use of a relaxation parameter.

3) The showing of the way in which a special canonical

form may be useful in reducing computations for the decomposition algorithm.

CHAPTER II

ITERATIVE SOLUTION METHODS

2.1 MATRIX PRELIMINARIES

Vector and Matrix Norms

It is useful to have a single number to measure the size of a vector or matrix. This number is called the norm of the vector or matrix and is denoted by $||\cdot||$, [17].

For vector norms the following relations must hold

$$||x|| > 0 \quad \text{unless} \quad x = 0$$

$$||\alpha x|| = |\alpha| \, ||x|| \quad \text{for any complex scalar} \quad \alpha$$

$$||x + y|| \leq ||x|| + ||y|| \quad .$$

Three vector norms that are commonly used are

$$||x||_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p} \quad (p = 1,2,\infty)$$

where $||x||_\infty$ is interpreted as $\max_i |x_i|$. The norm $||x||_2$ is the usual Euclidean length of the vector x. An important inequality is

$$|x^T y| \leq ||x||_2 ||y||_2$$

the Schwarz inequality.

Similarly, for the norm of a matrix A, denoted $||A||$ the following relations must hold

$$||A|| > 0 \quad \text{unless} \quad A = 0$$

$$||\alpha A|| = |\alpha| \; ||A|| \quad \text{for any complex scalar} \quad \alpha$$

$$||A + B|| \leq ||A|| + ||B||$$

$$||AB|| \leq ||A|| \; ||B|| \quad .$$

Corresponding to each vector norm the associated matrix norm defined by

$$||A|| = \max_{x \neq 0} \frac{||Ax||}{||x||}$$

satisfies the conditions for a matrix norm and is said to be subordinate to the vector norm.

Corresponding to the three vector norms the subordinate matrix norms are

$$||A||_1 = \max_j \sum_i |a_{ij}|$$

$$||A||_\infty = \max_i \sum_j |a_{ij}|$$

$$||A||_2 = \max_i [\lambda_i (A^H A)]^{1/2}$$

where $\lambda_i(A^H A)$ are the eigenvalues of $A^H A$ and A^H denotes the conjugate transpose of A.

From the definition of these norms it is apparent that

$$||Ax||_p \leq ||A||_p ||x||_p \; ; \quad p = 1,2,\infty$$

is satisfied for all x. Any matrix norm which satisfies this inequality is said to be consistent or compatible. Another matrix norm which is used frequently that is compatible with the vector norm $||\cdot||_2$ is the Euclidean norm. The Euclidean norm for a matrix A is defined by

$$||A||_E = [\sum_i \sum_j |a_{ij}|^2]^{1/2} \quad .$$

The $||A||_2$ norm is referred to as the spectral norm. Some useful relationships involving the spectral and Euclidean norms that can be developed are

$$||A||_2 \leq ||A||_E \leq n^{1/2} ||A||_2$$

where A is an $n \times n$ matrix. These inequalities follow from the fact that $A^H A$ is positive semidefinite and

$$\max|\lambda(A^H A)| \leq ||A||_E^2 = \mathrm{tr}(A^H A) \leq n \max|\lambda(A^H A)| \quad .$$

Also, if λ is an eigenvalue of A and x is a corresponding eigenvector, then for consistent matrix and vector norms

$$||Ax|| = |\lambda| \; ||x|| \leq ||A|| \; ||x||$$

$$|\lambda| \leq ||A|| \quad .$$

From this we can obtain

$$||A||_2^2 = \max|\lambda(A^H A)| \leq ||A^H A||_\infty \leq ||A||_1 ||A||_\infty \quad .$$

2.2 THE LYAPUNOV EQUATION AS A VECTOR EQUATION

The Kronecker Product

The useful notion of the Kronecker product and its elementary properties is presented here to facilitate later analysis of the Lyapunov equation.

The Kronecker product is discussed by Bellman [11] in which some useful elementary identities are stated. The Kronecker product of two matrices A and B is denoted by $A \otimes B$ and defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & & a_{2n}B \\ \vdots & & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nn}B \end{bmatrix}$$

In order to justify the use of the term product, it may be seen by using block matrix multiplication and the definition of Kronecker product that the following identities hold:

$$(A \otimes B) \otimes C = A \otimes (B \otimes C)$$

$$(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

$$(aA \otimes bB) = ab(A \otimes B) \quad \text{for scalars } a,b \ .$$

Eigenvalues of $A \otimes B$

The eigenvalues of $A \otimes B$ have a simple relationship to the eigenvalues of A and B. Suppose α and β are eigenvalues of A and B respectively, with corresponding eigenvectors x and y. Consider the Kronecker product

$$(Ax) \otimes (By) = (\alpha x) \otimes (\beta y)$$

which can be written alternatively as

$$(A \otimes B)(x \otimes y) = \alpha\beta(x \otimes y)$$

using the previous identities. Thus $\alpha\beta$ is an eigenvalue of $A \otimes B$ with corresponding eigenvector $(x \otimes y)$. Note, however, that for multiple eigenvalues of a defective matrix (i.e. not diagonalizable) the generalized eigenvectors or principal vectors of grades greater than one will not be of the form $x \otimes y$. A vector x_m is called a generalized eigenvector of rank m (or a principal vector of grade m) [17] of the matrix A and eigenvalue λ if $(A - \lambda I)^m x_m = 0$ but $(A - \lambda I)^{m-1} x_m \neq 0$.

Kronecker Products and Linear Equations

In studying the Lyapunov equation, it will be useful to examine the linear matrix equation

$$AXB = C$$

where $A \epsilon R^{nxn}$, $B \epsilon R^{mxm}$ and $X, C \epsilon R^{nxm}$. This equation can be decomposed into two separate problems

$$AX = Y \quad \text{and} \quad YB = C \quad .$$

If X,Y and C are composed of vector columns x_i, y_i and c_i respectively for $i = 1$ to m, then it is apparent that

$$\begin{bmatrix} A & & & \\ & A & & \\ & & \ddots & \\ & & & A \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

and

$$[y_1 y_2 \cdots y_m] \; B = [c_1 c_2 \cdots c_m] \quad .$$

Hence,

$$c_i = \sum_{k=1}^{m} b_{ki} y_k = [b_{1i} I \; b_{2i} I \cdots b_{mi} I] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad i = 1 \text{ to } m$$

where b_{ij} is the (i,j)th element of B. As i runs from one to m this can be rewritten as

$$\begin{bmatrix} b_{11}I & b_{21}I & \cdots & b_{m1}I \\ b_{12}I & b_{22}I & & \vdots \\ \vdots & & \ddots & \vdots \\ b_{1m}I & \cdots & \cdots & b_{mm}I \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} .$$

Now, by substitution for Y we obtain

$$\begin{bmatrix} b_{11}I & b_{21}I & \cdots & b_{m1}I \\ b_{12}I & b_{22}I & & \\ \vdots & \vdots & \ddots & \\ b_{1m}I & b_{2m}I & \cdots & b_{mm}I \end{bmatrix} \begin{bmatrix} A & & \\ & \ddots & \\ & & A \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}$$

or

$$\begin{bmatrix} b_{11}A & b_{21}A & \cdots & b_{m1}A \\ b_{12}A & b_{22}A & & \vdots \\ \vdots & & \ddots & \vdots \\ b_{1m}A & \cdots & \cdots & b_{mm}A \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} .$$

Now by definition of the Kronecker we can write

$$(B^T \otimes A)x_c = c_c$$

where

$$x_c = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad \text{and} \quad c_c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} \quad x_c, c_c \in R^{nm} .$$

We have now obtained a linear vector equation from a matrix linear equation. However, suppose that it is desired to construct x_R and c_R of the rows of the matrices X and C instead of the columns as we have done previously. This can be accomplished simply by considering the columns of X^T and C^T so that for the original problem we substitute the equivalent problem

$$B^T X^T A^T = C^T$$

and using our previous result we obtain

$$(A \otimes B^T) x_R = c_R$$

where

$$x_R = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ and } c_R = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

with

$$X^T = [x_1 x_2 \ \ldots \ x_n] \quad \text{and} \quad C^T = [c_1 c_2 \ \ldots \ c_n] \ .$$

Furthermore, the relation between x_c and x_r is given by

$$x_c = P x_R$$

with

$$P^T P = P P^T = I$$

where P is an orthogonal permutation matrix. P is defined by

$$P = \begin{bmatrix} U_{11} & U_{21} & \cdots & U_{n1} \\ U_{12} & U_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ U_{1m} & \cdots & & U_{nm} \end{bmatrix}$$

where each U_{ij} is a $n \times m$ matrix which has elements of zero except for the (i,j)th element which is a one. In the case $n = m$ P is then symmetric.

It has been shown previously that the corresponding sets of equations in terms of Kronecker products for $AXB = C$ is:

$$AXB = C \leftrightarrow (A \otimes B^T)x_R = c_R$$

$$AXB = C \leftrightarrow (B^T \otimes A)x_c = c_c \quad .$$

Hence,

$$(A \otimes B^T)x_R = P^T c_c$$

$$(A \otimes B^T)P^T x_c = P^T c_c$$

$$[P(A \otimes B^T)P^T]x_c = c_c$$

which implies that

$$(B^T \otimes A) = P(A \otimes B^T)P^T$$

or that $B^T \otimes A$ and $A \otimes B^T$ are similar and therefore have the same elementary divisors.

The Lyapunov Equation and Kronecker Products

Using the formulas for Kronecker products, we may rewrite the Lyapunov equation

$$A^T K + KA + Q = 0$$

in the form

$$K_A k_R = - q_R$$

where

$$K_A = A^T \otimes I + I \otimes A^T$$

and k_R and q_R are vectors comprised of the rows of the matrices K and Q respectively.

Eigenvalues of K_A

If A has eigenvalues λ_i then

$$A^T x_i = \lambda_i x_i$$

where x_i is an eigenvector of A^T corresponding to λ_i. Consider the Kronecker products

$$A^T x_i \otimes x_j = \lambda_i x_i \otimes x_j$$

$$x_i \otimes A^T x_j = \lambda_j x_i \otimes x_j$$

which when added give

$$(A^T \otimes I + I \otimes A^T)(x_i \otimes x_j) = (\lambda_i + \lambda_j)(x_i \otimes x_j) \quad .$$

Hence, the eigenvalues of K_A are simply $\lambda_i + \lambda_j$ counting multiplicities. If A possesses n linearly independent eigenvectors (i.e. A is not defective) then K_A has only linear elementary divisors and n^2 linearly independent eigenvectors

$$z_{ij} = x_i \otimes x_j \quad i = 1 \text{ to } n \; ; \; j = 1 \text{ to } n \; .$$

Givens [12] has shown that if A has a Jordan form with Jordan blocks of size p_i , $i = 1,2,\ldots,m$, where m is the number of Jordan blocks, then K_A has Jordan blocks of size $p_i + p_j - 1, p_i + p_j - 3, \ldots, |p_i - p_j| + 1$; $i = 1$ to m and $j = 1$ to m .

Norm Relationships for K_A

Since $K_A^T = A \otimes I + I \otimes A$, $A \varepsilon R^{nxn}$

$$K_A = \begin{bmatrix} a_{11}I + A & a_{12}I & \cdots & a_{1n}I \\ a_{21}I & a_{22}I + A & & a_{2n}I \\ \vdots & & \ddots & \vdots \\ a_{n1}I & a_{n2}I & \cdots & a_{nn}I + A \end{bmatrix} .$$

Suppose for $1 \leq r \leq n$ and $i = 1,2,\ldots,n$

$$||a_r|| \geq ||a_i||$$

where

$$A = [a_1 a_2 \ \ldots \ a_n] \quad ; \quad a_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix}$$

then

$$||K_A||_\infty = ||K_A^T||_1 = 2||a_r|| = 2||A||_1 \quad .$$

This is necessary since K_A^T has one column with every element of a_r appearing twice except for element a_{rr} which is replaced by $2a_{rr}$ and the remaining elements of the column being zero.

Similarly for the rows of K_A we can deduce that

$$||K_A||_1 = ||K_A^T||_\infty = 2||A||_\infty \quad .$$

For the Euclidean norm, if

$$A = B + D$$

where D is diagonal and B has a zero diagonal,

$$||A||_E^2 = ||B||_E^2 + ||D||_E^2 \quad .$$

Hence,

$$||K_A||_E^2 = ||K_B||_E^2 + ||K_D||_E^2$$

must hold.

Now $||K_B||_E^2 = 2n||B||_E^2$ since B has a zero diagonal, and

$$||D||_E^2 = \sum_{i=1}^{n} |a_{ii}|^2$$

$$||K_D||_E^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ii} + a_{jj}|^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\Big(|a_{ii}|^2 + |a_{jj}|^2 + 2|a_{ii}a_{jj}|\Big)$$

$$||K_D||_E^2 = 2n||D||_E^2 + 2\,tr^2(A) \quad .$$

Therefore

$$||K_A||_E^2 = 2n||A||_E^2 + 2(trA)^2 \quad .$$

The spectral norm of K_A has only an inequality relationship to the spectral norm of A;

$$||K_A||_2 = ||K_A^T||_2 \leq ||A \otimes I||_2 + ||I \otimes A||_2$$

$$||A \otimes I||_2^2 = \rho[(A \otimes I)^H(A \otimes I)] = \rho[(A^HA \otimes I)] = \rho(A^HA)$$

where

$$\rho(X) = \max_i[|\lambda_i(X)|] \quad .$$

Also

$$||I \otimes A||_2^2 = \rho[(I \otimes A)^H(I \otimes A)] = \rho(A^HA) = ||A||_2^2$$

and hence

$$||A||_2 \leq ||K_A||_2 \leq 2||A||_2 \quad .$$

2.3 ITERATIVE METHODS FOR THE LYAPUNOV EQUATION

Present iterative methods [6] for solving the Lyapunov equation

$$A^T K + KA + Q = 0 \ , \quad K \epsilon R^{nxn}$$

in continuous time transform the equation into a discrete time version which is then iterated to stationarity. The discrete time Lyapunov equation is

$$K_{m+1} = F^T K_m F + G \ , \quad K_o = 0 \quad .$$

Convergence of this iteration is ensured if all of the eigenvalues of F lie within the unit disk. However, the rate of convergence may be extremely slow, if F has eigenvalues close to the unit circle, but may be accelerated using the following iteration

$$S_{m+1} = F_m^T S_m F_m + S_m \ , \quad S_o = G$$

$$F_{m+1} = F_m^2 \ , \quad F_o = F$$

where $\{S_m\} \rightarrow K$ and $S_m = K_{2^m}$.

This iteration is nonstationary and can be shown to converge quadratically for the case when F is diagonalizable. To see this write K_m as

$$K_m = \sum_{i=0}^{m-1} (F^i)^T G F^i$$

then

$$K_{2m} = (F^m)^T K_m F_m + K_m$$

so that with $m = 2^p$ we obtain

$$K_{2^{p+1}} = (F^{2^p})^T K_{2^p} F^{2^p} + K_{2^p}$$

the described iteration above. To simplify the convergence analysis, the iterations may be written (using Kronecker product notation of Section 2.2) as vector iterations

$$k_{m+1} = (F \otimes F)^T k_m + g \ , \quad k_o = 0$$

$$s_{m+1} = [(F \otimes F)^{2^m}]^T s_m + s_m \ , \quad s_o = g$$

where k_m , s_m and g are the vectors formed from the rows of K_m , S_m and G respectively, such that the first row forms the first n elements of the vector, the second row forms the second n elements of the vector and so on. Now, defining the error at each step in the iterations quadratic convergence may be demonstrated. Since the solution k satisfies the equation

$$k = (F \otimes F)^T k + g$$

let

$$e_m = k - k_m \quad \text{and} \quad \hat{e}_m = k - s_m$$

then with

$$B = (F \otimes F)^T$$

$$e_{m+1} = B\, e_m$$

$$e_{m+r} = B^r e_m$$

so that with $m = 0$ and $r = 2^p$

$$\hat{e}_p = e_{2^p} = B^{2^p} e_o = B^{2^p} k \quad .$$

When F is diagonalizable and stable, B is diagonalizable and stable. This is true because there exists an X such that

$$F = X^{-1}\Lambda X \quad \text{where} \quad \Lambda \quad \text{is diagonal}$$

so that

$$B = (X^{-1} \otimes X^{-1})(\Lambda \otimes \Lambda)(X \otimes X)$$

$$B = (X \otimes X)^{-1}(\Lambda \otimes \Lambda)(X \otimes X) = T^{-1}DT$$

$$\lambda_{ij}(B) = \lambda_i(\Lambda)\ \lambda_j(\Lambda) \quad \text{for} \quad i = 1 \text{ to } n \quad \text{and} \quad j = 1 \text{ to } n \quad .$$

Let

$$D = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_N \end{bmatrix} \quad ; \quad N = n^2$$

and consider the following norms for any $\tilde{k}^T = [\tilde{k}_1, \tilde{k}_2, \ldots, \tilde{k}_N]$

$$||D^{2^{m+1}}\tilde{k}||_2^2 = \sum_{i=1}^{N} |d_i^{2^{m+1}}\tilde{k}_i|^2$$

$$||D^{2^m}\tilde{k}||_2^2 = \sum_{i=1}^{N} |d_i^{2^m}\tilde{k}_i|^2$$

$$||D^{2^m}\tilde{k}||_2^4 \geq \sum_{i=1}^{N} |d_i^{2^{m+1}}\tilde{k}_i|^2 |\tilde{k}_i|^2$$

since

$$||x||_2^4 \geq \sum_{i=1}^{n} x_i^4$$

and define

$$\alpha_i = \left\{ \begin{array}{ll} 0 , & \tilde{k}_i = 0 \\ \dfrac{1}{|\tilde{k}_i|} , & \tilde{k}_i \neq 0 \end{array} \right\} \quad ; \quad i = 1 \text{ to } N$$

$$\alpha = \max_i \alpha_i \quad ; \quad i = 1 \text{ to } N$$

hence

$$\alpha^2 ||D^{2^m}\tilde{k}||_2^4 \geq \sum_{i=1}^{N} |d_i^{2^{m+1}}\tilde{k}_i|^2 \alpha_i^2 |\tilde{k}_i|^2 = \sum_{i=1}^{N} |d_i^{2^{m+1}}\tilde{k}_i|^2$$

$$= ||D^{2^{m+1}}\tilde{k}||_2^2$$

or

$$\alpha||D^{2^m}\tilde{k}||_2^2 \geq ||D^{2^{m+1}}\tilde{k}||_2$$

$$\alpha||TT^{-1}D^{2^m}\tilde{k}||_2^2 = \alpha||D^{2^m}\tilde{k}||_2^2 \geq ||D^{2^{m+1}}\tilde{k}||_2 \quad .$$

Multiplying both sides by $||T^{-1}||_2$ and using matrix norm inequalities

$$||T||_2^2||T^{-1}||_2\alpha||T^{-1}D^{2^m}\tilde{k}||_2^2 \geq ||T^{-1}||_2||D^{2^{m+1}}\tilde{k}||_2$$

$$\geq ||T^{-1}D^{2^{m+1}}\tilde{k}||_2$$

with

$$\tilde{k} = Tk$$

$$(\alpha||T||_2^2||T^{-1}||)||B^{2^m}k||_2^2 \geq ||B^{2^{m+1}}k||$$

and finally

$$||\hat{e}_{m+1}||_2 \leq \beta(\hat{e}_o)||\hat{e}_m||_2^2 \qquad \forall_m$$

where

$$\beta(\hat{e}_o) = \alpha||T^{-1}||_2||T||_2^2$$

the desired result of quadratic convergence.

The methods that use this approach differ only in their transformations from continuous time to discrete time. One transformation is

$$F = e^{Ah} \quad , \quad G = \int_o^h (e^{A\sigma})^T Q e^{A\sigma} d\sigma \quad .$$

Davison and Man [7] use the approximation

$$G \tilde{=} Qh \; ; \quad F = e^{Ah}$$

where h is required to be small in order for the approximation to be good. Reducing h, however, leads to an increase in the number of iterations required which in turn limit the accuracy obtainable due to round off error. This problem can be circumvented to some degree if a better approximation for G is used by including more terms of the series expansion for G. Another transformation utilized by Smith [8] which is better uses a matrix linear fractional transformation. In this case the left half plane is mapped into the unit disk by

$$F = - (A - aI)^{-1}(A + aI)$$

$$G = \frac{(A^T - aI)^{-1}Q(A - aI)^{-1}}{2a}$$

where the choice of $a = [|\lambda_{min}(A)\lambda_{max}(A)|]^{1/2}$ for real $\lambda_i(A)$ minimizes the spectral radius of F. The operation count for this algorithm is $(3\sigma + r)n^3$ where σ is the number of iterations required.

2.4 ITERATIVE METHODS FOR LINEAR VECTOR EQUATIONS

Since the Lyapunov equation may be written as

$$K_A k_R = - q_R \ , \quad K_A \varepsilon R^{n^2}$$

it is simply a linear vector equation that may be solved by a number of numerical methods that are currently available. Note that if Q is symmetric the number of equations to solve can be reduced from n^2 to $n(n + 1)/2$.

For such linear systems of order n^2 direct methods will require $O(n^6)$ arithmetic operations with even the forward and backward substitution requiring $O(n^4)$ operations. Thus this approach can obviously not compete with methods developed specifically for the Lyapunov equation which require $O(n^3)$ operations. This remains essentially true even if sparsity is exploited in the direct methods. However, there is theoretical insight to be gained from this formulation of the Lyapunov equation. Therefore for the rest of this chapter the equation $Ax = b$ will be studied.

Iterative Methods for $Ax = b$

The iterative methods discussed here are referred to as norm reducing methods [18] since they start with an arbitrary first approximation to the solution and successively reduce the norm of the error. There has been much research in this area that has been directed to the solution of large sparse

systems such as arise in the solution of partial differential equations [13,14].

The basic idea of these methods is to split the A matrix of $Ax = b$ into two parts A_o and A_1 and form the iteration

$$A_o x_{K+1} = - A_1 x_K + b$$

where

$$A = A_o + A_1 \quad .$$

To be an effective algorithm the splitting should be of such a nature that x_{K+1} is easy to compute. Therefore A_o should be simple and invertible. Usually A_o is diagonal or triangular or block diagonal. To ensure that the algorithm converges to the correct solution for any initial vector x_o it is necessary and sufficient that

$$\rho(- A_o^{-1} A_1) = \rho(A_o^{-1} A_1) < 1 \quad .$$

If the iteration is subtracted from the true equation $A_o x = - A_1 x + b$ the form of the error at each step results:

$$A_o e_{K+1} = - A_1 e_K$$

where

$$e_K = x - x_K$$

or

$$e_{K+1} = - A_o^{-1} A_1 e_K$$

$$e_K = (- A_o^{-1}A_1)^K e_o$$

so that

$$||e_K|| \leq ||(A_o^{-1}A_1)^K|| \; ||e_o|| \quad .$$

Now, if $\rho(A_o^{-1}A_1) < 1$ then for some K not necessarily small

$$||(A_o^{-1} \; A_1)^K|| < 1$$

and it is entirely possible for some K_o iterations for $||(A_o^{-1}A_1)^K||$ to increase although eventually it must approach zero as K tends to infinity. The spectral radius is only the asymptotic rate of convergence. If $A_o^{-1}A_1$ has a full set of eigenvectors y_i and eigenvalues λ_i this is clearly seen since

$$e_K = c_1\lambda_1^K y_1 + \ldots + c_n\lambda_n^K y_n$$

where

$$e_o = c_1y_1 + c_2y_2 + \ldots + c_ny_n$$

$$c_i = \text{constants}$$

and the largest $|\lambda_i|$ will eventually be dominant and govern the convergence.

Obviously, different splittings result in different methods with different rates of convergence. For $A = L + D + U$ where L is strictly lower triangular, D is diagonal and

U is strictly upper triangular, the well known methods are classified by the following table.

$A_o(w)$	$A_1(w)$	Method
D	L + U	Jacobi
D + L	U	Gauss Seidel
D/w	$(1 - \frac{1}{w})D + L + U$	Simultaneous overrelaxation*
D/w + L	$(1 - \frac{1}{w})D + U$	Successive overrelaxation**
* $w \neq 0$ ** $0<w<2$	w is the relaxation parameter and $A = A_o(w) + A_1(w)$	

Postponing explanation of what w is for a moment, a brief discussion of the first two methods will be given. Firstly, one of the simplest splittings of A in which A_o is easily invertible is $A_o = D$, the Jacobi iteration. However, the Gauss Seidel iteration is just as easy to invert even though it does not appear so at first glance. The reason is that the components of x_{K+1}

$$x_{K+1}^T = [x_1^{K+1}, x_2^{K+1}, \ldots, x_n^{K+1}]$$

are computed in order from x_1^{K+1} to x_n^{K+1}. This allows the Gauss Seidel iteration to be written as

$$Dx_{K+1} = - Lx_{K+1} - Ux_K + b \quad .$$

Since L is strictly lower triangular, Lx_{K+1} can be included on the right hand side of the equation because in computing any component of x_{K+1} only components of x_{K+1} that have already been computed are present in Lx_{K+1}. Hence, A_o for the Gauss Seidel method is just as easy to invert as the A_o for the Jacobi method since all it involves is updating the components of x_K as soon as the $K + 1$ updates become available. This cuts the necessary storage in half from the Jacobi method since x_{K+1} and x_K always can be stored in n storage locations. The Jacobi method is required to keep all the components of x_K until the calculation of x_{K+1} is complete but does have the advantage that all of the components of x_{K+1} could be computed simultaneously on a suitable processor.

Turning to the next two methods and the idea of relaxation of algorithms, we observe that if w, the relaxation parameter, is equal to one that the simultaneous relaxation method reduces to the Jacobi method and the successive over-relaxation method reduces to the Gauss Seidel method, which suggests their origin. Relaxation techniques applied to the Jacobi and Gauss Seidel methods yield the latter two methods in the table, and try to inprove the convergence rate of the algorithms by attempting to reduce the spectral radius of $A_o^{-1}A_1$, $\rho(A_o^{-1}A)$. The way this is done will be illustrated using the Gauss Seidel method.

The Gauss Seidel method rearranged slightly is

$$x_{K+1} = - D^{-1}Lx_{K+1} - D^{-1}Ux_K + D^{-1}b \quad .$$

Trying to improve on this algorithm's estimate of what x_{K+1} should be, we define a new iteration

$$x_{K+1} = x_K + w\{\tilde{x}_{K+1} - x_K\} = (1 - w)x_K + w\tilde{x}_{K+1}$$

where $\tilde{x}_{K+1}$ is the old Gauss Seidel estimate of what x_{K+1} should be, given all the components of x_{K+1} and x_K that have been computed so far, so that

$$\tilde{x}_{K+1} = - D^{-1}Lx_{K+1} - D^{-1}Ux_K + D^{-1}b$$

and

$$x_{K+1} = x_K + w\{-D^{-1}Lx_{K+1} - D^{-1}Ux_K - x_K\} + wD^{-1}b$$

$$x_{K+1} = - wD^{-1}Lx_{K+1} + [I(1 - w) - wD^{-1}U]x_K + wD^{-1}b$$

or

$$(D + wL)x_{K+1} = [(1 - w)D - wU]x_K + wb$$

$$\underbrace{(\frac{D}{w} + L)}_{A_o}x_{K+1} = - [\underbrace{(1 - \frac{1}{w}) + U}_{-A_1}]x_K + b \quad .$$

Note w allows us to correct more or less than the original Gauss Seidel method and that the new x_{K+1} is a weighted mean of x_K and $\tilde{x}_{K+1}$.

Relaxation attempts to overcorrect if the algorithm is converging monotonically to the solution and undercorrect if the iterates oscillate about the solution, damping the oscillation. Obviously, if one component of the vector iterates is converging monotonically to its limit point while another is oscillating about its limit point, there are conflicting objectives to achieve using a single relaxation parameter w. In this case, optimal relaxation finds the best compromise. Very roughly speaking, if the oscillation of one component hurts the convergence the most, then the best policy would be to undercorrect at each step. The reverse would be true if a slow monotone convergence in one component was the most damaging to the convergence of the algorithm.

If relaxation is effective in reducing the spectral radius of $A_o^{-1}A_1$, it can lead to a significant reduction in the arithmetic computation. Determination of an optimum relaxation factor w_o is therefore of interest. The original iteration

$$A_o x_{K+1} = - A_1 x_K + b$$

can be written with relaxation as

$$x_{K+1} = Bx_K + c$$

where

$$B = B(w) = - A_o^{-1}(w) A_1(w)$$

$$c = c(w) = A_o^{-1}(w) b \quad .$$

For the four methods discussed $B(w)$ is given by:

$B(w)$	Method
$-D^{-1}(L + U)$	Jacobi
$-(D + L)^{-1}U$	Gauss Seidel
$I(1 - w) - wD^{-1}(L + U)$	Simultaneous overrelaxation
$(I + wD^{-1}L)^{-1}[(1 - w)I - wD^{-1}U]$	Successive overrelaxation

In the simultaneous overrelaxation method, the eigenvalues of $B(w)$ are

$$\lambda(B) = 1 - w + w\lambda(J)$$

where

$$J = -D^{-1}(L + U) = B(1) \quad .$$

If all the eigenvalues of J, $\lambda_i(J)$ are real and $\lambda_1(J) < \lambda_2(J) < \ldots < \lambda_n(J)$ then for $\lambda_i(J) < 1$ for all i or $\lambda_i(J) > 1$ for all i the optimum relaxation factor, w_o and the minimum spectral radius, $\rho[B(w_o)]$ are given by:

$$w_o = \frac{2}{2 - \lambda_1(J) - \lambda_n(J)}$$

$$\rho[B(w_o)] = \left| \frac{\lambda_n(J) - \lambda_1(J)}{2 - \lambda_n(J) - \lambda_1(J)} \right| < 1 \quad .$$

If $\lambda_n(J) = -\lambda_1(J) < 1$ then $w_o = 1$ and $\rho[B(w_o)] = \lambda_n(J)$. This is true since for $\lambda_1(J) < \lambda_2(J) <\ldots< \lambda_n(J) < 1$

$$w\lambda_i(J) + 1 - w \geq 1 \quad \text{for} \quad w \leq 0$$

therefore $w > 0$ and it follows that

$$w\lambda_1(J) + (1 - w) <\ldots< w\lambda_n(J) + (1 - w) < 1$$

and hence $|w_o\lambda_1(J) + (1 - w_o)| = |w_o\lambda_n(J) + (1 - w_o)|$ which implies

$$w_o(1 - \lambda_1(J)) - 1 = w_o(\lambda_n(J) - 1) + 1 = \rho[B(w_o)] \quad .$$

The case for $1 < \lambda_1(J) < \lambda_2(J) <\ldots< \lambda_n(J)$ follows similarly.

However, when the eigenvalues of J are complex with $\sigma_i < 1$ and $\sigma_j > 1$ for some i and j, where $\sigma_i \equiv \mathrm{Re}\{\lambda_i(J)\}$, then $\rho[B(w)] > 1$ for all w. This follows from the fact that

$$|\sigma_K w + 1 - w| \leq |\lambda_K(J)w + 1 - w| \quad \text{for all} \quad K$$

and that if

$$w = 0 \quad \text{then} \quad \rho[B(w)] = 1$$

$$w < 0 \quad \text{then} \quad w(\sigma_i - 1) + 1 > 1$$

$$w > 0 \quad \text{then} \quad w(\sigma_j - 1) + 1 > 1 \quad .$$

In general, when the eigenvalues of J are complex

$$|w| < 2/z$$

where

$$z = \max_{i,j}(|\lambda_i(J) - \lambda_j(J)|)$$

for $\rho[B(w)]$ to be less than one. This is necessary since all of the eigenvalues of $B(w)$ must lie within the unit disk.

For the successive overrelaxation method it is absolutely necessary that $0 < w < 2$. This result is Kahan's theorem [see Ref. 13]. It is quite simple to show since

$$\begin{aligned} \det B(w) &= \det\{(I + D^{-1}Lw)^{-1}[(1 - w)I - wD^{-1}U]\} \\ &= \det(I + D^{-1}Lw)^{-1} \det[(1 - w)I - wD^{-1}U] \\ &= \det[(1 - w)I - wD^{-1}U] = (1 - w)^n \end{aligned}$$

because $D^{-1}L$ is strictly lower triangular and $D^{-1}U$ strictly upper triangular. From this, we can obtain

$$|(1 - w)|^n = \prod_{i=1}^{n} |\lambda_i(B)| \leq |\lambda_n(B)|^n \; ; \quad B = B(w)$$

where

$$|\lambda_n(B)| = \max_i |\lambda_i(B)|$$

and hence

$$|1 - w| \leq \rho(B(w)) < 1$$

implies $0 < w < 2$.

These relationships give some insight into the nature of relaxation methods and a limited region in which to search for an optimal w.

2.5 HIERARCHIAL AND DECENTRALIZED FEATURES OF ITERATIVE DECOMPOSITION METHODS

To illustrate the relation between hierarchial coordination and iterative methods for $Ax = b$ consider the partitioned problem

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

and the associated block Jacobi iteration

$$\begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_1^{K+1} \\ x_2^{K+1} \end{bmatrix} = \begin{bmatrix} 0 & -A_{12} \\ -A_{21} & 0 \end{bmatrix} \begin{bmatrix} x_1^{K} \\ x_1^{K} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} .$$

This can be written in the form of subproblems with coordination depicted by Figure 2.5.1 which reveals the two level hierarchy of coordinator and subproblems and the decentralization of information about system parameters.

Note that neither the coordinator nor the subproblems "know" all of the A matrix and that new information from the coordinator for the subproblems may be sent to both subproblems simultaneously allowing the computations to be done in parallel.

Now consider the Gauss Seidel iteration

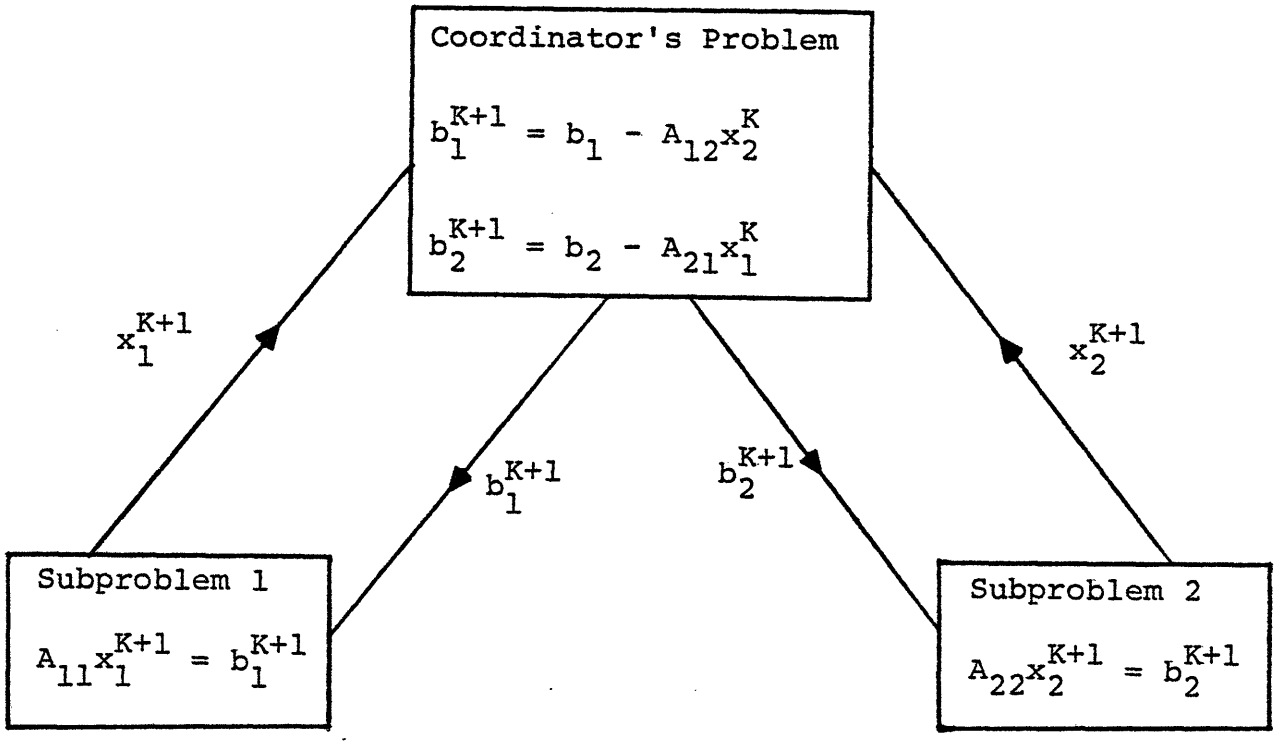

Coordinator's Problem
$b_1^{K+1} = b_1 - A_{12}x_2^K$
$b_2^{K+1} = b_2 - A_{21}x_1^K$
x_1^{K+1}
x_2^{K+1}
b_1^{K+1}
b_2^{K+1}
Subproblem 1
$A_{11}x_1^{K+1} = b_1^{K+1}$
Subproblem 2
$A_{22}x_2^{K+1} = b_2^{K+1}$

Figure 2.5.1

$$\begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_1^{K+1} \\ x_2^{K+1} \end{bmatrix} = \begin{bmatrix} 0 & -A_{12} \\ -A_{21} & 0 \end{bmatrix} \begin{bmatrix} x_1^{K+1} \\ x_2^{K} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} .$$

If this iteration is decomposed in fashion similar to the previous case, the same terminology may be adopted. Here, the information about the system (the A matrix) required by each of the subproblems and the coordinator remains the same but now the coordinator can no longer process new information from the subproblems in parallel, but must process the information sequentially. However, now the coordinator requires only half the storage space for x^K it required previously. Also, the Gauss Seidel iteration generally converges faster than the Jacobi because it utilizes the new information calculated in a more efficient manner. In terms of decentralized information about system parameters, both methods are the same. Their difference arises from the coordination scheme used, which changes the decomposition.

When both of these algorithms are relaxed, the coordinator requires more information about the system and the information about system parameters required by the subproblems remains unchanged. This can be seen by writting out the methods as before. For the simultaneous overrelaxation method, the iteration is

$$\begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_1^{K+1} \\ x_2^{K+1} \end{bmatrix} = w \begin{bmatrix} 0 & -A_{12} \\ -A_{21} & 0 \end{bmatrix} \begin{bmatrix} x_1^K \\ x_2^K \end{bmatrix} + w \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$+ (1 - w) \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_1^K \\ x_2^K \end{bmatrix} .$$

Decomposing this

$$A_{11}x_1^{K+1} = b_1^{K+1} \quad \ldots \quad \text{subproblem 1}$$

$$A_{22}x_2^{K+1} = b_2^{K+1} \quad \ldots \quad \text{subproblem 2}$$

the subproblems have exactly the same form and hence, require the same information about system parameters. The coordinator's problem becomes

$$b_1^{K+1} = - w(A_{11}x_1^K + A_{12}x_2^K) + A_{11}x_1^K + wb_1$$

$$b_2^{K+1} = - w(A_{21}x_1^K + A_{22}x_2^K) + A_{22}x_2^K + wb_2$$

and it is necessary for the coordinator to "know" all of A. Even so, the coordinator may still process information from the subproblems in parallel.

Sequential processing is however, required for the successive overrelaxation method which can be written as

$$\begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_1^{K+1} \\ x_2^{K+1} \end{bmatrix} = w \begin{bmatrix} 0 & -A_{12} \\ -A_{21} & 0 \end{bmatrix} \begin{bmatrix} x_1^{K+1} \\ \cdot \\ x_2^{K} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} w$$

$$+ (1 - w) \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} x_1^{K} \\ x_2^{K} \end{bmatrix} .$$

The subproblems and coordination problem are then defined as

$$A_{11}x_1^{K+1} = b_1^{K+1} \quad \ldots \text{ subproblem 1}$$

$$A_{22}x_2^{K+1} = b_2^{K+1} \quad \ldots \text{ subproblem 2}$$

$$b_1^{K+1} = - w[A_{11}x_1^K + A_{12}x_2^K - b_1] + A_{11}x_1^K$$

$$b_2^{K+1} = - w[A_{21}x_1^{K+1} + A_{22}x_2^K - b_2] + A_{22}x_2^K$$

coordinator's problem .

In this case, the coordinator again must "know" all of A which is bad from the decentralization of information viewpoint, but not surprisingly the storage required for x^K is cut in half. As mentioned before, it is also true that the

processing of information transferred must be done sequentially.

These types of decompositions and iterative techniques can be applied to the Lyapunov equation without writing the Lyapunov equation as a linear vector equation. This makes these iteration feasible because the dimension of the problem has not been increased. However, now the subproblems are Lyapunov type equations instead of vector equations.

2.6 CONDITIONING, STABILITY AND DECOMPOSITION METHODS

Conditioning and Stability

Computing the solution to mathematically exact analytical expression using high precision in the calculations may still result in a very inaccurate solution. This happens, barring outright blunders, for one or both of the following reasons: (1) the problem is illconditioned or (2) the algorithm used to compute the solution is unstable. A problem is defined to be illconditioned if a "small" perturbation in the data may result in a "large" change in the solution. Conversely, a problem is said to be well conditioned if a "small" perturbation in the data necessarily results in a "small" change in the solution.

It is important to note that conditioning of a problem is independent of the procedure used to obtain the solution. Stability or instability is a property of the algorithm used to compute the approximate solution. An algorithm is defined to be stable if the approximate solution it produces is the exact solution of the problem with slightly perturbed data. A well conditioned problem solved with a stable algorithm produces an approximate solution that is "near" the exact solution. In any other case we cannot make this guarantee [15].

In order to measure the conditioning of the problem of solving systems of linear equations, the notion of a condition number [15,16] is introduced. Consider the linear system of equations

$$Ax = b$$

where x is the unknown and A and b are the data. If perturbations in b are designated as δb and the corresponding changes in x as δx then

$$A(x + \delta x) = b + \delta b$$

and

$$A\delta x = \delta b \quad .$$

Using the relationships of vector and matrix norms, the following inequalities must hold

$$\frac{||b||}{||A||} \leq ||x|| \leq ||A^{-1}||\ ||b||$$

$$\frac{||\delta b||}{||A||} \leq ||\delta x|| \leq ||A^{-1}||\ ||\delta b||$$

which in turn imply

$$c_A^{-1}\ \frac{||\delta b||}{||b||} \leq \frac{||\delta x||}{||x||} \leq c_A\ \frac{||\delta b||}{||b||}$$

where $c_A = ||A||\ ||A^{-1}||$ the condition number of A with respect to inversion. If c_A is very large, we see that the norm of the relative error in the solution $||\delta x||/||x||$ may be very large or very small for a small relative error in the

norm of b. For many $b, \delta b$ pairs the upper bound may be very pessimistic, but there is always some pairs for which it is achieveable. However, for small c_A the norm of relative error in x will be small if the norm of the relative error in b is small. In this case, the equations are said to be well conditioned or A is well conditioned with respect to inversion. Note that $c_A \geq 1$ since $||AA^{-1}|| \leq ||A||\ ||A^{-1}||$. In this discussion the words large and small are not given a precise definition. This is because such a definition is not possible because it is problem dependent.

So far, we have only considered perturbations in b, but perturbations in A may be also analyzed. It should be pointed out though, that for any perturbation in A there exists an equivalent perturbation in b so that previous bounds still apply in this case also.

Suppose that now both A and b are perturbed then

$$(A + E)(x + \delta x) = b + \delta b$$

and

$$(A + E)\delta x = \delta b - Ex \quad .$$

The next difficulty encountered is the possibility that $(A + E)^{-1}$ may not exist. However, making the reasonable assumption

$$||A^{-1}E|| < ||A^{-1}||\ ||E|| < 1$$

the invertibility of $(A + E)$ is guaranteed.

To see this write $(A + E)^{-1}$ in the following form

$$(A + E)^{-1} = (I + A^{-1}E)^{-1}A^{-1}$$

so that $(A + E)^{-1}$ exists if and only if $(I + A^{-1}E)^{-1}$ exists. Letting $S = A^{-1}E$ we have $||S|| < 1$ hence $(I + S)$ is invertible and

$$(I + S)^{-1}(I + S) = I$$

or

$$(I + S)^{-1} = I - S(I + S)^{-1}$$

which leads to

$$||(I + S)^{-1}|| \leq ||I|| + ||S|| \; ||(I + S)^{-1}||$$

or

$$||(I + S)^{-1}|| \leq \frac{1}{1 - ||S||} \quad .$$

This allows us to bound $(A + E)^{-1}$ as

$$||(A + E)^{-1}|| \leq ||A^{-1}|| \; ||(I + A^{-1}E)^{-1}|| \leq \frac{||A^{-1}||}{1 - ||A^{-1}|| \; ||E||}$$

and hence, returning to the perturbations in x, we have

$$||\delta x|| \leq \frac{||A^{-1}||}{1 - ||A^{-1}|| \; ||E||} \; [||\delta b|| + ||E|| \; ||x||]$$

$$\frac{||\delta x||}{||x||} \leq \frac{||A^{-1}||}{1 - ||A^{-1}||\,||E||}\left[\frac{||\delta b||}{||b||}\;\frac{||b||}{||x||} + \frac{||E||}{||A||}\;||A||\right]$$

Using the inequality $||A||\,||x|| \geq ||b||$, we may obtain

$$\frac{||\delta x||}{||x||} \leq \frac{||A^{-1}||\,||A||}{1 - ||A^{-1}||\,||A||\,\frac{||E||}{||A||}}\left[\frac{||\delta b||}{||b||} + \frac{||E||}{||A||}\right]$$

or

$$\frac{||\delta x||}{||x||} \leq \frac{c_A}{1 - c_A\,\frac{||E||}{||A||}}\left[\frac{||\delta b||}{||b||} + \frac{||E||}{||A||}\right] .$$

Again, it is immediately apparent that c_A is the determining factor in bounding the relative error in x. Unless $c_A||E||/||A|| << 1$ the bound will be very large and the problem illconditioned. Also note that if $E = 0$ we return to original bound for perturbations in b only. If $\delta b = 0$ then

$$\frac{||\delta x||}{||x||} \leq \frac{c_A\,\frac{||E||}{||A||}}{1 - c_A\,\frac{||E||}{||A||}}$$

and the problem is only well conditioned if c_A is small.

One of the main reasons for obtaining these type of bounds is related to the philosophy of backward error analysis, as opposed to forward error analysis. Forward error analysis regards the computation in question as being described by a number of mathematical equations and hence, a

number arithmetic operations. Each arithmetic operation is associated with a maximum rounding error. Then, by considering each operation in turn, the maximum error is determined for the computation. Backward analysis is concerned with representing each arithmetic operation on the data as if it were an exact computation on slightly perturbed data and giving a bound on the effective perturbations in the data that the total computations of an algorithm make. Therefore backward analysis can tell us if an algorithm is stable, which is really what we are concerned with developing. After the effective perturbations have been determined, the perturbational bounds just derived may be used to bound the relative error.

Conditioning for Decomposition Methods

Since the conditioning of a problem is independent of the method used to solve it, decomposition methods cannot improve an illconditioned problem. However, by utilizing more information about the system, it may be possible to obtain a better bound than one specified with minimal information in the condition number. For example, consider the linear problem

$$Ax = \begin{bmatrix} 1 & 0 \\ 0 & 1000 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1000 \end{bmatrix} = b$$

where

$$||A||_\infty ||A^{-1}||_\infty = 1000 \quad .$$

For the perturbation

$$\delta b = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix} \quad ; \quad \frac{||\delta b||_\infty}{||b||_\infty} = 10^{-3}$$

but even for small δb perturbations

$$x = \begin{bmatrix} 1 \\ \\ 1 \end{bmatrix} \quad ; \quad \delta x = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix} \quad ; \quad \frac{||\delta x||_\infty}{||x||_\infty} = 1$$

so that

$$\frac{||\delta x||_\infty}{||x||_\infty} = ||A||_\infty ||A^{-1}||_\infty \frac{||\delta b||_\infty}{||b||_\infty} \quad .$$

However, suppose that in solving this problem as two subproblems

$$a_{11} x_1 = 1 \; x_1 = 1 = b_1$$

and

$$a_{22} x_2 = 1000 \; x_2 = 1000 = b_2$$

we could guarantee that

$$\frac{||\delta b_1||_\infty}{||b_1||_\infty} \leq 10^{-3} \quad \text{and} \quad \frac{||\delta b_2||_\infty}{||b_2||_\infty} \leq 10^{-3}$$

then since the condition numbers of a_{11} and a_{22} are unity

we could bound the error in x as

$$\frac{||\delta x||_\infty}{||x||_\infty} \leq \frac{||\delta x_1||_\infty}{||x_1||_\infty} + \frac{||\delta x_2||_\infty}{||x_2||_\infty} \leq c_{a_{11}} \frac{||\delta b_1||_\infty}{||b_1||_\infty}$$

$$+ c_{a_{22}} \frac{||\delta b_2||_\infty}{||b_2||_\infty} \leq 2 \cdot 10^{-3} \quad .$$

Thus, the additional information obtainable from the perturbational bounds on the subproblems and their condition numbers can result in a better bound by knowing which types of perturbations have been ruled out.

For illconditioned problems, decomposition methods will exhibit the illconditioning in several different ways, though they cannot remove the illconditioning. In fact, by decomposing the problem it is possible to have illconditioning in the subproblems whereas the original problem may be well conditioned. To illustrate these ideas, suppose that we are trying to solve $Ax = b$ where

$$A = \begin{bmatrix} 1+\varepsilon & 2 \\ 2 & 4 \end{bmatrix} ; \quad b = \begin{bmatrix} 1 \\ 1 \end{bmatrix} ; \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad .$$

Clearly, as $\varepsilon \to 0$ A becomes singular and there is no unique solution to the problem. The condition number of A is also very large. If the Gauss Seidel iteration is formed, each of the subproblems is well conditioned and can be

written

$$x_1^{K+1} = -\frac{2}{1+\varepsilon}\, x_2^K + \frac{1}{1+\varepsilon}$$

$$x_2^{K+1} = -\frac{1}{2}\, x_1^{K+1} + \frac{1}{4} \quad .$$

Substituting the first equation into the second, we have

$$x_2^{K+1} = \frac{1}{1+\varepsilon}\, x_2^K - \frac{1-\varepsilon}{4(1+\varepsilon)} \quad .$$

There are several observations to make. First, if $\varepsilon = 0$ (corresponding to a singular A matrix) then $\rho(A_o^{-1}A_1) = 1$ and the illconditioning of the original problem is "hidden" in the rate of convergence (divergence) of the coordination scheme and not in the conditioning of the subproblems. Next, if $\varepsilon = -1$, the subproblem for x_1^{K+1} is illconditioned (singular in fact) even though the A is invertible. For $-2 \leq \varepsilon < 0$, the method does not converge since $\rho(A_o^{-1}A_1) \geq 1$. We say in this case that the coordination is not successful (it is successful if $\rho(A_o^{-1}A_1) < 1$) since each of the sequences of solutions to the subproblems cannot be combined to obtain the centralized solution. However, the reason the method diverges for this case is the fact that the coupling between the subproblems has become too strong, relative to the subproblems which no longer govern the nature of the solution. This is due more to the inapplicability of the method than to illconditioning of the problem.

Alternatively, it is possible for the illconditioning of the original problem to be reflected in the conditioning of a subproblem rather than in the spectral radius of the iteration. In Chapter III this particular case occurs.

Finally, it is possible for all the subproblems to be well conditioned and the spectral radius to be less than one but the overall problem still illconditioned. This happens because illconditioning of the coordination scheme can occur in more than one way. The spectral radius may be less than one, but small perturbations in the coordinator's solution's that are returned to the subproblems may radically alter the subproblem's solutions. For the iteration

$$A_o\, x_{K+1} = -\, A_1 x_K + b = b_{K+1}$$

this says that small perturbations in b_{K+1} (the coordinator's solution) lead to large perturbations in x_{K+1} (the joint solution of the subproblems). If the sequence of $\{b_K\}$ cannot be computed exactly, there is a possibility that the relative error in x may be large. An example of this is the Jacobi iteration for $Ax = b$ where A is $n \times n$ and

$$A = \begin{bmatrix} 1 & -1 & -1 & \cdots & -1 \\ & 1 & -1 & & \vdots \\ & & \ddots & \ddots & \\ & & & \ddots & -1 \\ & & & & 1 \end{bmatrix} ;$$

$$A^{-1} = \begin{bmatrix} 1 & 1 & 2 & 4 & \cdots & 2^{n-2} \\ & 1 & 1 & 2 & & \\ & & \ddots & \ddots & & \\ & & & \ddots & 2 & 4 \\ & & & & 1 & 2 \\ & & & & 1 & 1 \\ & & & & & 1 \end{bmatrix}$$

with

$$||A||_\infty = n \quad \text{and} \quad ||A^{-1}||_\infty = 2^{n-1} \quad .$$

For the splitting $A_o = I$, $A_1 = A - I = U$ we have that $\rho(A_o^{-1}A_1) = 0$ and $||A_o^{-1}||_\infty ||A_o||_\infty = 1$ so that the iteration converges in n steps and each of the subproblems is ideally well conditioned. The Jacobi iteration is written as

$$x_{K+1} = (I - A)x_K + b = b_{K+1} \ , \quad x_o = 0$$

which certainly verifies that the subproblems can be solved without error given b_{K+1} from the coordinator. The only possible problem that could occur is that the coordinator might not accurately calculate b_K. This can be seen from the particular case where

$$b^T = [n, n-3, n-4, \ldots, 1, 0, -1]$$

$$\delta b^T = [1 \quad 1 \quad 1 \quad \ldots \quad 1]2^{-n+1}$$

$$x^T = [1 \quad -1 \quad -1 \quad \ldots \quad -1]$$

$$\delta x^T = [1 \quad 1/2 \quad 1/4 \quad \ldots \quad 2^{-n+1}$$

$$\frac{||\delta x||_\infty}{||x||_\infty} = 1 \quad \text{and} \quad \frac{||\delta b||_\infty}{||b||_\infty} = \frac{2^{-n+1}}{n} \quad .$$

If $n = 10$ this means that a uniform relative perturbation of b on the order of $.02\%$ results in a 100% relative error in the norm of x. Hence, if b_K is not calculated exactly large errors may result.

This last example shows that in order to obtain perturbational bounds for iterative methods that not only must the condition numbers of the subproblems and the contraction condition be accounted for, but also the rounding errors in the computation of the information supplied by the coordinator to the subproblems. If these errors in the coordination equations can be bounded or particular error directions be

eliminated it may then be possible to obtain accurate solutions even though the original problem is illconditioned.

Condition Number of K_A

Viewing the Lyapunov equation as a set of linear equations in n^2 variables, the conditioning of the problem is determined by the condition number of K_A. Athay [25] has examined this problem in an attempt to formulate the condition number of K_A in terms of quantities involving A. However, there seems to be no easy way to compute or approximate the condition number of K_A. The following inequality illustrates the nature of the problem.

Using the spectral norm,

$$C_{K_A} = ||K_A||_2||K_A^{-1}||_2 \geq \frac{\max|\lambda_i(K_A)|}{\min|\lambda_i(K_A)|} = \frac{\max|\lambda_i(A) + \lambda_j(A)|}{\min|\lambda_i(A) + \lambda_j(A)|}$$

where $\lambda_i(K_A)$ and $\lambda_i(A)$ are the eigenvalues of K_A and A respectively. Now suppose A is a 2 x 2 symmetric matrix with eigenvalues $\alpha \pm jw$ and $\alpha << w$, corresponding to a lightly damped system, which is not uncommon. It follows that

$$C_A = ||A||_2||A^{-1}||_2 = \frac{\max|\lambda_i(A)|}{\min|\lambda_i(A)|} = 1$$

while

$$C_{K_A} = ||K_A||_2||K_A^{-1}||_2 \geq \frac{2|\alpha + jw|}{\alpha} \simeq \frac{2w}{\alpha} >> 1$$

and it is clear that the condition number of A is not related to the condition number of K_A. The basic difficulty of approximating C_{K_A} is that of finding an upper bound on $||K_A^{-1}||$ that is easily computable from quantities involving A.

CHAPTER III

TWO TIME SCALE ITERATIVE DECOMPOSITION

3.1 Introduction

It is common in engineering practice to obtain approximate solutions to difficult problems by truncation of a series representation of the solution. These series approximations are often easier to obtain and therefore attractive. Perturbational techniques can be applied to characterize the nature of the solution's dependence on some relatively small parameter. In the case of singular perturbations, the formal expansion of the solution in a power series may produce a non-convergent series. This series may nevertheless, be useful as an asymptotic series expansion of the solution which may yield a sufficiently accurate approximation [19].

In control theory, the use of singular perturbation techniques resulting in asymptotic series expansion has been motivated by the desire to reduce the numerical computation associated with large scale problems for which analytical treatment is not feasible [23]. It has also been noted recently, that application of singular perturbation techniques to linear stochastic optimization and estimation problems result in hierarchically structured approximations to optimal controllers and filters [20,21,22]. Basically, the upper levels of the hierarchey develop their control strategers based on a reduced model of the system that predicts the long

term behavior while the lower levels of the hierarchy utilize a boundary layer model to predict the short term behavior. This can be extended to the multiple time scale case.

To obtain the reduced order models of the corresponding fast and slow modes of the system, the system is decomposed according to its dependence on a small parameter. By decomposing the original problem into decoupled subproblems and coordinating these solutions iteratively to a global solution, a hierarchical structure is imposed.

However, a decomposition of the global problem raises the questions of coordinability (convergence), rate of convergence, and convergence to the correct solution.

3.2 Decomposition Algorithm for the Lyapunov Equation

For the stable linear system

$$\dot{x} = Ax \ , \quad x(0) \quad \text{given}$$

we write the Lyapunov equation

$$A^T K + KA + Q = 0$$

where Q is positive semidefinite and K is the unknown. Assume x is partitioned into n_1 and n_2 dimensional vectors and that A,K and Q are partitioned conformably as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}, \quad K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} .$$

Consider the following algorithm, motivated by singular perturbation theory whose connection will be given later

$$\begin{bmatrix} A_{11}^T & A_{21}^T \\ A_{12}^T & A_{22}^T \end{bmatrix} \begin{bmatrix} K_{11}^m & K_{12}^{m-1} \\ K_{21}^m & K_{22}^m \end{bmatrix} + \begin{bmatrix} K_{11}^m & K_{12}^m \\ K_{21}^{m-1} & K_{22}^m \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = 0$$

Here K^m denotes the $m^{\underline{th}}$ iterate and the initial guess $K^o = 0$. To more clearly formulate the algorithm, the following subproblems are defined.

Fast Subproblem

$$A_{22}^T K_{22}^m + K_{22}^m A_{22} + Q_{22}^m = 0 \quad .$$

Slow Subproblem

$$\hat{A}_{11}^T K_{11}^m + K_{11}^m \hat{A}_{11} + Q_{11}^m = 0 \; ; \quad \hat{A}_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21} \; .$$

Coordinator's Problem

$$K_{12}^m A_{22} + K_{11}^m A_{12} + A_{21}^T K_{22}^m + Q_{12}^m = 0$$

$$Q_{22}^m = A_{12}^T K_{12}^{m-1} + K_{12}^{m-1} A_{12} + Q_{22}$$

$$Q_{12}^m = Q_{12} + A_{11}^T K_{12}^{m-1}$$

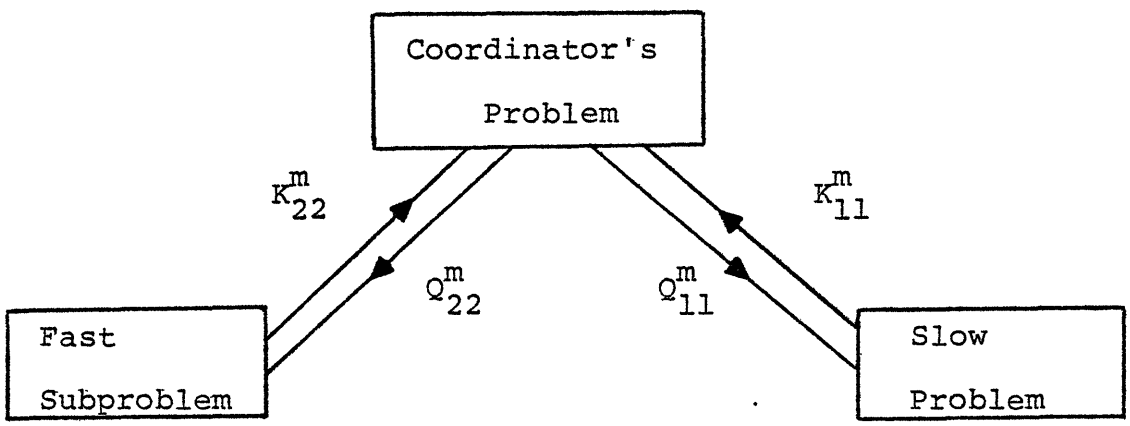
$$Q_{11}^m = Q_{11} + (A_{22}^{-1}A_{21})^T Q_{22}^m (A_{22}^{-1}A_{21}) - Q_{12}^m (A_{22}^{-1}A_{21}) - (A_{22}^{-1}A_{21})^T Q_{21}^m$$

or

$$Q_{11}^m = \hat{Q}_{11} - \hat{A}_{11}^T K_{12}^{m-1} (A_{22}^{-1}A_{21}) - (A_{22}^{-1}A_{21})^T K_{21}^{m-1} \hat{A}_{11}$$

$$\hat{Q}_{11} = Q_{11} - (A_{22}^{-1}A_{21})^T Q_{21} - Q_{12}(A_{22}^{-1}A_{21}) + (A_{22}^{-1}A_{21})^T Q_{22}(A_{22}^{-1}A_{21}).$$

This can be represented diagramatically (Figure 3.2.1) as



Two-Time Scale Decomposition Algorithm for the Lyapunov Equation

Figure 3.2.1

with the initialization of the algorithm beginning with $K_{12}^{o} = 0$ to obtain $Q_{22}^{1} = Q_{22}$ and $Q_{11}^{1} = \hat{Q}_{11}$. Next, both of the subproblems are solved and K_{11}^{1} and K_{22}^{1} are sent to the coordinator for its calculation of K_{12}^{1}, Q_{11}^{2} and Q_{22}^{2} and so forth. Note that the fast and slow subproblems are decoupled and that both the reduced order system

$$\dot{x}_1 = \hat{A}_{11}x_1$$

and the boundary layer system

$$\dot{x}_2 = A_{22}x_2$$

need be stable for the algorithm to be well defined at each step. However, even if the subproblems are well defined, the algorithm need not converge to the global solution. We will say that the coordination is <u>successful</u> with this decomposition method if and only if

$$\lim_{m\to\infty} K^{m} = K \quad .$$

<u>Theorem 3.2.1</u>

Define the linear operators

$L_1 : R^{nxn} \to R^{nxn}$ and $L_2 : R^{nxn} \to R^{nxn}$ by

the following equations

$$L_1(K) = \begin{bmatrix} A_{11}^T K_{11} + K_{11} A_{11} + A_{21}^T K_{21} + K_{12} A_{21} & A_{21}^T K_{22} + K_{11} A_{12} + K_{12} A_{22} \\ K_{22} A_{21} + A_{21}^T K_{11} + A_{22}^T K_{21} & A_{22}^T K_{22} + K_{22} A_{22} \end{bmatrix}$$

$$L_2(K) = \begin{bmatrix} 0 & A_{11}^T K_{12} \\ K_{21} A_{11} & A_{12}^T K_{12} + K_{21} A_{12} \end{bmatrix} \quad .$$

Assume that $\hat{A}_{11}$ and A_{22} are stable so that

$$\mathrm{Re}[\lambda_i(\hat{A}_{11})] < 0 \ , \quad \mathrm{Re}[\lambda_j(A_{22})] < 0 \quad \forall i,j \quad .$$

The Lyapunov equation is then successfully coordinated by the two time scale decomposition method if and only if

$$\rho(L_1^{-1} \cdot L_2) < 1 \quad .$$

Proof:

If $L_1(K) = 0$ then

$$A_{22}^T K_{22} + K_{22} A_{22} = 0$$

and hence $K_{22} = 0$ since A_{22} is stable.

Next

$$A_{21}^T K_{22} + K_{11} A_{12} + K_{12} A_{22} = 0$$

or

$$K_{12} = - K_{11} A_{12} A_{22}^{-1} \quad .$$

which implies that

$$A_{11}^T K_{11} + K_{11} A_{11} + A_{21}^T K_{21} + K_{12} A_{21} = \hat{A}_{11}^T K_{11} + K_{11} \hat{A}_{11} = 0$$

and thus, it is necessary that $K_{11} = 0$ and therefore, $K_{12} = 0$ because $\hat{A}_{11}$ is stable. Hence, if $L_1(K) = 0$ then $K = 0$ and L_1 has no zero eigenvalues and therefore must be invertible.

Now the decomposition algorithm can be written

$$L_1(K^m) + L_2(K^{m-1}) + Q = 0$$

or

$$K^m = - L_1^{-1} \cdot L_2(K^{m-1}) - L_1^{-1}(Q) \quad .$$

It is necessary and sufficient that $\rho(L_1^{-1} \cdot L_2) < 1$ for the iterates $\{K^m\}$ of this last equation to converge to K as m tends to infinity for any initial K^o and any positive semi-definite Q. Q.E.D.

The asymptotic rate of convergence is also given by $\rho(L_1^{-1} \cdot L_2)$. This can be seen if the error at each step is E^m so that

$$E^m = K^m - K$$

then

$$E^m = - L_1^{-1} \cdot L_2(E^{m-1}) = (-L_1^{-1} \cdot L_2)^m E^o \quad .$$

Eventually, the eigenvalue of $L_1^{-1} \cdot L_2$ of largest modulus equal to $\rho(L_1^{-1} \cdot L_2)$ becomes dominant and the "eigenvector" associated with this eigenvalue becomes the dominant

direction. Note that $L_1^{-1} \cdot L_2$ has in general $n_1^2 + n_2^2$ zero eigenvalues and $2n_1n_2$ nonzero eigenvalues. In the limit

$$||E^m|| = \rho(L_1^{-1} \cdot L_2)||E^{m-1}||$$

for any matrix norm but as noted before, it may take many iterations before this asymptotic convergence rate is achieved. The calculation of $\rho(L_1^{-1} \cdot L_2)$ is not feasible as a convergence test since it involves an $n(n + 1)/2$ order eigenvalue problem. It may be reduced to an n order eigenvalue problem after some algebraic manipulation, but in any event, this would involve as much effort as solving the original problem. However, a few examples will provide insight as to when the algorithm may be expected to converge.

Scalar Example

Writing out the decomposition algorithm,

$$K_{11}^m = \frac{-q_{11}^m}{2\hat{a}_{11}}$$

$$K_{22}^m = \frac{-q_{22}^m}{2a_{22}}$$

$$K_{12}^m = \left(\frac{-1}{a_{22}}\right)[a_{12}K_{11}^m + a_{21}K_{22}^m + a_{11}K_{12}^{m-1} + q_{12}]$$

$$q_{11}^m = \frac{-2\hat{a}_{11}a_{21}}{a_{22}} K_{12}^{m-1} + \hat{q}_{11}$$

$$q_{22}^m = 2a_{12}K_{12}^{m-1} + q_{22}$$

and by substitution and algebraic manipulation

$$K_{11}^m = \frac{a_{21}}{a_{22}} K_{12}^{m-1} + c_1 \quad ; \quad c_1 \text{ a constant}$$

$$K_{22}^m = -\frac{a_{12}}{a_{22}} K_{12}^{m-1} + c_2 \quad ; \quad c_2 \text{ a constant}$$

$$K_{12}^m = -\frac{a_1}{a_2} K_{12}^{m-1} + c_3 \quad ; \quad c_3 \text{ a constant} \quad .$$

It is now clear that the equation for K_{12}^m governs the convergence of the whole iteration and hence

$$\rho(L_1^{-1} \cdot L_2) = \left|\frac{a_{11}}{a_{22}}\right| \quad .$$

Decoupled Example

For $A_{12} = 0 \quad ; \quad A_{21} = 0$

$$A_{11}^T K_{11}^m + K_{11}^m A_{11} + Q_{11} = 0$$

$$A_{22}^T K_{22}^m + K_{22}^m A_{22} + Q_{22} = 0$$

$$K_{12}^m = -A_{11}^T K_{12}^{m-1} A_{22}^{-1} - Q_{12}$$

so that K_{11}^m and K_{22}^m are constant and only K_{12}^m varies from step to step. The asymptotic convergence rate of $\{K_{12}^m\}$

is given by

$$\rho(L_1^{-1} \cdot L_2) = \rho(-A_{11}^T \otimes A_{22}^{-T}) = \max_{i,j} \left| \frac{\lambda_i(A_{11})}{\lambda_j(A_{22})} \right| .$$

Hence

$$\rho(L_1^{-1} \cdot L_2) < 1 \quad \text{if and only if}$$

$$\max_i |\lambda_i(A_{11})| < \min_j |\lambda_j(A_{22})|$$

which means that subsystem A_{11} represents a slow system relative to subsystem A_{22} which represents a fast subsystem. It is here that the two time scale nature of this algorithm becomes apparent. Note, in fact, that only one of A_{12} or A_{21} need be zero for this separation of time scales condition to apply.

In order to obtain a similar condition for convergence in the general case when A is not block triangular, it is first necessary to discuss the subproblem conditioning.

Subproblem Conditioning

The conditioning of the subproblems is considered as side issues since the purpose of this section is to derive a bound for $||K_{ii}||$ in terms of $||Q_{ii}||$ and $||A_{ii}||$. The conditioning of the subproblems arises in connection with bounding

$$||K_{A_{ii}}^{-1}|| = ||(A_{ii}^T \otimes I + I \otimes A_{ii}^T)^{-1}||$$

and since $C_{K_{A_{ii}}} = ||K_{A_{ii}}||\ ||K_{A_{ii}}^{-1}||$ the condition number of $K_{A_{ii}}$ for each subproblem is directly related to this bound. This bound is then used to obtain a sufficient condition in terms of matrix norms for the convergence of the algorithm.

Each of the subproblems is in the form of a Lyapunov equation, so we will not bother with subscripts. To avoid confusing notation let P be the unknown of the Lyapunov equation

$$A^T P + PA + Q = 0 \quad , \quad A \in R^{nxn} \quad .$$

Using the Kronecker product notation, this can be written as

$$K_A p = - q$$

where $K_A = A^T \otimes I + I \otimes A^T$ and $p = p_R$ and $p = q_R$ discussed in Chapter II, Section 2.2. Now using some basic matrix norm relationships, we may write

$$||p||_2 \leq ||K_A^{-1}||_2 ||q||_2$$

or reforming the matrices P and Q

$$||P||_E \leq ||K_A^{-1}||_2 ||Q||_E$$

and

$$||P||_2 \leq n^{1/2} ||K_A^{-1}||_2 ||Q||_2 \quad .$$

Define μ_i to be the condition number of K_A with respect to

various matrix norms so that

$$\mu_i = ||K_A||_i ||K_A^{-1}||_i ; \quad i = 1,2,\infty,E$$

where E denotes the Euclidean norm. Manipulating matrix norms we find

$$||K_A^{-1}||_2^2 \leq ||K_A^{-1}||_1 ||K_A^{-1}||_\infty$$

or

$$||K_A^{-1}||_2^2 \leq \frac{\mu_1 \mu_\infty}{||K_A||_1 ||K_A||_\infty}$$

and since $||K_A||_{1,\infty} = 2||A||_{\infty,1}$ we may write

$$||K_A^{-1}||_2 \leq \frac{(\mu_1 \mu_\infty)^{1/2}}{2(||A||_1 ||A||_\infty)^{1/2}} \leq \frac{(\mu_1 \mu_\infty)^{1/2}}{2||A||_2} .$$

If the Euclidean norm is used then

$$||K_A^{-1}||_2 \leq ||K_A^{-1}||_E \leq \frac{\mu_E}{||K_A||_E}$$

and since $||K_A||_E^2 = 2n||A||_E^2 + 2(trA)^2$

$$||K_A^{-1}||_E \leq \frac{\mu_E}{\sqrt{2n}\ ||A||_E} = \frac{\mu_E}{2||A||_E}\sqrt{\frac{2}{n}} .$$

We have now obtained bounds on $||K_A^{-1}||$ in terms of its condition number and $||A||$. By substituting these bounds for $||K_A^{-1}||$ into the appropriate inequalities involving $||P||$ and $||Q||$ we obtain

$$||P||_2 \leq (n\mu_1\mu_\infty)^{1/2} \frac{||Q||_2}{2||A||_2}$$

and

$$||P||_E \leq (\frac{2}{n})^{1/2}\mu_E \frac{||Q||_E}{2||A||_E} .$$

Note also that

$$(\frac{2}{n})^{1/2}\mu_2 \leq (\frac{2}{n})^{1/2}\mu_E \leq \sqrt{2}\ (n\mu_1\mu_\infty)^{1/2}$$

which shows the maximum size of $(\frac{2}{n})^{1/2}\mu_E$ as compared to $(n\mu_1\mu_\infty)^{1/2}$.

Sufficient Convergence Condition for the General Case

Using the bounds just developed, a sufficient but not necessary convergence condition can be derived. Not being a necessary condition, it leads to a pessimistic estimate of when the algorithm will converge, but seems to be the only condition obtainable. If we define

$\alpha_1, \alpha_2, \beta_1$ and β_2 as

$$\alpha_1 = (n_1\hat{\mu}_1\hat{\mu}_\infty)^{1/2}$$

$$\alpha_2 = (n_2\mu_1\mu_\infty)^{1/2}$$

$$\beta_1 = (2/n_1)^{1/2}\hat{\mu}_E$$

$$\beta_2 = (2/n_2)^{1/2}\mu_E$$

where

$$\hat{\mu}_i = ||K_{\hat{A}_{11}}||_i ||K^{-1}_{\hat{A}_{11}}||_i \quad , \; i = 1,2,\infty,E$$

$$\mu_i = ||K_{A_{22}}||_i ||K^{-1}_{A_{22}}||_i \quad , \; i = 1,2,\infty,E$$

and

$$\hat{A}_{11} \quad \text{is} \quad n_1 \times n_1$$

$$A_{22} \quad \text{is} \quad n_2 \times n_2$$

then, from the subproblems of the iteration

$$||K^m_{11}||_2 \leq \frac{\alpha_1 ||Q^m_{11}||_2}{2||\hat{A}_{11}||_2}$$

$$||K^m_{11}||_E \leq \frac{\beta_1 ||Q^m_{11}||_E}{2||\hat{A}_{11}||_E}$$

and

$$||K^m_{22}||_2 \leq \frac{\alpha_2 ||Q^m_{22}||_2}{2||A_{22}||_2}$$

$$||K^m_{22}||_E \leq \frac{\beta_2 ||Q^m_{22}||_E}{2||A_{22}||_E} \quad .$$

The coordinator's problem for Q^m_{22} and Q^m_{11} leads to

$$||Q^m_{22}|| \leq 2||A_{12}|| \; ||K^{m-1}_{12}|| + ||Q_{22}||$$

$$||Q_{11}^m|| \leq 2||\hat{A}_{11}||\ ||K_{12}^{m-1}||\ ||A_{22}^{-1}A_{21}|| + ||\hat{Q}_{11}|| \ .$$

If we let γ_1 denote α_1 or β_1 depending on which matrix norm is used and likewise γ_2 for α_2 or β_2 then using the last two inequalities we have

$$||K_{11}^m|| \leq \gamma_1||A_{22}^{-1}A_{21}||\ ||K_{12}^{m-1}|| + \frac{\gamma_1||\hat{Q}_{11}||}{2||\hat{A}_{11}||}$$

$$||K_{22}^m|| \leq \frac{\gamma_2||A_{12}||\ ||K_{12}^{m-1}||}{||A_{22}||} + \frac{\gamma_2||Q_{22}||}{2||A_{22}||} \ .$$

Now these inequalities are used in the coordinators equation for K_{12}^m

$$||K_{12}^m|| \leq ||A_{22}^{-1}|| \left[||A_{12}||\ ||K_{11}^m|| + ||A_{21}||\ ||K_{22}^m|| \right.$$

$$\left. + ||A_{11}||\ ||K_{12}^{m-1}|| + ||Q_{12}|| \right]$$

to obtain

$$||K_{12}^m|| \leq ||A_{22}^{-1}|| \left[\gamma_1||A_{12}||\ ||A_{22}^{-1}A_{21}|| + \frac{\gamma_2||A_{21}||\ ||A_{12}||}{||A_{22}||} \right.$$

$$\left. + ||A_{11}|| \right] ||K_{12}^{m-1}|| + \left[\frac{\gamma_1||A_{12}||\ ||\hat{Q}_{11}||}{2||\hat{A}_{11}||} \right.$$

$$\left. + \frac{\gamma_2||A_{21}||\ ||Q_{22}||}{2||A_{22}||} \right] ||A_{22}^{-1}||$$

or

$$||K_{12}^m|| \leq ||A_{22}^{-1}|| \left[(\gamma_1 + \gamma_2)||A_{12}||\ ||A_{22}^{-1}||\ ||A_{21}|| + ||A_{11}||\right] ||K_{12}^{m-1}|| + c\ ,\ c = \text{constant}.$$

Therefore, if

$$||A_{22}^{-1}|| \left[(\gamma_1 + \gamma_2)||A_{12}||\ ||A_{22}^{-1}||\ ||A_{21}|| + ||A_{11}||\right] < 1$$

the algorithm is a contraction mapping and $\{K_{12}^m\}$ must converge to K_{12} and so $\{K_{11}^m\}$ and $\{K_{22}^m\}$ converge to K_{11} and K_{22} respectively and the whole algorithm converges. When A is a 2 x 2 matrix, both γ_1 and γ_2 are unity and the condition becomes

$$\frac{1}{|a_{22}|}\left[|a_{11}| + \frac{2|a_{12}|\ |a_{21}|}{|a_{22}|}\right] < 1$$

instead of

$$\frac{|a_{11}|}{|a_{22}|} < 1$$

which is the necessary and sufficient condition. This shows that the matrix norm condition may be not satisfied even for cases that would converge fairly rapidly and is therefore pessimistic for many cases.

If γ_1 and γ_2 are close to one, which implies that the subproblems are well conditioned, we might conjecture that

$$||A_{22}^{-1}||\ ||A_{11}|| < 1$$

might be a reasonable condition which under which convergence could be expected. It is interesting to note that as A_{11} approaches a singular matrix, the $\rho(L_1^{-1} \cdot L_2) < 1$ condition may not be violated. Of course, the solution of the $\hat{A}_{11}$ subproblem does not exist in this case. However, A_{22} becoming singular does affect the spectral radius so that $\rho(L_1^{-1} \cdot L_2) > 1$ and the algorithm diverges. However, when the algorithm does converge, it would be helpful to know how much computational effort is required.

Number of Operations

This section gives a comparison of the number of operations required to compute the approximate solution to the Lyapunov equation by the decomposition algorithm and the general method used to compute the subproblems.

The following assumptions are made:

(1) Q is symmetric

(2) The state is decomposed into two substates of equal dimension, i.e. $n_1 = n_2 = n/2$.

(3) $f(n)$ is the number of operations to solve the Lyapunov equation and $f(n) = \alpha n^3$ where α is a positive constant.

(4) Additions are not counted only multiplications.

Breaking down the algorithm according to subproblems and

coordination, we see that

Forming	Operations
$\hat{A}_{11}$ and $A_{22}^{-1}A_{21}$	$\frac{3n^3}{8}$
$\hat{Q}_{11}$	$\frac{3n^3}{8}$
Q_{11}^{m+1}	$\frac{n^3}{4}$
Q_{22}^{m+1}	$\frac{n^3}{8}$
Solving for	
K_{11}^{m+1}	$\frac{\alpha}{8} n^3$
K_{22}^{m+1}	$\frac{\alpha}{8} n^3$
K_{12}^{m+1} and K_{21}^{m+1}	$\frac{\alpha}{2} n^3$.

Let σ = number of iterations, then the total operation count for symmetric Q is $T = \text{t.o.c.} = \frac{3n^3}{8} + \sigma[\frac{7n^3}{8} + \alpha(\frac{1}{4} n^3)]$.

For the algorithm to effectively reduce the computations involved

$$T < \alpha n^3$$

or

$$\sigma < \frac{8\alpha - b}{2\alpha + 7} < 4 \quad .$$

This shows that neglecting sparsity if only a few iterations are required the computation can be reduced.

3.3 Numerical Results for a Simple Model Coupled Harmonic Oscillators

As a test of when $||A_{11}||\ ||A_{22}^{-1}||$ might be a reliable indicator of convergence, the spectral radius of $L_1^{-1} \cdot L_2$ was computed numerically for a variety of couplings and subsystem pole placements for the system given by

$$A = \left[\begin{array}{cc:cc} 0 & 1 & 0 & 0 \\ -w_1^2 & -2\phi_1 w_1 & \varepsilon_1 & 0 \\ \hdashline 0 & 0 & 0 & 1 \\ \varepsilon_2 & 0 & -w_2^2 & -2\phi_2 w_2 \end{array}\right] .$$

This system matrix represents coupled harmonic oscillators where w is the natural frequency and ϕ is the damping ratio. Define r as

$$r = \max_{i,j} \left| \frac{\lambda_i(A_{11})}{\lambda_j(A_{22})} \right| .$$

In this test, r is set to a value of $1/5$ and the damping ratios of the two subsystems A_{11} and A_{22} are varied.

The spectral radius, $\rho(L_1^{-1} \cdot L_2)$ was generally smaller when the two subsystems had poles with the same angle in the

complex plane.

Also, when $|\varepsilon_1\varepsilon_2| \leq w_1^2 w_2^2$, all cases gave $\rho(L_1^{-1} \cdot L_2) \leq 1$. If $|\varepsilon_1\varepsilon_2| \geq w_1^2 w_2^2$, $\hat{A}_{11}$ is singular or A is unstable. The coupling was varied as

$$\varepsilon_1\varepsilon_2 = \alpha_i w_1^2 w_2^2$$

where

$$\alpha_i = 1,\ 1/2,\ -1/2,\ -1,\ -2 \quad .$$

This was done because in the detailed analytical expression for the spectral radius (not given here) for this specific example ε_1 and ε_2 always appear together as the product $\varepsilon_1\varepsilon_2$.

Only the cases for $\alpha_i = \pm 1$ are tabulated in Table 3.3.1. For $\varepsilon_1\varepsilon_2 = -2w_1^2 w_2^2$ cases in which $\rho(L_1^{-1} \cdot L_2) > 1$ were obtained. This limits how strongly the subsystems may be coupled. The results generally indicate that r is a reliable indicator of convergence when $|\varepsilon_1\varepsilon_2| \leq w_1^2 w_2^2$ and the subsystems do not have large angles between their complex poles.

Numerical Calculation of $\rho(L_1^{-1} \cdot L_2)$

The calcualation of the spectral radius can be accomplished by finding the eigenvalues of a matrix of order n where $n = \dim(L_1)$. The algorithm written in operator form

Table 3.3.1

$\omega_1^2 = 1$ $\quad \sigma_s \pm j\omega_s$ eigenvalues of A_{11}

$\omega_2^2 = 25$

$r = \frac{1}{5}$ $\quad \sigma_F \pm j\omega_F$ eigenvalues of A_{22}

$2\phi_1\hat{\omega}_1$	$2\sigma_2\omega_2$	$\frac{\varepsilon_1\varepsilon_2}{(\omega_1\omega_2)^2}$	σ_s	ω_s	σ_F	ω_F	$\rho(L_1^{-1}, L_2)$
.5	4	1	-.25	.968	-2	4.583	.320
2	4	1	-1	0	-2	4.583	.367
1.5	1	1	-.75	.661	-.5	4.975	.425
.5	1	1	-.25	.968	-.5	4.975	.274
1	9	1	-.5	.866	-4.5	2.179	.360
1.5	9	1	-.75	.661	-4.5	2.179	.327
.5	4	-1	-.25	.968	-2	4.583	.488
2	4	-1	-1	0	-2	4.583	.289
1.5	1	-1	-.75	.661	-.5	4.975	.425
.5	1	-1	-.25	.968	-.5	4.975	.458
1.5	9	-1	-.75	.661	-4.5	2.179	.217
1	9	-1	-.5	.866	-4.5	2.179	.248
.5	10	1	-.25	.968	-5	0	.800*
.5	10	-1	-.25	.968	-5	0	.553**

* $\max \rho(L_1^{-1}, L_2)$ for $\varepsilon_1\varepsilon_2/\omega_1^2\omega_2^2 = 1$

** $\max \rho(L_1^{-1}, L_2)$ for $\varepsilon_1\varepsilon_2/\omega_1^2\omega_2^2 = -1$

$\min \rho(L_1^{-1}, L_2) = .200$ for $\varepsilon_1\varepsilon_2/\omega_1^2\omega_2^2 = \pm 1$

$$L_1(K^{m+1}) = - L_2(K^m) - Q$$

corresponds to a matrix iteration

$$M_1 k^{m+1} = - M_2 k^m - q$$

where k^m and q are vectors formed from the rows of K^m and Q respectively. Taking into account the symmetry of K and Q, we define the matrix iteration by using the Kronecker products and identities found in Section 2.2.

$$M_1 k = \begin{bmatrix} A_{11}^T \otimes I + I \otimes A_{11}^T & I \otimes A_{21}^T + (A_{21}^T \otimes I)P & 0 \\ I \otimes A_{12}^T & I \otimes A_{22}^T & A_{21}^T \otimes I \\ 0 & 0 & A_{22}^T \otimes I + A_{22}^T \end{bmatrix} \begin{bmatrix} k_{11_R} \\ k_{12_R} \\ k_{22_R} \end{bmatrix}$$

$$M_2 k = \begin{bmatrix} 0 & 0 & 0 \\ 0 & A_{11}^T \otimes I & 0 \\ 0 & A_{12}^T \otimes I + (I \otimes A_{12}^T)P & 0 \end{bmatrix} \begin{bmatrix} k_{11_R} \\ k_{12_R} \\ k_{22_R} \end{bmatrix}$$

where P is the permutation matrix associated with the order of the matrices $A_{ii} \otimes I$. In the case where A_{ii} is 2×2

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Now note that $M_1^{-1} \cdot M_2$ is much larger than $n \times n$ and that

$$\rho(M_1^{-1} \cdot M_2) = \rho(L_1^{-1} \cdot L_2)$$

since they are the spectral radii of the same iteration. If M_1 and M_2 are partitioned so that

$$M_1^{-1} = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix} = X$$

and

$$M_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & N_1 & 0 \\ 0 & N_2 & 0 \end{bmatrix}$$

then

$$M_1^{-1}M_2 = \begin{bmatrix} 0 & X_{12}N_1+X_{13}N_2 & 0 \\ 0 & X_{22}N_1+X_{23}N_2 & 0 \\ 0 & X_{32}N_1+X_{33}N_2 & 0 \end{bmatrix}$$

so that

$$\rho(M_1^{-1} \cdot M_2) = \rho(X_{22}N_1 + X_{23}N_2) \quad .$$

Define J as $J = X_{22}N_1 + X_{23}N_2$.

To determine $M_1^{-1} = X$, M_1 is partitioned as

$$M_1 = \left[\begin{array}{c|cc} M_{11} & M_{12} & 0 \\ \hline M_{21} & M_{22} & M_{23} \\ 0 & 0 & 0 \end{array}\right] = \left[\begin{array}{c|cc} A_{11}^T \otimes I + I \otimes A_{11}^T & I \otimes A_{21}^T + (A_{21}^T \otimes I)P & 0 \\ \hline I \otimes A_{12}^T & I \otimes A_{22}^T & A_{21}^T \otimes I \\ 0 & 0 & A_{22}^T \otimes I + I \otimes A_{22}^T \end{array}\right]$$

and it follows that M_{11}^{-1} , M_{22}^{-1} and M_{33}^{-1} exist for stable A_{11} and A_{22}. It can then be shown that

$$X_{22} = (M_{22} - M_{21}M_{11}^{-1}M_{12})^{-1}$$

$$X_{23} = - X_{22}M_{23}M_{33}^{-1} \quad .$$

Finally, using these last two formulas, the spectral radius of $L_1^{-1} \cdot L_2$ can be written as

$$\rho(L_1^{-1} \cdot L_2) = \rho([M_{22} - M_{21}M_{11}^{-1}M_{12}]^{-1}[N_1 - M_{23}M_{33}^{-1}N_2]) = \rho(J) \quad .$$

Note that here J is only an $n \times n$ matrix. Using this formula, the spectral radius for different pole placements was calculated.

3.4 Singular Perturbation Theory and Multiple Time Scales

Relation to Singular Perturbation Theory

Up to this point, the connection of this algorithm to the ideas of singular perturbation theory [19-24,29] has been somewhat obscure. After some preliminaries about singular perturbations, the connection will be made clear.

Consider the partitioned linear system

$$\begin{bmatrix} \dot{x}_1 \\ \varepsilon\dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad .$$

It is conventional to assume that ε is a small positive parameter, $\varepsilon << 1$, and that $||A_{11}||$ and $||A_{22}||$ are approximately the same magnitude. With this system, there are two associated systems called the degenerate system and the boundary layer system [29]. The degenerate system is formed by setting $\varepsilon = 0$ and is given by

$$\dot{x}_1 = (A_{11} - A_{12}A_{22}^{-1}A_{21})x_1$$

$$\dot{x}_2 = (-A_{22}^{-1}A_{21})x_1 \quad .$$

The degenerate system predicts the long term behavior of the system. In order to predict the short term behavior of the system, the boundary layer system is formed using the stretching transformation

$$\tau = t/\varepsilon$$

and making a change of variables in the original equations. To obtain the boundary layer system ε is set to zero in the transformed equations. This results in

$$\frac{dx_1}{d\tau} = 0$$

$$\frac{dx_2}{d\tau} = A_{21}x_1 + A_{22}x_2 \quad .$$

Note here that x_1 in this system is a constant. If A_{22} is assumed to be stable, we may interpret these two systems as an approximation to the original system over different time intervals. We see that the eigenvalues of $A(\varepsilon)$ where

$$A(\varepsilon) = \begin{bmatrix} A_{11} & A_{12} \\ \frac{A_{21}}{\varepsilon} & \frac{A_{22}}{\varepsilon} \end{bmatrix}$$

approach the eigenvalues of $\hat{A}_{11}$ and $\frac{A_{22}}{\varepsilon}$ where

$$\hat{A}_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21} \quad .$$

Now, notice that in the boundary layer system since x_1 is a constant that $(\frac{A_{21}x_1}{\varepsilon})$ can be regarded as a constant input to the stable system

$$\dot{x}_2 = \frac{A_{22}}{\varepsilon} x_2 + u \ , \quad u = \frac{A_{21}x_1}{\varepsilon} \quad .$$

Therefore, in the original system, when ε is near zero, the

eigenvalues of A_{22}/ε will be much greater in magnitude than the eigenvalues of $\hat{A}_{11}$ and $x_1(t)$ will remain essentially constant while the transients due to the initial condition $x_2(0)$ die out. Thus, $x_2(t)$ approaches $-A_{22}^{-1}A_{21}x_1(0)$ and from this time on, the system may clearly be modeled as the degenerate system. Here it is important to notice the separation of the time scales of the substates $x_1(t)$ and $x_2(t)$ in addition to the separation of eigenvalues of $A(\varepsilon)$ when ε is near zero. Systems may have eigenvalues that are widely separated, but the substates defined may be of mixed modes corresponding to both fast and slow eigenvalues. Hence, a similarity transformation is necessary in these cases in order that the substates defined should have different time scales. It is also important to realize that we refer to $x_2(t)$ as being a fast substate (or has a fast component) only over the interval that the boundary layer system model applies, since as t becomes large

$$x_2(t) = - A_{22}^{-1}A_{21}x_1(t) \quad .$$

Kokotovic and Chow [24] define a system partitioned as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

to possess a two time scale property if it can be decomposed by a similarity transformation into two subsystems

$$\begin{bmatrix} \dot{x}_s \\ \dot{x}_f \end{bmatrix} = \begin{bmatrix} A_s & 0 \\ 0 & A_f \end{bmatrix} \begin{bmatrix} x_s \\ s_f \end{bmatrix}$$

where

$$|\lambda(A_s)|_{max} << |\lambda(A_f)|_{min}$$

which is satisfied if

$$||A_f^{-1}|| << ||A_s||^{-1} \quad .$$

After extended argument, this last condition can be shown [31] to be satisfied if

$$||A_{22}^{-1}|| \leq \frac{1}{3} [||A_{11}|| + ||A_{12}|| \; ||A_{22}^{-1}A_{21}||]^{-1}$$

and

$$||A_{22}^{-1}|| << ||A_{11}||^{-1} \quad .$$

These last conditions will guarantee the system to possess the two time scale property. Note however, that this is only a sufficient condition and not strictly necessary, which can be seen by letting A be defined as

$$A = \begin{bmatrix} -10 & 9 \\ 9 & -10 \end{bmatrix}$$

In this case, A has eigenvalues $-1, -19$

but does not satisfy the two time scale inequalities of Kokotovic and Chow.

A somewhat different, but related definition is that of a singularly perturbed structure. This is concerned more with the magnitudes of the block elements $A_{ij}(\varepsilon)$ than with the eigenvalues $A(\varepsilon)$ which are imposed by the singular perturbation method. We define a system to have a singularly perturbed structure if it can be put in the form

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ \frac{A_{21}}{\varepsilon} & \frac{A_{22}}{\varepsilon} \end{bmatrix}$$

where

$$\varepsilon << 1 \quad \text{and} \quad ||A_{11}|| \simeq ||A_{22}|| \quad .$$

Alternatively, instead of using the conventional assumption $\varepsilon << 1$ and $||A_{11}|| \simeq ||A_{22}||$, we may assume that

$$\varepsilon = 1 \quad \text{and} \quad ||A_{11}|| << ||A_{22}||$$

and use this as an alternate definition. Note also, that if approaches zero, this is equivalent to $||A_{21}||$ becoming large relative to $||A_{11}||$ for $\varepsilon = 1$. To clarify the relationships between two time scale systems and systems with singularly perturbed structures, the following examples are given. Consider two system matrices A and F where

$$A = \begin{bmatrix} -10 & 9 \\ 9 & -10 \end{bmatrix} \text{ and } \quad F = \begin{bmatrix} 0 & 1 \\ -9 & -6 \end{bmatrix} .$$

As before, A has eigenvalues -1, -19, whereas F has a multiple eigenvalue of -3. Notice that A does not have a singularly perturbed structure since

$$||-10|| \not\ll ||-10||$$

but that F does since

$$||0|| << ||6||$$

corresponding to a very small ε in the conventional sense. Hence, systems that do not possess widely separated eigenvalues may still fit the singularly perturbed structure induced by the singular perturbation for some range of ε. On the other hand, systems that do possess a two time scale property may not exhibit fast and slow substates without first performing a similarity transformation (i.e. linearly recombine the states to form a new set of states) so that its two time scale nature cannot be used to advantage.

In the A matrix of the first example, not even reindexing or scaling will put the system in a singularly perturbed structure and it often must be physical insight that provides the key as to how to define the right set of states so that

singular perturbations can be used.

Relation of Algorithm and Asymptotic Series

Consider again the partitioned linear system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ \frac{A_{21}}{\varepsilon} & \frac{A_{22}}{\varepsilon} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

where $||A_{11}|| << ||A_{22}||$ so that $\varepsilon = 1$ is small parameter relative to $||A_{22}||/||A_{11}||$. The Lyapunov equation for this system is

$$\begin{bmatrix} A_{11}^T & \frac{A_{21}^T}{\varepsilon} \\ A_{12}^T & \frac{A_{22}^T}{\varepsilon} \end{bmatrix} \begin{bmatrix} K_{11}(\varepsilon) & K_{12}(\varepsilon) \\ K_{21}(\varepsilon) & K_{22}(\varepsilon) \end{bmatrix} + \begin{bmatrix} K_{11}(\varepsilon) & K_{12}(\varepsilon) \\ K_{21}(\varepsilon) & K_{22}(\varepsilon) \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ \frac{A_{21}}{\varepsilon} & \frac{A_{22}}{\varepsilon} \end{bmatrix}$$

$$= \begin{bmatrix} A_{11}^T & A_{21}^T \\ A_{12}^T & A_{22}^T \end{bmatrix} \begin{bmatrix} K_{11}(\varepsilon) & \varepsilon \hat{K}_{12}(\varepsilon) \\ \hat{K}_{21}(\varepsilon) & \hat{K}_{22}(\varepsilon) \end{bmatrix} + \begin{bmatrix} K_{11}(\varepsilon) & \hat{K}_{12}(\varepsilon) \\ \varepsilon \hat{K}_{21}(\varepsilon) & \hat{K}_{22}(\varepsilon) \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$= - \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

where $\hat{K}_{12}(\varepsilon) = \frac{1}{\varepsilon} K_{12}(\varepsilon)$, $\hat{K}_{21}(\varepsilon) = \frac{1}{\varepsilon} K_{21}(\varepsilon)$ and $\hat{K}_{22}(\varepsilon) = \frac{1}{\varepsilon} K_{22}(\varepsilon)$.

Define

$$\tilde{K}(\varepsilon) = \begin{bmatrix} K_{11}(\varepsilon) & \tilde{K}_{12}(\varepsilon) \\ \tilde{K}_{21}(\varepsilon) & \tilde{K}_{22}(\varepsilon) \end{bmatrix}$$

then, the Lyapunov equation can be written as

$$L_1(\tilde{K}(\varepsilon)) + \varepsilon L_2(\tilde{K}(\varepsilon)) = -Q$$

where L_1 and L_2 are given as in the algorithm as

$$L_1(K) = \begin{bmatrix} A_{11}^T K_{11} + K_{11} A_{11} + A_{21}^T K_{21} + K_{12} A_{21} & A_{21}^T K_{22} + K_{11} A_{12} + K_{12} A_{22} \\ K_{22} A_{21} + A_{21}^T K_{11} + A_{22}^T K_{21} & A_{22}^T K_{22} + K_{22} A_{22} \end{bmatrix}$$

$$L_2(K) = \begin{bmatrix} 0 & A_{11}^T K_{12} \\ K_{21} A_{11} & A_{12}^T K_{12} + K_{21} A_{12} \end{bmatrix} .$$

Now, we try to expand $\tilde{K}(\varepsilon)$ in an asymptotic series in ε about $\varepsilon = 0$ and evaluate it at $\varepsilon = 1$. A sufficient condition that $\tilde{K}(0)$ exist and be unique is the requirement that A_{22} be stable [23]. Therefore, we may define K^m as

$$K^m = \sum_{r=0}^{m} \frac{1}{r!} \left[\frac{\partial^r \tilde{K}(\varepsilon)}{\partial \varepsilon^r}\right]_{\varepsilon=0}$$

which is the $(m+1)^{\underline{th}}$ partial sum of the power series expansion in ε of $\tilde{K}(\varepsilon)$ evaluated at $\varepsilon = 1$. To calculate the derivatives of $\tilde{K}(\varepsilon)$, we differentiate successively the Lyapunov equation in operator from and get

$$\left[\frac{\partial \tilde{K}(\varepsilon)}{\partial \varepsilon^r}\right]_{\varepsilon=0} = - r\, L_1^{-1} \cdot L_2 \left[\frac{\partial^{r-1} \tilde{K}(\varepsilon)}{\partial \varepsilon^{r-1}}\right]_{\varepsilon=0}$$

or

$$\left[\frac{\partial^r \tilde{K}(\varepsilon)}{\partial \varepsilon^r}\right]_{\varepsilon=0} = r!(- L_1^{-1} \cdot L_2)^r [L_1^{-1}(Q)] \quad .$$

Using this last equality in the expression for K^m, we have that

$$K^m = \sum_{r=0}^{m} (- L_1^{-1} \cdot L_2)^r [L_1^{-1}(Q)] \quad .$$

Hence, the $m^{\underline{th}}$ partial sum of the asymptotic series is simply the $m^{\underline{th}}$ iterate of the two time scale decomposition algoithm of this chapter. This correspondence is the result of the linearity of L_1 and L_2 and does not extend to the nonlinear case, so that the extension of this algorithm to the Riccati equation, developed in Section 3.5, is not equivalent to an asymptotic series expansion.

However, having shown this equivalence in the linear case,

some remarks on the convergence of the algorithm and the two time scale property along with singularly perturbed structures is in order. Since singular perturbation methods are usually associated in the control area with multiple time scale systems, it is natural to assume that the necessary and sufficient condition for the algorithm to converge, namely $\rho(L_1^{-1}, L_2) < 1$, implies a two time scale system. This is not the case. Examination of our two previous examples where

$$A = \begin{bmatrix} -10 & 9 \\ 9 & -10 \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 0 & 1 \\ -9 & -6 \end{bmatrix}$$

and represent system matrices show that for A

$$\rho_A(L_1^{-1} \cdot L_2) = \left|\frac{-10}{-10}\right| = 1$$

while for F

$$\rho_F(L_1^{-1} \cdot L_2) = \left|\frac{0}{-6}\right| = 0 \quad .$$

Yet, A having eigenvalues of -1 and -19 has the two time scale property while F having the multiple eigenvalue of -3 does not. However, F does possess a singularly perturbed structure. Hence, under reasonable assumptions about the coupling terms, the conjecture that

$$||A_{11}||\ ||A_{22}^{-1}|| << 1$$

is a reliable indicator of convergence implies that the system possesses a singularly perturbed structure often associated with two time scale systems. However, as seen before, these two properties are not always associated. It is nevertheless, convenient to refer to the algorithm as a two time scale method since every two time scale system can be put into a singularly perturbed structure by the proper definition of the states (i.e. similarity transformation).

3.5 Decomposition Algorithm for the Riccati Equation

The decomposition algorithm for the Lyapunov equation can be extended to the Riccati equation using some additional constraints. The system and associated cost functional considered are respectively

$$\dot{x} = Ax + Bu$$

$$J(u) = \int_0^\infty (x^T Q x + u^T R u)\,dt \quad .$$

Assume that $[A,B]$ controllable and $[A,\sqrt{Q}]$ observable with $Q \geq 0$ and $R > 0$, then

$$A^T K + KA + Q - KBR^{-1}B^T K = 0$$

has a unique solution $K > 0$. Partitioning the Riccati equation, let

$$K = K_1 + K_2 = \begin{bmatrix} K_{11} & 0 \\ K_{21} & K_{22} \end{bmatrix} + \begin{bmatrix} 0 & K_{12} \\ 0 & 0 \end{bmatrix}$$

then the algorithm is defined by

$$A^T K_1^m + (K_1^m)^T A + Q^m - (K_1^m)^T B R^{-1} B^T K_1^m = 0$$

where

$$Q^m = Q + A^T K_2^{m-1} + (K_2^{m-1})^T A - (K_2^{m-1})^T B R^{-1} B^T K_2^{m-1}$$

$$-(K_1^{m-1})^T B R^{-1} B^T K_2^{m-1} - (K_2^{m-1})^T B R^{-1} B^T K_1^{m-1}$$

with K_2^o being initialized as zero. After tedious algebraic manipulation, which can be found in the Appendix in Chapter V, the algorithm can be written in a form analogous to the linear (Lyapunov) case.

Fast Subproblem

$$A_{22}^T K_{22}^m + K_{22}^m A_{22} + Q_{22}^m - K_{22}^m B_2 R^{-1} B_2^T K_{22}^m = 0 \quad .$$

Slow Subproblem

$$\hat{A}_{11}^T K_{11}^m + K_{11}^m \hat{A}_{11} + \hat{Q}_{11}^m - (K_{11}^m \hat{B}_1 + S^m)(R^m)^{-1}(K_{11}^m \hat{B}_1 + S^m)^T = 0$$

Coordinator's Problem

$$K_{12}^m (A_{22} - B_2 R^{-1} B_2^T K_{22}^m) + K_{11}^m (A_{12} - B_1 R^{-1} B_2^T K_{22}^m) + A_{21}^T K_{22}^m + Q_{12}^m = 0$$

$$Q_{12}^m = Q_{12} + [A_{11}^T - (K_{11}^{m-1} B_1 + K_{12}^{m-1} B_2) R^{-1} B_1^T] K_{12}^{m-1}$$

$$Q_{22}^m = Q_{22} + A_{12}^T K_{12}^{m-1} + K_{21}^{m-1} A_{12} - K_{21}^{m-1} B_1 R^{-1} B_1^T K_{12}^{m-1}$$

$$- K_{21}^{m-1} B_1 R^{-1} B_2^T K_{22}^{m-1} - K_{22}^{m-1} B_2 R^{-1} B_1^T K_{12}^{m-1}$$

$$Q_{11}^m = Q_{11} + (A_{22}^{-1} A_{21})^T Q_{22}^m (A_{22}^{-1} A_{21}) - Q_{12}^m (A_{22}^{-1} A_{21}) - (A_{22}^{-1} A_{21})^T Q_{21}^m$$

$$S^m = [(A_{22}^{-1} A_{21})^T Q_{22}^m - Q_{12}^m](A_{22}^{-1} B_2)$$

$$R^m = R + (A_{22}^{-1} B_2)^T Q_{22}^m (A_{22}^{-1} B_2)$$

where

$$\hat{A}_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21} \quad \text{and} \quad \hat{B}_1 = B_1 - A_{12}A_{22}^{-1}B_2 \quad .$$

Also, define $\hat{Q}_{11}$ as in the Lyapunov algoirthm as

$$\hat{Q}_{11} = Q_{11} + (A_{22}^{-1}A_{21})^T Q_{22}(A_{22}^{-1}A_{21}) - Q_{12}(A_{22}^{-1}A_{21}) - (A_{22}^{-1}A_{21})^T Q_{21}.$$

Now, the algoirthm is initialized with $K_{12}^o = 0$ in the coordination equations for Q_{12}^1 and Q_{22}^1. With $K_{12}^o = 0$ we find that K^1 is the usual signular perturbation approximation to the solution K. Note that if A_{22} is singular, the equations must be written in a different form.

For the algorithm to be well-defined, the subproblem and coordination equations must all have solutions. Sufficient conditions for this to happen are the conditions $[\hat{A}_{11},\hat{B}_1]$ and $[A_{22},B_2]$ controllable, $[\hat{A}_{11},\sqrt{\hat{Q}_{11}}]$ and $[A_{22},\sqrt{Q_{22}}]$ observable and the requirement that $Q^m \geq 0$ for all m. This is exactly the requirement that the boundary layer system and the degenerate system are both controllable and observable, where observability refers to the cost observability of the state in the quadratic cost functional [29,32]. Also, the original system is assumed to be controllable and observable.

The nonlinear nature of the iteration precludes a global analysis, but local convergence conditions are quite simple. The coordination strategy of the decomposition algorithm is <u>locally successful</u> if there exists a neighborhood of K the solution such that if K^m is an element of that neighborhood,

then $\lim_{m\to\infty} K^m = K$.

Theorem 3.5.1

Define the linear operators $\bar{L}_1 : R^{nxn} \to R^{nxn}$ and $\bar{L}_2 : R^{nxn} \to R^{nxn}$ by the linear operators L_1 and L_2 respectively of Theorem 3.2.1 where A is replaced by $\bar{A}$ and

$$\bar{A} = A - BR^{-1}B^TK \quad .$$

Assume $[\hat{A}_{11},\hat{B}_1]$ and $[A_{22},B_2]$ controllable, $[\hat{A}_{11},\sqrt{\hat{Q}_{11}}]$ and $[A_{22},\sqrt{Q}_{22}]$ observable and

$$Q^m \Big|_{K^m=K} > 0 \quad .$$

The algorithm has a locally successful coordination strategy if

$$\rho(\bar{L}_1^{-1} \cdot \bar{L}_2) < 1$$

and only if

$$\rho(\bar{L}_1^{-1} \cdot \bar{L}_2) \leq 1 \quad .$$

Proof:

Let $K^m = K + \Delta K_1^m + \Delta K_2^{m-1}$. Then, the linearized equation

$$\bar{A}^T(\Delta K_1^m + \Delta K_2^{m-1}) + (\Delta K_1^m + \Delta K_2^{m-1})\bar{A} = 0$$

is obtained and must converge to $\Delta K^m = 0$ by Theorem 3.2.1 for K^m in some neighborhood of K. For $||\Delta K^m||$ sufficiently small, the convergence of the algorithm is given by

the linearized equation. Also, for $||\Delta K^m||$ sufficiently small (i.e. K^m in some neighborhood of K) $Q^m \geq 0$ so that the subproblems are well defined. Thus, when $\rho(\bar{L}_1^{-1} \cdot \bar{L}_2) < 1$, there exists a neighborhood about K such that the algorithm converges Q.E.D. The gap between the sufficient condition and the necessary condition arises because for $\rho(\bar{L}_1^{-1} \cdot L_2) = 1$, the nonlinear terms not included in the linearized equation in ΔK^m govern the convergence.

It is important here to realize that it is the closed loop matrix $\bar{A}$ that determines the convergence of the algorithm in the neighborhood of K, and not open loop matrix A. Therefore, the feedback may alter the natural separation of modes in the physical system in either a favorable or unfavorable fashion. Note, also that for systems that possess a singularly perturbed form so that

$$A_{21} = \frac{1}{\varepsilon}\tilde{A}_{21}, \quad A_{22} = \frac{1}{\varepsilon}\tilde{A}_{22}, \quad B_2 = \frac{1}{\varepsilon}\tilde{B}_2$$

there always exists an ε sufficiently small so that the algorithm converges. This follows from the asymptotic behavior of $K_{12} = \varepsilon\tilde{K}_{12}$ and $K_{22} = \varepsilon\tilde{K}_{22}$ where $\tilde{K}_{12}$ and $\tilde{K}_{22}$ have finite limits as $\varepsilon \to 0^+$ so that

$$Q^m \Big|_{K^m=K} > 0$$

and

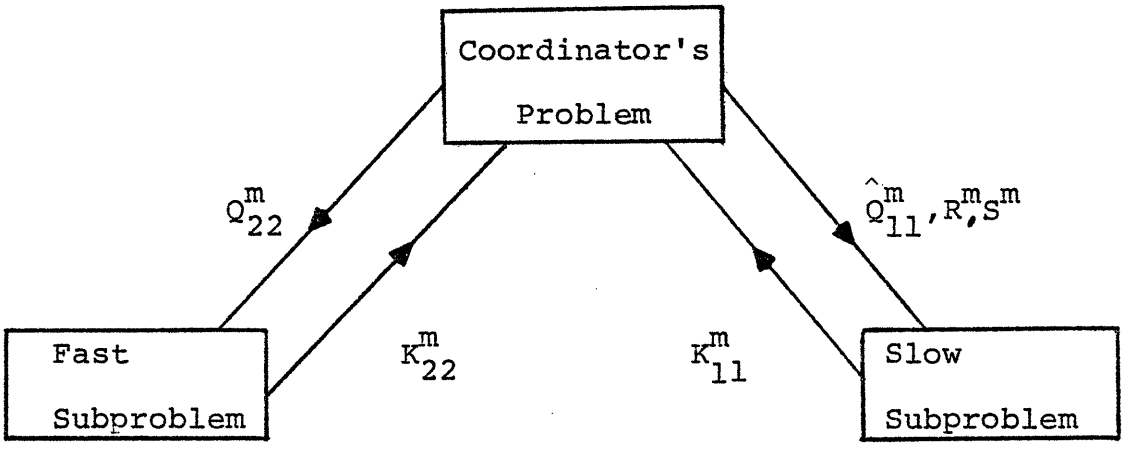
$$\rho(\bar{L}_1^{-1} \cdot \bar{L}_2) < 1$$

are readily verified from the fact that K^1 is the singular perturbation approximation to K which is known to approach K as $\varepsilon \to 0^+$ [23,29].

3.6 Hierarchial Structure of the Algorithm

It has been mentioned before that singular perturbation methods applied to control and estimation problems yield hierarchically structured filters and controllers that are asymptotically optimal [20,21,22]. The hierarchial structure of the filters or controllers result from using the gains computed from the singularly perturbed control and filter Riccati equations and processing information hierarchically. Information from the slow substate is passed to the controller or filter, but not vice versa. However, this is not at all the same hierarchical structure of the decomposition algorithm. In the decomposition algorithm, the hierarchy is not divided between the fast and slow subproblems, but between the coordination scheme and the subproblems. The decomposition algorithm which computes the gains to be used by the filters and controllers is an off-line hierarchy, while the filters and controllers represent an on-line hierarchy which process the information of observations or outputs. In the decomposition algorithm, the coordinator representing the supremal decision maker guides the subproblem or infimal solutions to the global solution. This is depicted in Figure 3.6.1 for the Riccati equation and back in Figure 3.2.1 for the Lyapunov equation.

In terms of decentralization of information, the fast subproblem requires the least information about the rest of the system needing only A_{22}, the fast boundary layer model.



Two-Time Scale
Decomposition Algorithm for
the Riccati Equation

Figure 3.6.1

The slow subproblem requires the slow degenerate model $\hat{A}_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ and is uncoupled from the fast subproblem. The coordinator, however, needs information about the total system or the full order model.

The basic novelty of this algorithm is the decomposition of the global problem into subproblems

$$\dot{x}_1 = \hat{A}_{11}x_1 + \hat{B}_1u_1$$

$$J_1 = \int_0^\infty [x_1^T u_1^T] \begin{bmatrix} \hat{Q}_{11}^m & S^m \\ (S^m)^T & R^m \end{bmatrix} \begin{bmatrix} x_1 \\ u_1 \end{bmatrix} dt$$

and

$$\dot{x}_2 = A_{22}x_2 + B_2u_2$$

$$J_2 = \int_0^\infty (x_2^T Q_{22}^m x_2 + u_2^T R u_2)dt$$

with $u = u_1 + u_2$ on the basis of multiple time scales rather than the usual weak coupling approach. This decomposition can be viewed as two reduced order aggregate models of the overall system. From this perspective, the slow or degenerate model approximates the fast subsystem by its steady state, while the fast or boundary layer system approximates the slow subsystem by its initial value a constant. This approach could be adapted to the case where the state dynamics are nonlinear or the general two point value boundary problem of optimal control.

CHAPTER IV

WEAK COUPLING ITERATIVE DECOMPOSITION

4.1 WEAK COUPLING ALGORITHM FOR THE LYAPUNOV EQUATION

For large systems composed of smaller sybsystems that are interconnected weakly, different decompositions are practical [4,23,25]. Several variations of a weak coupling algorithm proposed by Athay [25] for the Lyapunov equation will be described here.

Basic Weak Coupling Algorithm

For the linear system

$$\dot{x} = Ax$$

where

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ \vdots & & & \vdots \\ & A_{22} & & \\ & & \ddots & \\ A_{N1} & \cdots & & A_{NN} \end{bmatrix}$$

we associate the Lyapunov equation

$$A^T K + KA + Q = 0 \quad .$$

Conformably partitioning K and Q the Lyapunov equation is

$$\sum_{m=1}^{N} (A_{mi}^T K_{mj} + K_{im} A_{mj}) = - Q_{ij} \quad .$$

$$A_D = \begin{bmatrix} A_{11} & & & \\ & A_{22} & & \\ & & \ddots & \\ & & & A_{NN} \end{bmatrix} \quad \text{and} \quad A_c = A - A_D \quad .$$

The matrix A_D represents the individual subsystems and A_c represents the coupling among them. This splitting of the system matrix induces a splitting of the Lyapunov operator L_A defined by

$$L_A(X) \triangleq A^T X + XA \quad .$$

Letting

$$L_1(X) \triangleq L_{A_D}(X) \quad \text{and} \quad L_2(X) \triangleq L_{A_c}(X)$$

the Lyapunov equation can be written in the following manner

$$L_1(K) + L_2(K) + Q = 0 \quad .$$

If the coupling among the subsystems is weak $L_2(K)$ should be small compared to the contribution of $L_1(K)$ so that the approximate solution is given by

$$K \simeq L_1^{-1}(Q) \quad .$$

To include the effects of the coupling, we form the iteration

$$L_1(K^{m+1}) = - Q - L_2(K^m)$$

or

$$K^{m+1} = - L_1^{-1}(Q) - L_1^{-1} \cdot L_2(K^m) \quad .$$

Provided L_1^{-1} exists, the convergence of this iteration occurs if and only if

$$\rho(L_1^{-1} \; L_2) < 1$$

where $\rho(\cdot)$ denotes spectral radius. Alternatively, this iteration can be written in the form

$$A_{ii}^T K_{ij}^{m+1} + K_{ij}^{m+1} A_{jj} + Q_{ij}^{m+1} = 0 \; , \quad \begin{matrix} i=1,2,\ldots,N \\ j=i,i+1,\ldots,N \end{matrix}$$

where

$$Q_{ij}^{m+1} = Q_{ij} + \sum_{\substack{r=1 \\ r\neq i}}^{N} A_{ri}^T K_{rj}^m + \sum_{\substack{r=1 \\ r\neq j}}^{N} K_{ir}^m A_{rj}$$

which represents $N(N + 1)/2$ uncoupled generalized Lyapunov equations (i.e. of the form $A^TX + XB + C = 0$). Note that if A_{ii} is stable then L_1^{-1} exists and the iteration is well defined. Also, this iteration, if written as a linear vector equation iteration will be recognized as a Jacobi iteration [13,14].

Relation to Power Series Expansion

There is a direct relationship between this iteration and the approximations found by using regular perturbation theory [23,19].

Consider the linear system

$$\dot{x} = Ax$$

where

$$A = A(\varepsilon) = \begin{bmatrix} A_{11} & \varepsilon A_{12} & \cdots & \varepsilon A_{1N} \\ \varepsilon A_{21} & A_{22} & & \varepsilon A_{2N} \\ \vdots & & \ddots & \vdots \\ \varepsilon A_{N1} & \varepsilon A_{N2} & \cdots & A_{NN} \end{bmatrix} = A_D + \varepsilon A_c \quad .$$

The Lyapunov equation for this system is

$$L_1(K(\varepsilon)) + \varepsilon L_2(K(\varepsilon)) = -Q$$

or

$$K(\varepsilon) = -\varepsilon L_1^{-1} \cdot L_2(K(\varepsilon)) - L_1^{-1}(Q) \quad .$$

Differentiating this last equation successively and evaluating the derivatives at $\varepsilon = 0$, we obtain

$$\left[\frac{\partial^m K(\varepsilon)}{\partial \varepsilon^m}\right]_{\varepsilon=0} = m![(-L_1^{-1} \cdot L_2)^m(K(0))] \quad .$$

Hence, the sum of the first m terms of the power series expansion of $K(\varepsilon)$ about $\varepsilon = 0$ and evaluated at $\varepsilon = 1$ is

$$K^m = \sum_{i=0}^{m-1} [-(-L_1^{-1} \cdot L_2)^i \cdot L_1^{-1}(Q)]$$

where K^m is the $m^{\underline{th}}$ iterate of the basic weak coupling algorithm.

Gauss Seidel Variations of the Weak Coupling Algorithm

The Jacobi iteration can be modified to a Gauss Seidel iteration by simply utilizing each K_{ij}^{m+1} as soon as it is available instead of K_{ij}^{m} in the basic weak coupling algorithm (Jacobi iteration). Depending on how the indices i and j are varied to select the different subproblems, it is possible to form different versions of the Gauss Seidel iteration. These versions may converge at different rates since the order of updating the block matrices of K induces a slightly different splitting of the Lyapunov operator and hence may change the spectral radius of $L_1^{-1} \cdot L_2$.

Two possible orderings can be seen to be effective for special forms of A , and seem to be likely candidates for an algorithm to handle a general A matrix. To illustrate the relationships of these variations, the 2 x 2 block case will be shown.

Jacobi

$$A_{11}^T K_{11}^{m+1} + K_{11}^{m+1} A_{11} = - [Q_{11} + A_{21}^T K_{21}^m + K_{12}^m A_{21}]$$

$$A_{11}^T K_{12}^{m+1} + K_{12}^{m+1} A_{22} = - [Q_{12} + A_{21}^T K_{22}^m + K_{11}^m A_{12}]$$

$$A_{22}^T K_{22}^{m+1} + K_{22}^{m+1} A_{22} = - [Q_{22} + A_{12}^T K_{12}^m + K_{21}^m A_{12}]$$

Gauss Seidel #1

$$A_{11}^T K_{11}^{m+1} + K_{11}^{m+1} A_{11} = - [Q_{11} + A_{21}^T K_{21}^m + K_{12}^m A_{21}]$$

$$A_{11}^T K_{12}^{m+1} + K_{12}^{m+1} A_{22} = - [Q_{12} + A_{21}^T K_{22}^m + K_{11}^{m+1} A_{12}]$$

$$A_{22}^T K_{22}^{m+1} + K_{22}^{m+1} A_{22} = - [Q_{22} + A_{12}^T K_{12}^{m+1} + K_{21}^{m+1} A_{12}]$$

Gauss Seidel #2

$$A_{11}^T K_{11}^{m+1} + K_{11}^{m+1} A_{11} = - [Q_{11} + A_{21}^T K_{21}^m + K_{12}^m A_{21}]$$

$$A_{22}^T K_{22}^{m+1} + K_{22}^{m+1} A_{22} = - [Q_{22} + A_{12}^T K_{12}^m + K_{21}^m A_{12}]$$

$$A_{11}^T K_{12}^{m+1} + K_{12}^{m+1} A_{22} = - [Q_{12} + A_{21}^T K_{22}^{m+1} + K_{11}^{m+1} A_{12}] \quad .$$

In all of the above iterations Q is assumed to be symmetric and hence K is symmetric.

For a block diagonal A (i.e. $A_{12} = 0$ and $A_{21} = 0$) all three methods converge in one step. However, for a block triangular A, the Jacobi iteration converges in three steps. Note that it makes no difference what order the equations in the Jacobi iteration are solved. For Gauss Seidel #1 iteration, the method converges in one step for upper triangular A ($A_{21} = 0$), but takes three steps to converge exactly for a lower triangular A ($A_{12} = 0$). The Gauss Seidel #1 iteration can also be written in terms of linear operators

$$L_1(K^{m+1}) = - L_2(K^m) - Q$$

where

$$L_A(X) = A^T X + XA$$

and

$$L_1(K) \triangleq L_{A_1}(K) \; ; \quad L_2(K) \triangleq L_{A_2}(K)$$

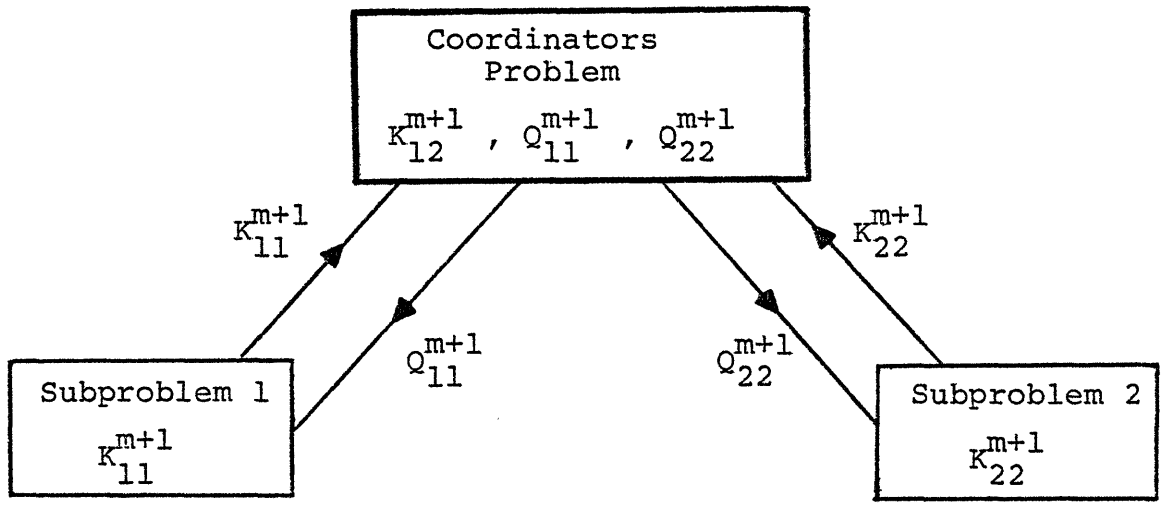
with A_1 being upper block triangular and A_2 being strictly lower block triangular such that

$$A = A_1 + A_2 \quad .$$

Note that the case in which it takes three iterations to converge could be reduced to one by requiring A_1 to be lower block triangular and A_2 strictly upper block triangular with again $A = A_1 + A_2$.

For the Gauss Seidel #2 iteration, a triangular A requires two steps to obtain the solution. This iteration might be used in the general case where the coupling matrices (A_{12} and A_{21}) are roughly the same "size". The Gauss Seidel #1 iteration in one of its two forms might be used for A matrices that came "close" to being block triangular. One further comment on the Gauss Seidel #2 iteration is that it is in the form of two uncoupled subproblems which are Lyapunov equations for the A_{11} and A_{22} subsystems with a coordination problem to calculate the effect of the coupling in K_{12}. This identification will be made again with the

Riccati equation in a later section. The following diagram (Figure 4.1.1) illustrates this particular identification with the decomposition.



$$Q_{11}^{m+1} = Q_{11} + A_{21}^{T} K_{21}^{m} + K_{12}^{m} A_{21}$$

$$Q_{22}^{m+1} = Q_{22} + A_{12}^{T} K_{12}^{m} + K_{21}^{m} A_{12}$$

Gauss Seidel #2 Algorithm -

Weak Coupling Decomposition for Lyapunov Equation

Figure 4.1.1

4.2 RELAXATION OF ALGORITHMS

To enhance the convergence characteristics of the various algorithms, relaxation techniques [13,14] can be applied to attempt to reduce the spectral radius of the iteration.

The spectral radius is the asymptotic convergence rate and reducing it may or may not significantly improve the average rate of convergence. Hopefully, the iteration reaches its asymptotic rate quickly. All of the algorithms for the Lyapunov equation in the previous sections can be written in the following form,

$$L_1(K^{m+1}) = - [L_3(K^{m+1}) + L_2(K^m) + Q]$$

where $L(K) = L_1(K) + L_2(K) + L_3(K)$ and where L_3 is a linear operator that uses only the elements of K^{m+1} which have been computed previously in computing the elements of $L_1(K^{m+1})$.

Solving for K^{m+1} we have

$$K^{m+1} = - [L_1^{-1} \cdot L_3(K^{m+1}) + L_1^{-1} \cdot L_2(K^m) + L_1^{-1}(Q)] \quad .$$

Applying relaxation to this iteration, we find

$$K^{m+1} = - w[L_1^{-1} \cdot L_3(K^{m+1}) + L_1^{-1} \cdot L_2(K^m) + L^{-1}(Q)]$$

$$+ (1 - w)K^m \quad .$$

Here, w is a relaxation parameter to be selected to try to

reduce the spectral radius of the iteration operator

$$(I + wL_1^{-1} \cdot L_3)^{-1}[(1 - w)I - wL_1^{-1} \cdot L_2] \quad .$$

For a Jacobi iteration L_3 is zero so that the spectral radius becomes

$$\rho\{(1 - w)I - wL_1^{-1} \cdot L_2\} = \max_i |(1 - w) - w\lambda_i(L_1^{-1} \cdot L_2)|$$

where $\lambda_i(L_1^{-1} \cdot L_2)$ is the $i^{\underline{th}}$ eigenvalue of $L_1^{-1} \cdot L_2$. Note, if $w = 1$, we are back to the original algorithm.

To calculate the spectral radius for the different iterations, it is necessary to write the linear operatiors in matrix form using the Kronecker product notation. However, in actually using the algorithms, the computation will be done from the standpoint of solving smaller Lyapunov and Sylvester equations (generalized Lyapunov equations). In terms of linear operators, this can be written,

$$L_1(K^{m+1}) = - w[L_2(K^{m+1}) + L_2(K^m) + Q]$$

$$+ (1 - w)L_1(K^m)$$

where $L_1(K^{m+1})$ is decomposed into a number of subproblems so that L_1^{-1} is not formed.

Relaxation of Gauss Seidel Algorithms

Writing the relaxed algorithms in terms of the subproblems we obtain

$$A_{ii}^T K_{ij}^{m+1} + K_{ij}^{m+1} A_{jj} + Q_{ij}^{m+1} = 0$$

$$Q_{ij}^{m+1} = w[Q_{ij} + \sum_{r=1}^{N} A_{ri}^T K_{rj}^{\alpha} + K_{ir}^{\beta} A_{rj}] - A_{ii}^T K_{ij}^m - K_{ij}^m A_{jj}$$

where α and β take on the values m or $m+1$ depending on the ordering of the subproblems.

In the Gauss Seidel #1 iteration $i = 1,2,\ldots,N$ and $j = i,i+1,\ldots,N$ is the indexing scheme. For the Gauss Seidel #2 iteration the indexing is $i = 1,2,\ldots,N$ and $j = i + k$ where $k = 0,1,\ldots,N - i$.

This second scheme decomposes the original problem into N coupled subproblems of Lyapunov type and $N(N-1)/2$ coupled coordination problems (Sylvester equations). The coupling, of course, is only produced by utilizing all information as it becomes available, the basic difference between the Jacobi and Gauss Seidel iterations, and does not arise from the use of relaxation.

Operation Count for Algorithms

Let $f(n)$ denote the number of operations required to solve a Lyapunov equation order n and $g(n)$ denote the number of operations to solve the generalized Lyapunov equation order n. Assuming Q is symmetric, there are N subsystems of dimension n and

$$f(n) = \alpha n^3 \; ; \quad \alpha \text{ constant}$$

$$g(n) = \beta n^3 \quad ; \quad \beta \text{ constant}$$
$$\beta > \alpha$$

the number of operations (multiplications) required per iteration may be calculated.

Calculating		Operations
$N(N+1)/2$	Q_{ij}^m's	$N(N+1)[2Nn^3]/2$
$N(N-1)/2$	K_{ij}^m's	$N(N-1)[\beta n^3]/2$
N	K_{ii}^m's	$N[\alpha n^3]$

For σ iterations, the total operation count T is

$$T = [N^2(N+1) + N(N-1)\frac{\beta}{2} + N\alpha]n^3\sigma$$

$$T = [N^2 + (1 + \frac{\beta}{2})N + (\alpha - \frac{\beta}{2})]Nn^3\sigma \quad .$$

For the decomposition algorithm to reduce the number of multiplications required as compared with the subproblem method

$$T < \alpha(Nn)^3$$

or

$$\sigma[1 + (1 + \frac{\beta}{2})\frac{1}{N} + (\alpha - \frac{\beta}{2})\frac{1}{N^2}] < \alpha \quad .$$

For most N this will approximately require that

$$\sigma < \alpha \quad .$$

However, suppose that A_{ij} , $i \neq j$ are sparse having only one nonzero row, which may be a fairly common situation with large scale systems. Then the number of multiplications required to compute Q^m_{ij} is only $2Nn^2$ instead of $2Nn^3$. Then the total operation count is

$$T = [\frac{N^2}{n} + (1 + \frac{\beta}{2})N + (\alpha - \frac{\beta}{2})]Nn^3\sigma$$

and to reduce the computations for most large N will require approximately

$$\frac{\sigma}{n} < \alpha \quad .$$

This shows if the size of the subsystems is appreciable, the computations can be greatly reduced. Also, if Q is block diagonal, it might be suspected that the resolution operation count for K_{ij} could be reduced. This unfortunately does not make a significant difference since even if Q is block diagonal Q^m will not be block diagonal. Hence, even using saved real Schur decompositions of A_{ij} as in Bartels-Stewart algorithm [30], the operation count is still order n^3 for the subproblems since Q^m_{ij} is not sparse and must be transformed at each resolution with an operation count order n^3. Nevertheless, as will be seen later, the assumption that $f(n) = \alpha n^3$ may be changed to $f(n) = \alpha n^2$ for special cases and significant reductions in computation will be possible. Therefore,

we conclude, if only the approximate solution corresponding to the first few terms in the Taylor series is needed, the number of computations will be reduced even in the general case. If sparsity is exploited or a particular canonical form is utilized, the reduction of computation will be significant.

4.3 POWER SYSTEM MODEL AND NUMERICAL RESULTS

Power System Example

To gain insight into the computational aspects of solving the Lyapunov equation, by the weak coupling algorithm with relaxation, a simple power system example was studied. Solution of the Lyapunov equation may serve as a measure of coherency between generators that form an interconnected system [25,26]. This is important since machines that form a coherent group tend to swing together under a major network disturbance.

A linearized model of a three generator infinite bus system was used, where each generator was represented by a constant voltage behind transient reactance. The linearized equations for this model are

$$M_1\dot{\Delta w}_1 + \Delta w_1/R_1 w_o = - Y_{11}\delta_1 + Y_{12}\delta_2 + Y_{13}\delta_3$$

$$\dot{\delta}_1 = w_1$$

$$M_2\dot{\Delta w}_2 + \Delta w_2/R_2 w_o = Y_{12}\delta_1 - Y_{22}\delta_2 + Y_{23}\delta_3$$

$$\dot{\delta}_2 = \Delta w_2$$

$$M_3\dot{\Delta w}_3 + \Delta w_3/R_3 w_o = Y_{13}\delta_1 + Y_{23}\delta_2 - Y_{33}\delta_3$$

$$\dot{\delta}_3 = \Delta w_3$$

where

Δw_i = frequency deviation of machine i

M_i = moment of inertia of machine i

δ_i = perturbation of machine i's rotor angle from the operating point

R_i = droop of machine i

Y_{ij} = transfer admittance between machines i and j

Y_{ii} = self admittance .

Athay [25] adapted this model from Lee [26], who studied the identification of coherent groups for transient stability analysis, by adding the damping terms in order to guarantee the existence of the solution to the Lyapunov equation. The model used here is the same as the one used by Athay except that a parameter was added to vary the coupling between generators, but keeping the admittance to the infinite bus constant.

This is represented as in Figure 4.3.1

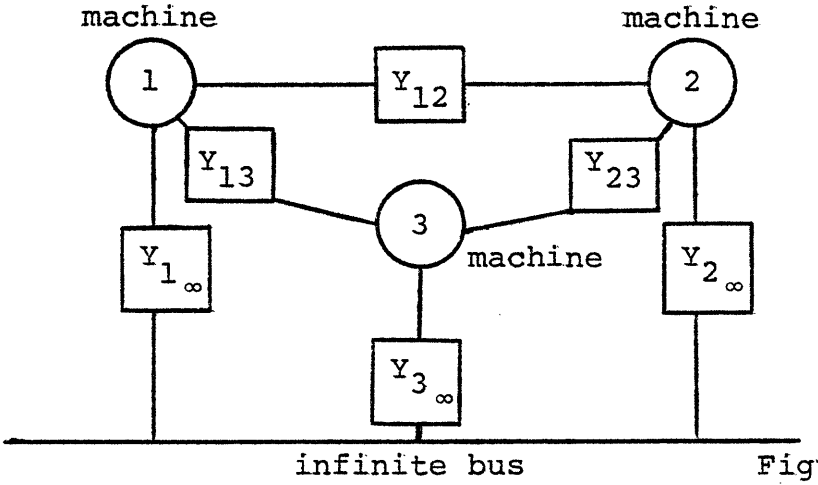


Figure 4.3.1

The per unit values used were

$$Y_{12} = \varepsilon \qquad Y_{1_\infty} = .3$$

$$Y_{13} = .9\varepsilon \qquad Y_{2_\infty} = .4$$

$$Y_{23} = 1.2\varepsilon \qquad Y_{3_\infty} = .2$$

$$Y_{11} = Y_{1_\infty} + Y_{12} + Y_{13}$$

$$Y_{22} = Y_{2_\infty} + Y_{12} + Y_{23}$$

$$Y_{33} = Y_{3_\infty} + Y_{13} + Y_{33}$$

$$w_o = 377 \text{ rad/sec}$$

$$M_1 = .1326 \qquad R_1 = .01$$

$$M_2 = .1592 \qquad R_2 = .01$$

$$M_3 = .1194 \qquad R_3 = .01 \quad .$$

Here ε is varried from one to zero. At $\varepsilon = 1$, the model is identical to Athay's and the coupling between machines is the strongest. Athay tried to solve the Lyapunov equation for this case by using Jacobi iteration (i.e. w = 1) or the iterative decoupling algorithm as he refers to it.

He found that the algorithm diverged because the spectral radius of the iteration was greater than one. This

happened because the generators were too tightly coupled as verified later by simulation. However, using relaxation on the Jacobi or Gauss Seidel iterations, it is possible to obtain convergence.

Numerical Results

The effectiveness of relaxation techniques was evaluated for both the Jacobi and Gauss Seidel algorithms by calculating the spectral radius for different values of the relaxation factor w and different values of ε the coupling factor. To obtain a visual estimate of the relative effects of the variations in the coupling parameter ε, a simulation of the system was done. Step inputs were applied to one machine and the response of all three machines in rotor angle perturbation and frequency deviation were recorded. The initial conditions in each simulation were set to zero. These simulations yield a comparison of the strength of the coupling and the convergence rate of the algorithms.

Results indicate that convergence though slow, may be achieved for even strongly coupled systems and as expected, a better convergence rate for more weakly coupled systems. Generally, the Gauss Seidel iteration converged faster if the Jacobi converged, but diverged faster if the Jacobi diverged. Also, for all cases examined in this example, the optimum relaxation factor was always less than one and greater than zero. For strongly coupled systems, it was the

smallest and approached one the more weakly coupled the system.

It should be mentioned that the spectral radius was computed for the Gauss Seidel #11 iteration. The spectral radius as a function of w and ε is tabulated in Table 4.3.2. In the plots that follow, a step input was applied to machine #1. The units of time are in seconds and the perturbation of rotor angle and frequency deviation are in radians and radians per second respectively.

WEAK COUPLING DECOMPOSITION EXAMPLE

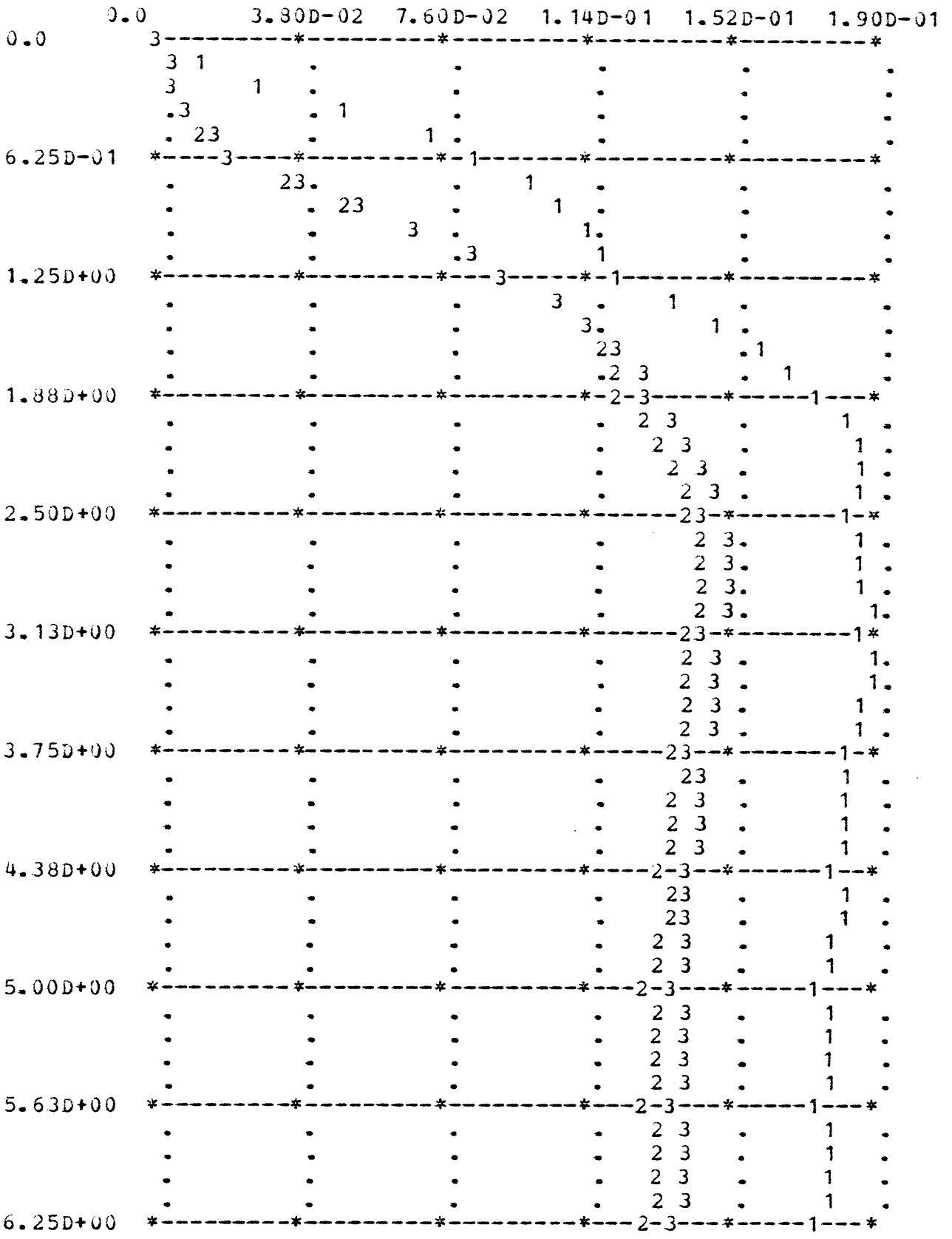
Relaxation factor - w
Coupling factor - ε
Spectral Radius - sr

Jacobi		Gauss Seidel		
w	sr	w	sr	ε
1.0	1.44	1.0	2.79	1.0
1.0	1.24	1.0	1.97	0.75
1.0	.99	1.0	1.19	0.50
1.0	0.64	1.0	0.46	0.25
1.0	0	1.0	0	0
0.5	0.94	0.6	0.90	1.0
0.6	0.90	0.8	0.80	0.75
0.8	0.83	0.9	0.72	0.50
1.0	0.64	1.0	0.46	0.25

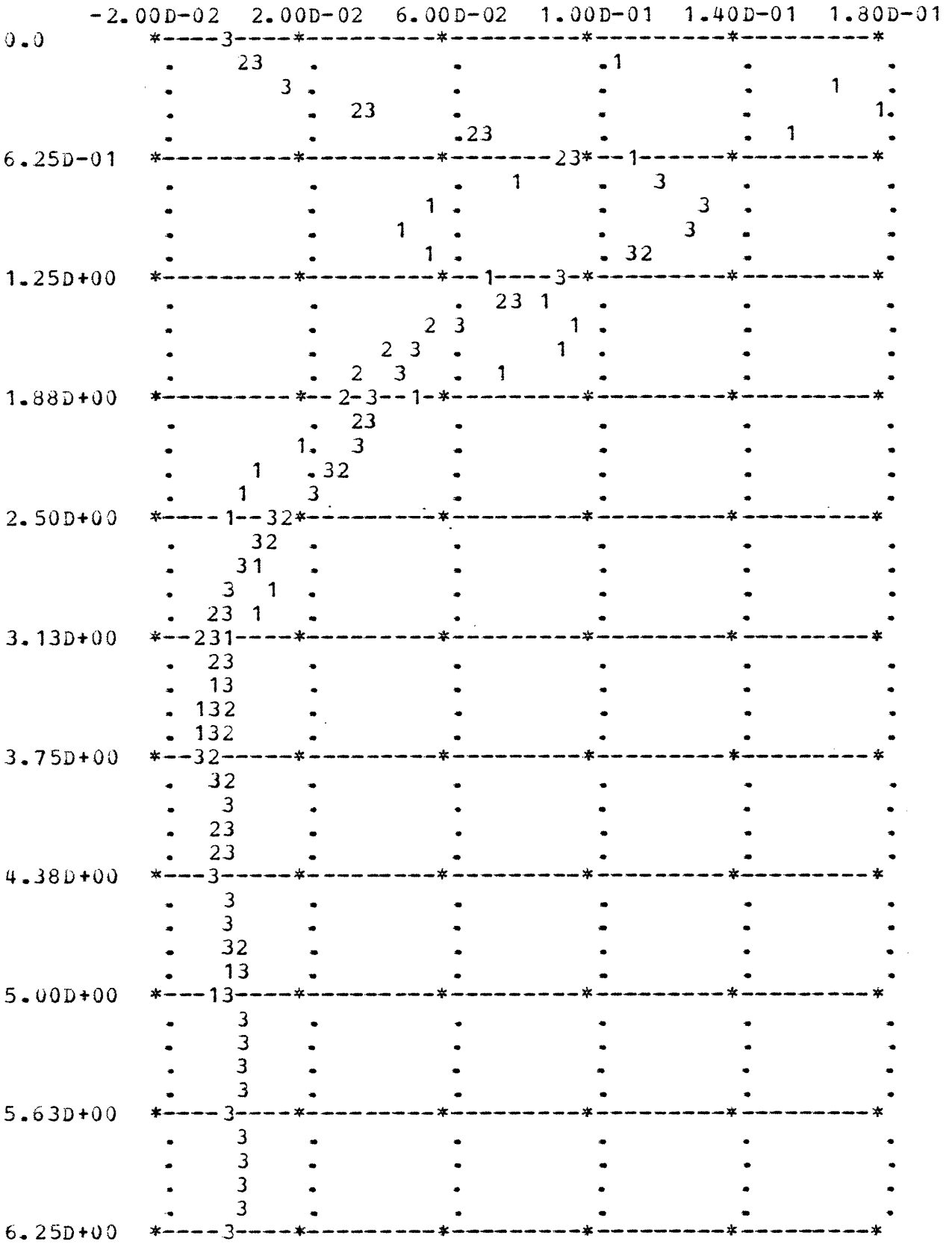
Three Machines Interconnected
Forming 6 x 6 System

Table 4.3.2

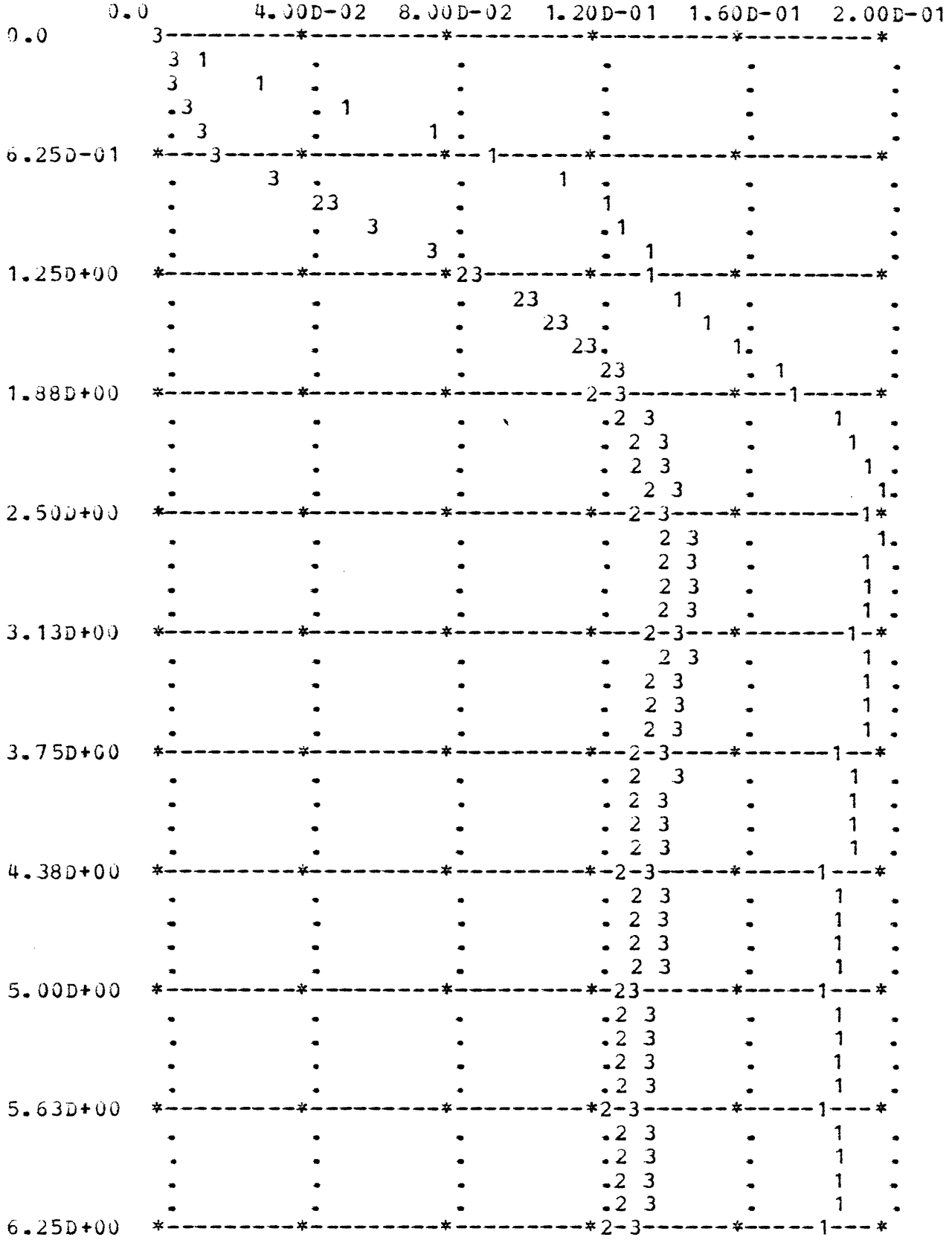
DELTA 1,2,3 VS. TIME
EPSILON= 1.00



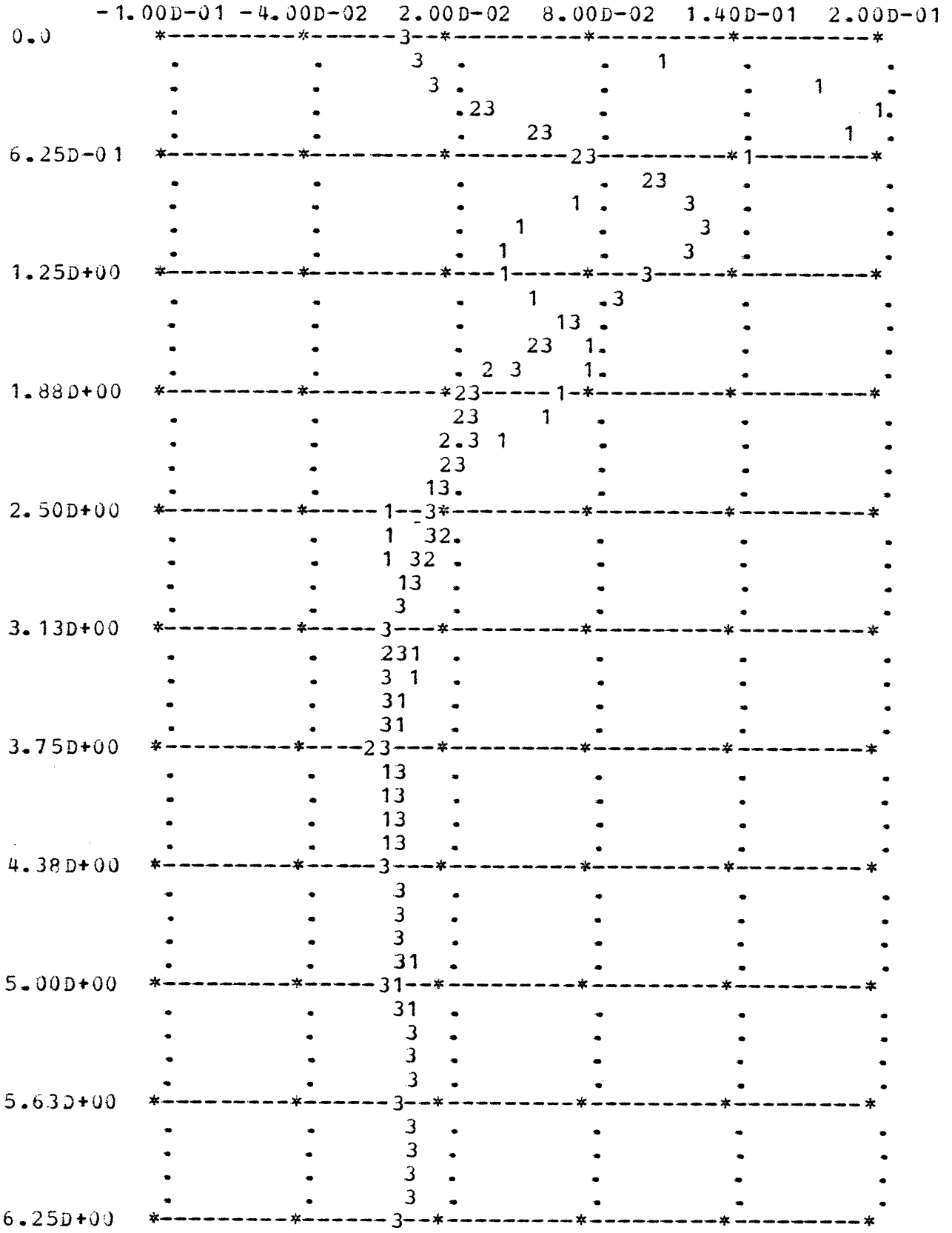
FREQ. 1,2,3 VS. TIME
EPSILON= 1.00



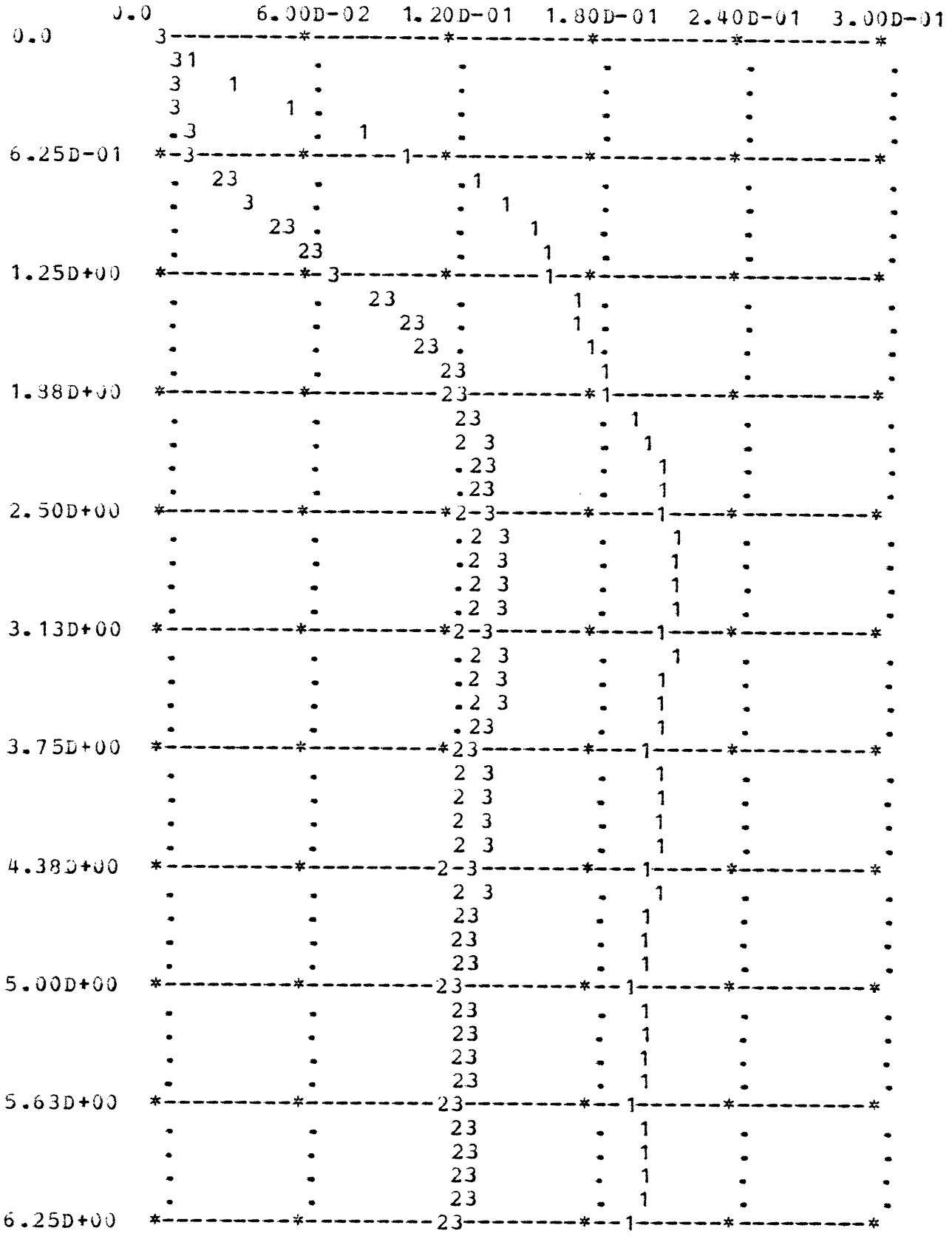
DELTA 1,2,3 VS. TIME
EPSILON= 0.75



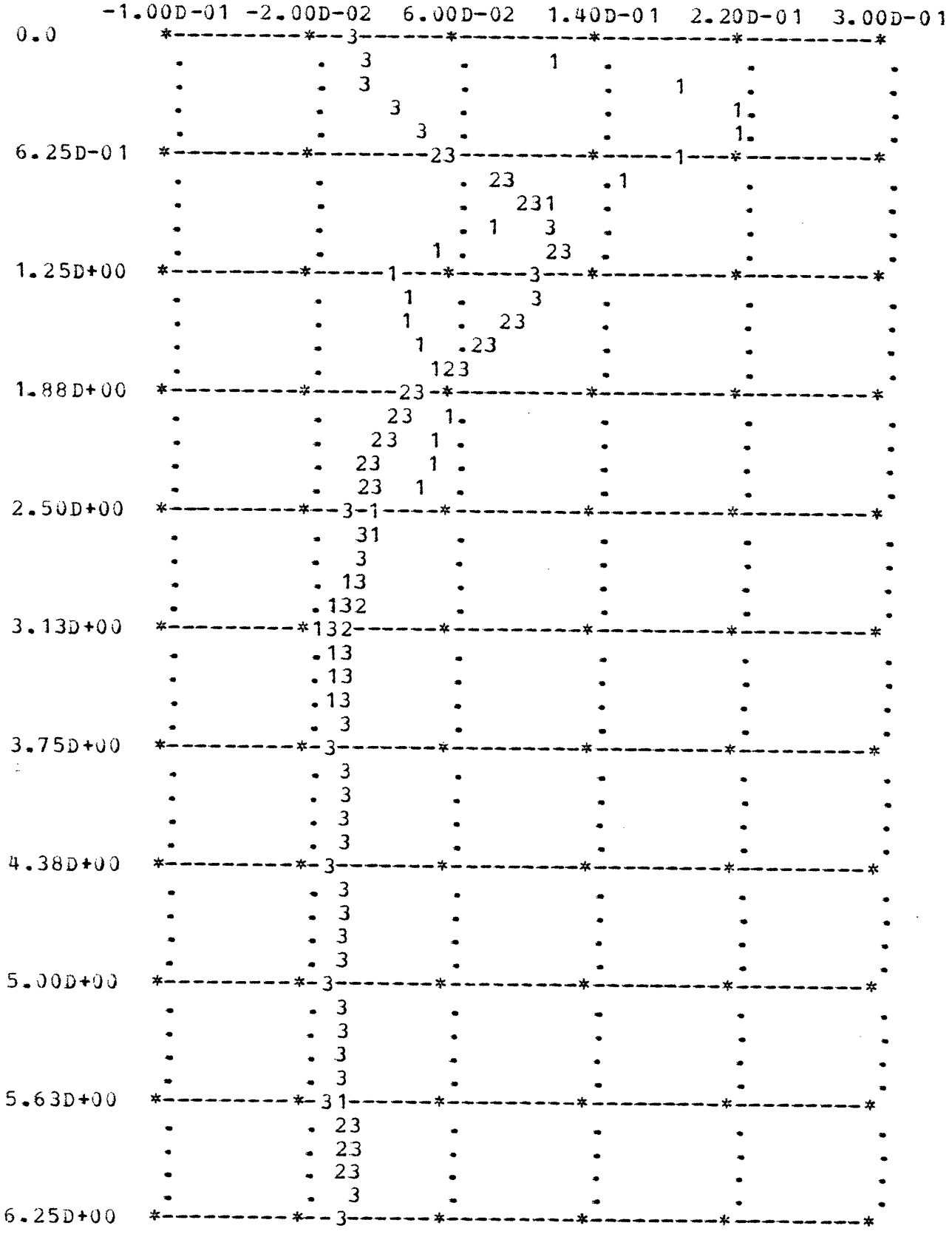
FREQ. 1,2,3 VS. TIME
EPSILON= 0.75



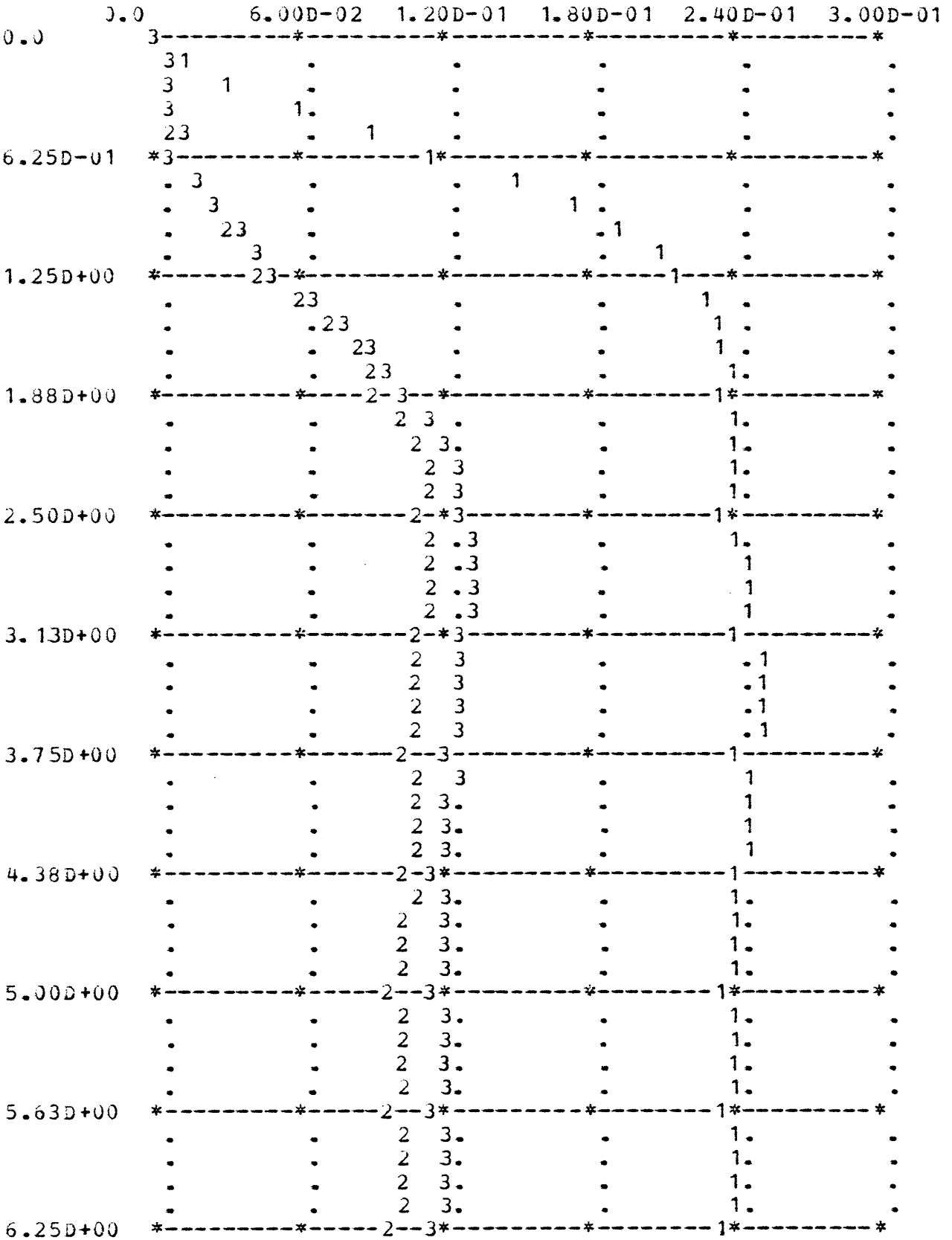
DELTA 1,2,3 VS. TIME
EPSILON= 0.50



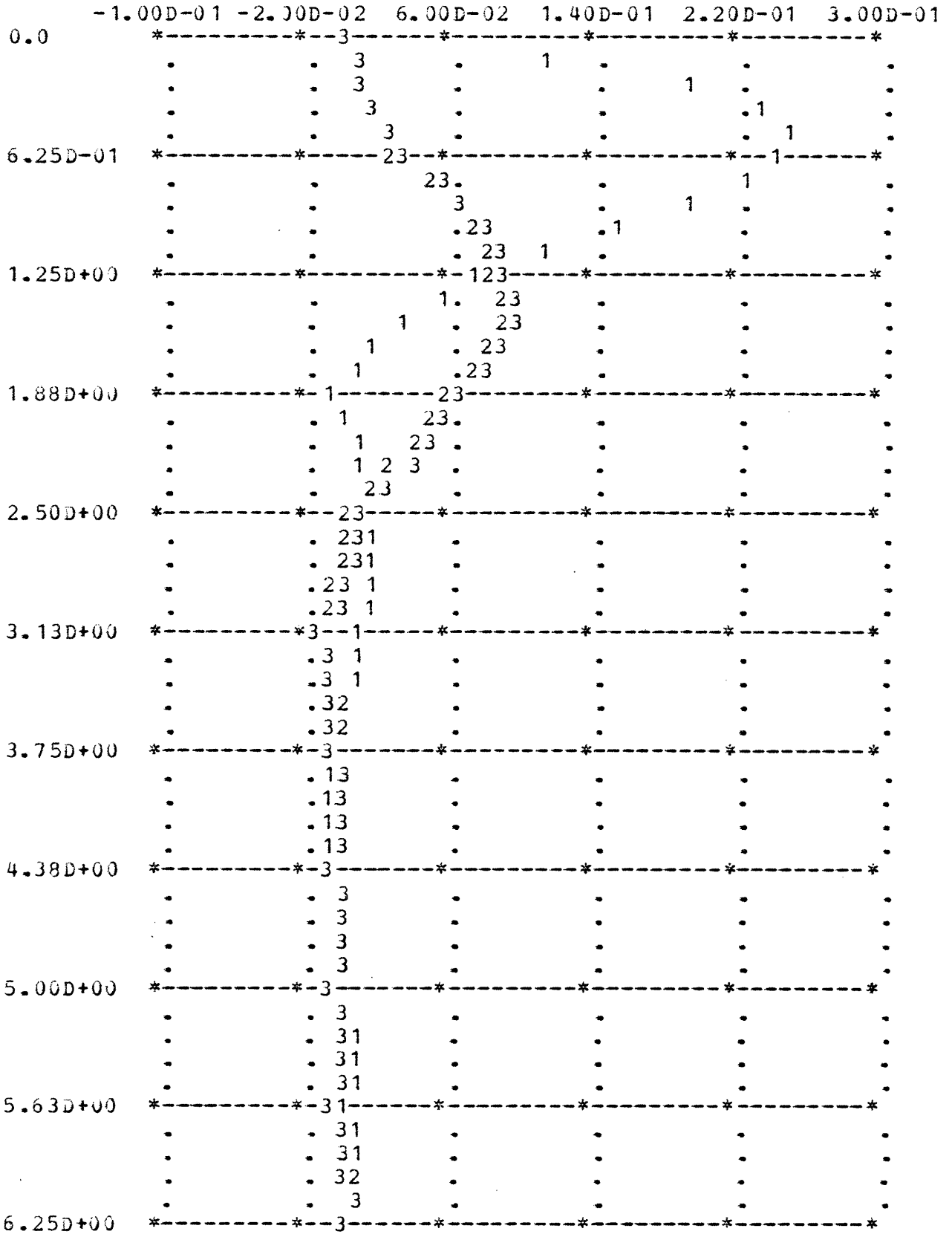
FREQ. 1,2,3 VS. TIME
EPSILON= 0.50



DELTA 1,2,3 VS. TIME
EPSILON= 0.25



FREQ. 1,2,3 VS. TIME
EPSILON= 0.25



4.4 Interconnected Single Input-Single Output Systems

To demonstrate the advantages of the weak coupling decomposition algorithm, consider a large scale system composed of interconnected single-input, single-output (SISO) subsystems. One of the common representations for these subsystems is the transfer function from which a companion form state space realization is easily constructed.

Suppose the transfer function is given by

$$T(s) = \frac{K(s^m + \alpha_m s^{m-1} + \ldots + \alpha_2 s + \alpha_1)}{s^n + a_n s^{n-1} + \ldots + a_2 s + a_1} = \frac{Y(s)}{U(s)} \quad .$$

This input-output relationship can be described in state space form as

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & & 1 & & \\ \vdots & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_1 & -a_2 & \cdots & & -a_n \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ K \end{bmatrix} u$$

companion form

$$y = [\alpha_1 \alpha_2 \ \cdots \ \alpha_m \ 1]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad .$$

With each subsystem in the above form, the overall large system matrix is of the form

$$A = \begin{bmatrix} A_{11} & A_{12} & & A_{1N} \\ A_{21} & A_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \vdots & & & \vdots \\ A_{N1} & \cdots & \cdots & A_{NN} \end{bmatrix}$$

where A_{ii} is in the above companion form and A_{ij}, $i \neq j$ is of the form

$$A_{ij} = \begin{bmatrix} 0 & 0 \cdots 0 \\ \vdots & 0 \cdots 0 \\ \vdots & \vdots \\ 0 & 0 \cdots 0 \\ c_s & c_{s+1} \cdots c_t \end{bmatrix} ; \quad s \text{ and } t \text{ integers } s \leq t$$

since the systems are interconnected only through their single controllable input.

The weak coupling decomposition algorithm for the Lyapunov equation is (in Jacobi form)

$$A_{ii}^T K_{ij}^{m+1} + K_{ij}^{m+1} A_{jj} + Q_{ij}^{m+1} = 0$$

$$Q_{ij}^{m+1} = \sum_{\substack{r=1 \\ r \neq i}}^{N} A_{ri}^T K_{rj}^m + \sum_{\substack{r=1 \\ r \neq j}}^{N} K_{ir}^m A_{rj} + Q_{ij} \quad .$$

Notice that only off diagonal block elements occur in the computation of Q_{ij}^{m+1} and therefore, each term of the sum

for a given r, is of the form

$$(C^T X + YD)_r = Z_r$$

where C and D are zero except for their last rows. The component equations can then be written as

$$(z_{ij})_r = \left(\sum_{p=1}^{\theta} c_{pi} x_{pj} + \sum_{p=1}^{\sigma} y_{ip} d_{pj}\right)_r$$

$$(z_{ij})_r = (c_{\theta i} x_{\theta j} + y_{i\sigma} d_{\sigma j})_r$$

θ, σ = values depending on the sizes of the subsystems interconnected .

Thus, only the last row of X and the last column of Y are needed to generate Z. Hence, in order to compute Q_{ij}^{m+1} only the last row and column of K_{ij}^m for all i and j is required.

In an algorithm given by Molinari [27], the last row and column of K_{ij}^m is solved for given Q_{ij}^m and the remainder of K_{ij}^m generated. The total operation count in evaluation of K_{ij}^m is 5 rs multiplications when K_{ij}^m is an r x s matrix.

Using this algorithm, the complete solution need only be generated at the last step in the decomposition algorithm. This allows all computations to be done with rows or columns of K_{ij}^m. Reducing the number of computations significantly,

however, is due to the use of Molinari's algorithm.

For N subsystems of dimension n we have:

evaluation of Q_{ij} $= 2n^2(N - 1)$ multiplications

evaluation of last row and column K_{ij} ; $i \neq j$

$= 3n^2$ multiplications

evaluation of last row and column $K_{ii} = \frac{3}{4} n^2$ multiplications.

For one iteration

$\frac{N(N + 1)}{2}$ Q_{ij}'s $\rightarrow n^2N(N^2 - 1)$ multiplications

$\frac{N(N - 1)}{2}$ K_{ij}'s ; $i \neq j \rightarrow \frac{3}{2} n^2N(N - 1)$ multiplications

N K_{ii}'s $\rightarrow \frac{3}{4} n^2N$ multiplications .

Hence, the total operation count after σ iterations is

$$(Nn)^2 + \frac{\sigma Nn^2}{4} [4N^2 + 6N - 7] < (nN)^2(2N\sigma + 1)$$

where the extra $(nN)^2$ comes from generating the full solution from the last row and column of the final iterate.

When the algorithm is of the Gauss Seidel type and relaxation techniques are applied similar results in operation counts can be attained with only a slightly higher number of operations.

Here it is important to note that the overall system is not in companion form, but only its interconnected subsystems.

Hence, algorithms that take advantage of special canonical forms are of interest even though the large scale system does not exhibit the exact form required for the algorithm.

A general system matrix may be transformed into companion form and Molinari's algorithm used to solve the Lyapunov equation in $5n^3$ multiplication where n is the dimension of the system. The existence of the companion form is guaranteed if the system is controllable for some input B matrix. However, the transformation to companion form is numerically unstable and the roundoff error is too great for even modest size systems.

Nevertheless, if the companion representation is already available, this algorithm appears particularly attractive. For multiple input-multiple output representations in standard controllable form [3] Molinari's algorithm may be adapted and used in conjunction with the decomposition algorithm. This is the case, since in standard controllable form the system is represented as

$$A = \begin{bmatrix} 0_m & I_m & 0 & \cdots & 0_m \\ 0_m & 0_m & I_m & & 0_m \\ \vdots & & \ddots & \ddots & \vdots \\ 0_m & 0 & \cdots & 0_m & I_m \\ -a_1 I_m & -a_2 I_m & \cdots & & -a_m I_m \end{bmatrix} \qquad B = \begin{bmatrix} 0_m \\ 0_m \\ 0_m \\ \vdots \\ 0_m \\ I_m \end{bmatrix}$$

$$C = [R_o R_1 \cdots R_{n-1}]$$

where

$$R(s) = C(Is - A)^{-1}B$$

$$p(s) = s^n + a_n s^{n-1} + \ldots + a_2 s + a_1$$

$$p(s)R(s) = (R_o + R_1 s + \ldots + R_{n-1} s^{n-1})$$

and $p(s)$ is the monic least common multiple of the denominators of the $r_{ij}(s)$.

For completeness, a brief sketch of Molinari's algorithm follows.

Molinari's Algorithm for Companion Form Matrices in the Solution of $A^T K + KB + Q = 0$.

Suppose

$$A = \left[\begin{array}{c|ccc} 0 & & I & \\ \hline -a_1 & -a_2 & -a_3 \ldots & -a_n \end{array}\right] ;$$

$$B = \left[\begin{array}{c|ccc} 0 & & I & \\ \hline -b_1 & -b_2 & -b_3 \ldots & -b_m \end{array}\right]$$

A n x n matrix

B m x m matrix

C n x m matrix .

Denote the characteristic polynomials of A and B by

$$\Delta_a(s) = s^n + a_n s^{n-1} + \ldots + a_1$$

$$\Delta_b(s) = s^m + b_m s^{m-1} + \ldots + b_1 \quad .$$

Writing the equation in component form

$$\sum_{r=1}^{n} a_{ri} k_{rj} + \sum_{r=1}^{m} k_{ir} b_{rj} = - q_{ij} \qquad \begin{matrix} i = 1,2,\ldots,n \\ j = 1,2,\ldots,m \end{matrix} \quad .$$

For $A = [a_{ij}]$; $B = [b_{ij}]$ we have

$$a_{ij} = \begin{cases} 1 & ; \ i \neq n \text{ and } j = i+1 \\ - a_j & ; \ i = n \\ 0 & ; \text{ otherwise} \end{cases}$$

$$b_{ij} = \begin{cases} 1 & ; \ i \neq m \text{ and } j = i + 1 \\ - b_j & ; \ i = m \\ 0 & ; \text{ otherwise} \end{cases} \quad .$$

Now

$$\sum_{r=1}^{n} a_{ri} k_{rj} = \begin{cases} k_{i-1,j} - a_i k_{nj} & ; \ i \neq 1 \\ - a_1 k_{nj} & ; \ i = 1 \end{cases}$$

$$\sum_{r=1}^{m} k_{ir} b_{rj} = \begin{cases} k_{i,j-1} - b_j k_{im} & ; \ j \neq 1 \\ - b_1 k_{im} & ; \ j = 1 \end{cases} \quad .$$

Defining $k_{ij} \triangleq 0$ if $i = 0$ or $j = 0$, then, the component form of the equation becomes

$$a_i k_{nj} + b_j k_{im} - k_{i-1,j} - k_{i,j-1} = q_{ij}$$

$$i = 1,2,\ldots,n$$

$$j = 1,2,\ldots,m \quad .$$

If all the component equations for which $i + j =$ constant are summed according to

$$\sum_{i=\sigma}^{\theta} (-1)^i (a_i k_{nj} + b_j k_{im} - k_{i-1,j} - k_{i,j-1}) = \sum_{i=\sigma}^{\theta} (-1)^i q_{ij}$$

where

$$\sigma = \begin{cases} 1 & ; \quad N \leq m \\ N + 1 - m & ; \quad N > m \end{cases}$$

$$\theta = \begin{cases} N & ; \quad N \leq n \\ n & ; \quad N > n \end{cases}$$

the equations simplify greatly.

To see this, consider the sum letting $p = i - 1$

$$g(N) = \sum_{i=\sigma}^{\theta} (-1)^i (k_{i-1,N+1-i} + k_{i,N-i}) = \sum_{p=\sigma-1}^{\theta-1} (-1)^{p+1} (k_{p,N-p})$$

$$+ \sum_{i=\sigma}^{\theta} (-1)^i (k_{i,N-i})$$

$$g(N) = (-1)^{\theta} k_{\theta,N-\theta} + (-1)^{\sigma} k_{\sigma-1,N-\sigma+1} \quad .$$

Then

$$g(N) = \begin{cases} 0 & ; N \leq m \text{ and } N \leq n \\ (-1)^{N+1-m} k_{N-m,m} & ; N > m \text{ and } N \leq n \\ (-1)^{n} k_{n,N-n} & ; N \leq m \text{ and } N > n \\ (-1)^{N+1-m} k_{N-m,m} + (-1)^{n} k_{n,N-n} & ; N > m \text{ and } N > n \end{cases} .$$

Define

$$h(N) = \begin{cases} \sum_{i=\sigma}^{\theta} (-1)^{i} q_{i,N+1-i} & ; N = 1,2,\ldots,m+n-1 \\ 0 & ; N = m+n \end{cases} .$$

The Sylvester equation may now be written as

$$\sum_{i=\sigma}^{\theta} (-1)^{i} (a_i k_{n,N+1-i} + b_{N+1-i} k_{im}) - g(N) = h(N) \quad .$$

This in turn may be written in matrix form by defining an $(n + m) \times n$ matrix $F(x,n,m)$

$$F(x,n,m) = \begin{bmatrix} x_1 & & & \\ x_2 & x_1 & & \\ \vdots & x_2 & \ddots & \\ \vdots & \vdots & & x_1 \\ x_{n+1} & \vdots & & x_2 \\ & x_{n+1} & & \\ & & & \\ & & & x_{n+1} \end{bmatrix}$$

with x an $n + 1$ component vector.

Now, the Sylvester equation is

$$[F(c,n,m) \vdots F(d,m,n)] \begin{bmatrix} \lambda \\ \cdots \\ \mu \end{bmatrix} = h$$

where

$$c^T = [-a_1, a_2, \ldots, (-1)^n a_n, (-1)^{n+1}]$$

$$d^T = [b_1, b_2, \ldots, b_m, 1]$$

$$\lambda^T = [k_{n1}, k_{n2}, \ldots, k_{nm}]$$

$$\mu^T = [-k_{1m}, k_{2m}, \ldots, (-1)^n k_{nm}]$$

$$h^T = [h(1), h(2), \ldots, h(n + m)] \quad .$$

This $(n + m)$-dimensional set of linear equations is solved in $3mn$ multiplications using a special algorithm that takes advantage of the structure of F. The remainder of K is generated from the last row and column using

$$k_{i,j-1} = q_{ij} + a_j k_{im} - B^T k_{ij} \; ; \; j = m, \ldots, 2$$

or

$$k_{j-1,i} = q_{ji} + b_j k_{ni} - A^T k_{ji} \; ; \; j = n, \ldots, 2 \quad .$$

4.5 Decomposition Algorithm for the Riccati Equation [4]

The Riccati equation like the Lyapunov equation, for the system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u$$

can be decomposed for a weakly coupled system. If A_{12} and A_{21} are "small", the solution of the Riccati equation should be "near" the solution of the Riccati equation from the block diagonal A made up of A_{11} and A_{22}. This suggests an iteration, similar to the previously discussed weak coupling algorithms for the Riccati equation.

Subproblem #1

$$A_{11}^T K_{11}^m + K_{11}^m A_{11} + Q_{11}^m - K_{11}^m B_1 R^{-1} B_1^T K_{11}^m = 0 \quad .$$

Subproblem #2

$$A_{22}^T K_{22}^m + K_{22}^m A_{22} + Q_{22}^m - K_{22}^m B_2 R^{-1} B_2^T K_{22}^m = 0 \quad .$$

Coordinator's Problem (Jacobi Iteration)

$$(A_{11} - B_1 R^{-1} B_1^T K_{11}^{m-1})^T K_{12}^m + K_{12}^m (A_{22} - B_2 R^{-1} B_2^T K_{22}^{m-1}) + Q_{12}^m = 0 \quad .$$

For a Jacobi type iteration

$$Q_{11}^m = Q_{11} + A_{21}^T K_{21}^{m-1} + K_{12}^{m-1} A_{21} - K_{12}^{m-1} B_2 R^{-1} B_2^T K_{21}^{m-1}$$

$$Q_{12}^{m} = A_{21}^{T} K_{22}^{m-1} + K_{11}^{m-1} A_{12}$$

$$Q_{22}^{m} = Q_{22} + A_{12}^{T} K_{12}^{m-1} + K_{21}^{m-1} A_{12} - K_{21}^{m-1} B_1 R^{-1} B_1^{T} K_{12}^{m-1}$$

must hold.

To change this to a Gauss-Seidel iteration, the coordinator's problem changes to

Coordinator's Problem (Gauss-Seidel):

$$(A_{11} - B_1 R^{-1} B_1^{T} K_{11}^{m})^{T} K_{12}^{m} + K_{12}^{m} (A_{22} - B_2 R^{-1} B_2^{T} K_{22}^{m}) + Q_{12}^{m} = 0$$

$$Q_{12}^{i} = A_{21}^{T} K_{22}^{m} + K_{11}^{m} A_{12}$$

and the subproblem remains the same, but must be solved before the coordination process. This is directly analogous to the Gauss-Seidel #2 iteration for the Lyapunov equation in the weak coupling section.

In a region sufficiently close to the solution K, the linearized equation governs the convergence rate of the algorithm, as in the two-time scale case.

CHAPTER V

CONCLUSIONS AND FUTURE RESEARCH

Two decompositions for the Lyapunov and Riccati equations have been presented. These decompositions make possible the solution of problems not feasible by standard methods, due to the large amount of storage required for manipulation of arrays during computation.

Using the decomposition algorithms, size of the arrays to be manipulated is reduced and computational savings can be attained when the number of iterations need be few or when the subsystem possess a special canonical form which may not be exhibited by the overall system. Finally, these algorithms recommend themselves where the regulator problem need be solved on-line,due to system parameter variations,by a number of small computers no one of which is able to solve the global problem.

Future Research

Much of the literature concerned with relaxation techniques has been with matrices that arise from partial differential equations that are discretized. These matrices form a very different class of matrices from the matrices that arise from state space descriptions of dynamical systems.

Hence, many of the theorems related to solving this class of large linear systems are inapplicable because they

depend on this type of special structure.

Specifically, the results on the selection of relaxation parameters is difficult to adapt and possibly other criteria such as a measure of the coupling might be substituted or selection of a relaxation factor might be done dynamically during the iteration. Also, other canonical forms besides the companion form might be used in decomposition methods to reduce computations.

APPENDIX

Two Time Scale Decomposition Algorithm for Riccati Equation

A.1 Computation of Q^m

Let

$$L = BR^{-1}B^T \tag{A.1.1}$$

$$\begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} = \begin{bmatrix} B_1R^{-1}B_1^T & B_1R^{-1}B_2^T \\ B_2R^{-1}B_1^T & B_2R^{-1}B_2^T \end{bmatrix} \tag{A.1.2}$$

$$(K_2^{m-1})^T A = \begin{bmatrix} 0 & 0 \\ K_{21}^{m-1} & 0 \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ K_{21}^{m-1}A_{11} & K_{21}^{m-1}A_{12} \end{bmatrix} \tag{A.1.3}$$

$$A^T K_2^{m-1} + (K_2^{m-1})^T A = \begin{bmatrix} 0 & A_{11}^T K_{12}^{m-1} \\ K_{21}^{m-1}A_{11} & A_{12}^T K_{12}^{m-1} + K_{21}^{m-1}A_{12} \end{bmatrix} \tag{A.1.4}$$

$$(K_2^{m-1})^T L K_2^{m-1} = \begin{bmatrix} 0 & 0 \\ K_{21}^{m-1} & 0 \end{bmatrix} \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} 0 & K_{12}^{m-1} \\ 0 & 0 \end{bmatrix} \tag{A.1.5}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & K_{21}^{m-1}L_{11}K_{12}^{m-1} \end{bmatrix}$$

$$(K_2^{m-1})^T L K_1^{m-1} = \begin{bmatrix} 0 & 0 \\ K_{21}^{m-1} & 0 \end{bmatrix} \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} K_{11}^{m-1} & 0 \\ K_{21}^{m-1} & K_{22}^{m-1} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ K_{21}^{m-1}L_{11} & K_{21}^{m-1}L_{12} \end{bmatrix} \begin{bmatrix} K_{11}^{m-1} & 0 \\ K_{21}^{m-1} & K_{22}^{m-1} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ K_{21}^{m-1}L_{11}K_{11}^{m-1} + K_{21}^{m-1}L_{12}K_{21}^{m-1} & K_{21}^{m-1}L_{12}K_{22}^{m-1} \end{bmatrix} \tag{A.1.6}$$

$$(K_1^{m-1})^T L K_2^{m-1} + (K_2^{m-1})^T L K_1^{m-1} =$$

$$\begin{bmatrix} 0 & K_{11}^{m-1}L_{11}K_{12}^{m-1} + K_{12}^{m-1}L_{21}K_{12}^{m-1} \\ K_{21}^{m-1}L_{11}K_{11}^{m-1} + K_{21}^{m-1}L_{12}K_{21}^{m-1} & K_{22}^{m-1}L_{21}K_{12}^{m-1} + K_{21}^{m-1}L_{12}K_{22}^{m-1} \end{bmatrix} \tag{A.1.7}$$

Since

$$Q^m = Q + A^T K_2^{m-1} + (K_2^{m-1})^T A - (K_2^{m-1})^T L K_2^{m-1} - (K_1^{m-1})^T L K_2^{m-1} - (K_2^{m-1})^T L K_1^{m-1} \tag{A.1.8}$$

we see that

$$\hat{Q}_{11}^m = Q_{11} \tag{A.1.9}$$

$$Q_{12}^m = Q_{12} + A_{11}^T K_{12}^{m-1} - K_{11}^{m-1} L_{11} K_{12}^{m-1} - K_{12}^{m-1} L_{21} K_{12}^{m-1} \tag{A.1.10}$$

$$Q_{22}^m = Q_{22}+A_{12}^T K_{12}^{m-1}+K_{21}^{m-1}A_{12}-K_{21}^{m-1}L_{11}K_{12}^{m-1}-K_{22}^{m-1}L_{21}K_{12}^{m-1}-K_{21}^{m-1}L_{12}K_{22}^{m-1} \tag{A.1.11}$$

Here, notice that $\tilde{Q}_{11}^m$ in (A.1.9) is different than $\hat{Q}_{11}^m$ because of the coupling between the slow subproblem and the coordination subproblem as yet to be taken care of, but (A.1.10) and (A.1.11) are in the desired form.

A.2 Computation of $A^T K_1^m+(K_1^m)^T A-(K_1^m)^T L K_1^m+Q^m = 0$

$$(K_1^m)^T A = \begin{bmatrix} K_{11}^m & K_{12}^m \\ 0 & K_{22}^m \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{A.2.1}$$

$$= \begin{bmatrix} K_{11}^m A_{11}+K_{12}^m A_{21} & K_{11}^m A_{12}+K_{12}^m A_{22} \\ K_{22}^m A_{21} & K_{22}^m A_{22} \end{bmatrix}$$

$$A^T K_1^m+(K_1^m)^T A =$$

$$\begin{bmatrix} A_{11}^T K_{11}^m+K_{11}^m A_{11}+A_{21}^T K_{21}^m+K_{12}^m A_{21} & A_{21}^T K_{22}^m+K_{11}^m A_{12}+K_{12}^m A_{22} \\ A_{12}^T K_{11}^m+A_{22}^T K_{21}^m+K_{22}^m A_{21} & A_{22}^T K_{22}^m+K_{22}^m A_{22} \end{bmatrix} \tag{A.2.2}$$

$$(K_1^m)^T L K_1^m =$$

$$\begin{bmatrix} K_{11}^m L_{11}+K_{12}^m L_{21} & K_{11}^m L_{12}+K_{12}^m L_{22} \\ K_{22}^m L_{21} & K_{22}^m L_{22} \end{bmatrix}\begin{bmatrix} K_{11}^m & 0 \\ K_{21}^m & K_{22}^m \end{bmatrix} \tag{A.2.3}$$

$$(K_1^m)^T L K_1^m =$$

$$\begin{bmatrix} K_{11}^m L_{11} K_{11}^m + K_{12}^m L_{21} K_{11}^m + K_{11}^m L_{12} K_{21}^m + K_{12}^m L_{22} K_{21}^m & K_{11}^m L_{12} K_{22}^m + K_{12}^m L_{22} K_{22}^m \\ K_{22}^m L_{21} K_{11}^m + K_{22}^m L_{22} K_{21}^m & K_{22}^m L_{22} K_{22}^m \end{bmatrix}$$

from this we can substitute in for L and obtain

$$K_{11}^m L_{11} K_{11}^m + K_{12} L_{21} K_{11}^m + K_{11}^m L_{12} K_{21}^m + K_{12}^m L_{22} K_{21}^m$$

$$= K_{11}^m (L_{11} K_{11}^m + L_{12} K_{21}^m) + K_{12}^m (L_{21} K_{11}^m + L_{22} K_{21}^m)$$

$$= K_{11}^m (B_1 R^{-1} B_1^T K_{11}^m + B_1 R^{-1} B_2^T K_{21}^m) + K_{12}^m (B_2 R^{-1} B_1^T K_{11}^m + B_2 R^{-1} B_2^T K_{21}^m)$$

$$= K_{11}^m B_1 R^{-1} (B_1^T K_{11}^m + B_2^T K_{21}^m) + K_{12}^m B_2 R^{-1} (B_1^T K_{11}^m + B_2^T K_{21}^m)$$

$$= (B_1^T K_{11}^m + B_2^T K_{21}^m)^T R^{-1} (B_1 K_{11}^m + B_2^T K_{21}^m) \qquad \text{(A.2.4)}$$

$$K_{11}^m L_{12} K_{22}^m + K_{12}^m L_{22} K_{22}^m = (K_{11}^m B_1 R^{-1} B_2^T + K_{12}^m B_2 R^{-1} B_2^T) K_{22}^m$$

$$= (B_1^T K_{11}^m + B_2^T K_{21}^m)^T R^{-1} (B_2^T K_{22}^m) \qquad \text{(A.2.5)}$$

$$K_{22}^m L_{22} K_{22}^m = K_{22}^m B_2 R^{-1} B_2^T K_{22}^m \qquad \text{(A.2.6)}$$

Let

$$G_1 = (B_1^T K_{11}^m + B_2^T K_{21}^m) \qquad \text{(A.2.7)}$$

$$G_2 = B_2^T K_{22}^m \qquad \text{(A.2.8)}$$

Now the algorithm may be written as

$$A_{11}^T K_{11}^m + K_{11}^m A_{11} + A_{21}^T K_{21}^m + K_{12}^m A_{21} + Q_{11}^m - G_1^T R^{-1} G_1 = 0 \quad \text{(A.2.9)}$$

$$A_{21}^T K_{22}^m + K_{11}^m A_{12} + K_{12}^m A_{22} + Q_{12}^m - G_1^T R^{-1} G_2 = 0 \quad \text{(A.2.10)}$$

$$A_{22}^T K_{22}^m + K_{22}^m A_{22} + Q_{22}^m - G_2^T R^{-1} G_2 = 0 \quad \text{(A.2.11)}$$

We have both (A.2.10) and (A.2.11) in the desired form only (A.2.9) is not as needed.

A.3 Slow Subproblem Derivation

Solving (A.2.10) for K_{12}^m

$$K_{12}^m = [G_1^T R^{-1} G_2 - A_{21}^T K_{22}^m - K_{11}^m A_{12} - Q_{12}^m] A_{22}^{-1} \quad \text{(A.3.1)}$$

$$K_{12}^m = X - K_{11}^m A_{12} A_{22}^{-1} \quad \text{(A.3.2)}$$

$$X = [G_1^T R^{-1} G_2 - A_{21}^T K_{22}^m - Q_{12}^m] A_{22}^{-1} \quad \text{(A.3.3)}$$

Looking at the terms of (A.2.9) we see

$$K_{12}^m A_{21} = X A_{21} - K_{11}^m A_{12} A_{22}^{-1} A_{21} \quad \text{(A.3.4)}$$

$$\begin{aligned} K_{11}^m A_{11} + K_{12}^m A_{21} &= K_{11}^m (A_{11} - A_{12} A_{22}^{-1} A_{21}) + X A_{21} \\ &= K_{11}^m \hat{A}_{11} + X A_{21} \end{aligned} \quad \text{(A.3.5)}$$

$$XA_{21} = (G_1^T R^{-1} G_2 - Q_{12}^m) A_{22}^{-1} A_{21} - A_{21}^T K_{22}^m A_{22}^{-1} A_{21} \qquad (A.3.6)$$

$$XA_{21} = Y - A_{21}^T K_{22}^m A_{22}^{-1} A_{21} \qquad (A.3.7)$$

$$Y = (G_1^T R^{-1} G_2 - Q_{12}^m) A_{22}^{-1} A_{21} \qquad (A.3.8)$$

$$A_{21}^T X + XA_{21} = Y + Y^T - (A_{22}^{-1} A_{21})^T (A_{22}^T K_{22}^m + K_{22}^m A_{22}) A_{22}^{-1} A_{21} \qquad (A.3.9)$$

Now substituting in (A.3.9) from (A.2.11) for $A_{22}^T K_{22}^m + K_{22}^m A_{22}$ we have

$$A_{21}^T X + XA_{21} = Y + Y^T - (A_{22}^{-1} A_{21})^T (G_2^T R^{-1} G_2 - Q_{22}^m)(A_{22}^{-1} A_{21}) \qquad (A.3.10)$$

Letting

$$A_o = A_{22}^{-1} A_{21} \qquad (A.3.11)$$

$$\begin{aligned} A_{21}^T X + XA_{21} = {} & A_o^T (G_2 R^{-1} G_1 - Q_{21}^m) + (G_1^T R^{-1} G - Q_{12}^m) A_o \\ & - A_o^T (G_2^T R^{-1} G_2 - Q_{22}^m) A_o \end{aligned} \qquad (A.3.12)$$

Now (A.2.9) may be written as

$$\hat{A}_{11}^T K_{11}^m + K_{11}^m \hat{A}_{11} + \hat{Q}_{11}^m - G_1^T R^{-1} G_1 - A_o^T G_2^T R^{-1} G_2 + A_o^T G_2^T R^{-1} G_1 + G_1^T R^{-1} G_2 A_o = 0 \qquad (A.3.13)$$

where

$$\hat{Q}_{11}^m = Q_{11} + A_o^T Q_{22}^m A_o - A_o^T Q_{21}^m - Q_{12}^m A_o \qquad (A.3.14)$$

which is the required form for $\hat{Q}_{11}^m$. Factoring the terms containing G_1 and G_2 in (A.3.13)

$$\hat{A}_{11}^T K_{11}^m + K_{11}^m \hat{A}_{11} + \hat{Q}_{11}^m - (G_1 - G_2 A_o)^T R^{-1} (G_1 - G_2 A_o) \qquad (A.3.15)$$

The last term in (A.3.15) is not yet in the form required. From (A.2.7), (A.2.8) and (A.3.11) we have

$$G_1 - G_2 A_o = B_1^T K_{11}^m + B_2^T K_{21}^m - B_2^T K_{22}^m A_{22}^{-1} A_{21} \qquad (A.3.16)$$

and from (A.3.1) we get

$$K_{21}^m = (A_{22}^{-1})^T [G_2^T R^{-1} G_1 - K_{22}^m A_{21} - A_{12}^T K_{11}^m - Q_{21}^m] \qquad (A.3.17)$$

Substituting (A.3.17) in (A.3.16) and collecting terms

$$G_1 - G_2 A_o = (B_1 - A_{12} A_{22}^{-1} B_2)^T K_{11}^m - B_2^T K_{22}^m A_{22}^{-1} A_{21} - (A_{22}^{-1} B_2)^T K_{22}^m A_{21} + (A_{22}^{-1} B_2)^T [G_2^T R^{-1} G_1 - Q_{21}^m] \qquad (A.3.18)$$

$$G_1 - G_2 A_o = \hat{B}_1^T K_{11}^m - B_2^T [(A_{22}^{-1})^T K_{22}^m + K_{22}^m A_{22}^{-1}] A_{21} + (A_{22}^{-1} B_2^T) [G_2^T R^{-1} G_1 - Q_{21}^m] \qquad (A.3.19)$$

$$= \hat{B}_1^T K_{11}^m - (A_{22}^{-1} B_2)^T [K_{22}^m A_{22} + A_{22}^T K_{22}^m] (A_{22}^{-1} A_{21}) + (A_{22}^{-1} B_2^T) [G_2^T R^{-1} G_1 - Q_{21}^m]$$

Substituting from (A.2.11) again

$$(G_1 - G_2 A_o) = \hat{B}_1^T K_{11}^m - (A_{22}^{-1} B_2)^T [(G_2^T R^{-1} G_2 - Q_{22}^i) A_{22}^{-1} A_{21} - (G_2^T R^{-1} G_1 - Q_{21}^m)] \qquad (A.3.20)$$

$$G_1-G_2A_o = \hat{B}_1^T K_{11}^m + (A_{22}^{-1}B_2)^T [G_2^T R^{-1}(G_1-G_2A_o) + Q_{22}^m A_o - Q_{21}^m] \quad (A.3.21)$$

$$[I-(A_{22}^{-1}B_2)^T G_2^T R^{-1}](G_1-G_2A_o) = \hat{B}^T K_{11}^m + (A_{22}^{-1}B_2)^T (Q_{22}^m A_o - Q_{21}^m) \quad (A.3.22)$$

$$G_1-G_2A_o = [I-(A_{22}^{-1}B_2)^T G_2^T R^{-1}]^{-1} [\hat{B}^T K_{11}^m + (A_{22}^{-1}B_2)^T (Q_{22}^m A_o - Q_{21}^m)] \quad (A.3.23)$$

Now

$$S^m = (A_o^T Q_{22}^m - Q_{12}^m)(A_{22}^{-1}B_2) \quad (A.3.24)$$

and

$$M = [I-(A_{22}^{-1}B_2)^T G_2^T R^{-1}]^{-1} \quad (A.3.25)$$

so that

$$(G_1-G_2A_o) = M(\hat{B}^T K_{11}^m + (S^m)^T) \quad (A.3.26)$$

For

$$M^{-1} = I-(A_{22}^{-1}B_2)^T K_{22}^m B_2 R^{-1} \quad (A.3.27)$$

we can substitute for K_{22}^m from (A.2.11) to get

$$M^{-1} = I-(A_{22}^{-1}B_2)^T [-A_{22}^T K_{22}^m - Q_{22}^m + G_2^T R^{-1} G_2] A_{22}^{-1} B_2 R^{-1}$$

$$M^{-1} = [R+(A_{22}^{-1}B_2)^T Q_{22}^m (A_{22}^{-1}B_2)]R^{-1} + [I-(A_{22}^{-1}B_2)^T G_2^T R^{-1}] G_2 A_{22}^{-1} B_2 R^{-1} \quad (A.3.28)$$

with

$$R^m = R+(A_{22}^{-1}B_2)^T Q_{22}^m (A_{22}^{-1}B_2) \quad (A.3.29)$$

$$M^{-1} = R^m R^{-1} + M^{-1} G_2 A_{22}^{-1} B_2 R^{-1} \quad (A.3.30)$$

from which

$$M^{-1} = R^m R^{-1}(I-G_2A_{22}^{-1}B_2R^{-1})^{-1} \qquad (A.3.31)$$

$$= R^m(R-G_2A_{22}^{-1}B_2)^{-1}$$

$$= R^m(I-R^{-1}G_2A_{22}^{-1}B_2)^{-1}R^{-1} = R^m M^T R^{-1}$$

or

$$M = R(M^{-1})^T(R^m)^{-1}$$

$$M^T R^{-1} M = (R^m)^{-1} \qquad (A.3.32)$$

Now from (A.3.26) we have

$$(G_1-G_2A_o)^T R^{-1}(G_1-G_2A_o) = (K_{11}^m\hat{B}_1+S^m)M^T R^{-1} M(K_{11}^m\hat{B}_1+S^m)^T \qquad (A.3.33)$$

and

$$(G_1-G_2A_o)^T R^{-1}(G_1-G_2A_o) = (K_{11}^m\hat{B}_1+S^m)(R^m)^{-1}(K_{11}^m\hat{B}_1+S^m)^T \qquad (A.3.34)$$

using (A.3.32) so that (A.3.15) becomes

$$\hat{A}_{11}^T K_{11}^m + K_{11}^m\hat{A}_{11} + \hat{Q}_{11}^m - (K_{11}^m\hat{B}_1+S^m)(R^m)^{-1}(K_{11}^m\hat{B}_1+S^m)^T = 0 \qquad (A.3.35)$$

the desired result.

REFERENCES

1. N.R. Sandell, Jr., P. Varaiya, and M. Athans, "A Survey of Decentralized Methods for Large Scale Systems", in Proc. Eng. Found. Conf., Sys. Eng. for Power: Status & Pros., 8-75.

2. M. Athans, "The Role and Use of the Stochastic Linear-Quadratic-Gaussian Problem in Control System Design", IEEE Trans. AC-16, Dec. 1971.

3. R.W. Brockett, Finite Dimensional Linear Systems, John Wiley and Sons, Inc., 1970.

4. N.R. Sandell, Jr., "Decomposition vs. Decentralization in Large-Scale System Theory", in Proc. Eng. Found. Conf., Sys. Eng. for Power: Status & Prospects, Aug. 1975.

5. A.J. Laub and F.N. Bailey, "An Iterative Coordination Approach to Decentralized Decision Problems", Systems Control Report No. 7702, Systems Control Group, University of Toronto, Canada, Feb. 1977.

6. P. Hagander, "Numerical Solution of $A^TS + SA + Q = 0$ ", Information Sciences, 4, 1972.

7. E.J. Davison and F.T. Man, "The Numerical Solution of $AQ + QA = - C$", IEEE Trans. AC-13, August, 1968.

8. R.A. Smith, "Matrix Equation $XA + BX = C$", SIAM Journal Appl. Math., 16, 1968.

9. D.L. Kleinman, "On An Iterative Technique for Riccati Equation Computations", IEEE Trans. AC-13, Feb. 1968.

10. N.R. Sandell, Jr., "On Newton's Method for Riccati Equation Solution", IEEE Trans. AC-19, June, 1974.

11. R. Bellman, Introduction to Matrix Analysis, McGraw-Hill, 1960.

12. W. Givens, "Elementary Divisors and Some Properties of the Lyapunov Mapping", Rep. ANL-6465, Argonne National Laboratory, Argonne, IL, 1961.

13. R.S. Varga, Matrix Iterative Analysis, Prentice Hall, 1962.

14. D.M. Young, Iterative Solution of Large Linear Systems, Academic Press, 1971.

15. G.W. Stewart, Introduction to Matrix Computations, Academic Press, 1973.

16. J.H. Wilkinson, Rounding Errors in Algebraic Processes, Prentice-Hall, 1963.

17. J.H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, 1965.

18. A.H. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell Publishing Co., 1965.

19. J.D. Cole, Perturbation Methods in Applied Mathematics, Blaisdell, 1968.

20. A.H. Haddad, "Linear Filtering of Singularly Perturbed Systems", IEEE Trans. AC-21, No. 4, 1976.

21. D. Teneketzis and N.R. Sandell, Jr., "Linear Regulators for Stochastic Systems by a Multiple Time Scales Method", IEEE Trans. AC-22, August, 1977.

22. A.H. Haddad and P.V. Kokotovic, "Stochastic Control of Linear Singularly Perturbed Systems", IEEE Trans. AC-22, October, 1977.

23. P.V. Kokotovic, "Feedback Design of Large Linear Systems", in J.B. Cruz, ed., Feedback Systems, McGraw-Hill, 1972.

24. J.H. Chow and P.V. Kokotivic, "Eigenvalue Placement in Two-Time Scale Systems", IFAC Symposium on Large Scale Systesm Theory and Applications, June 16-20, 1976, Udine, Italy.

25. T. Athay, Numerical Analysis of the Lyapunov Equation With Application to Interconnected Power Systems, Thesis for Engineers Degree, MIT, June, 1976.

26. Lee, Stephen, Transient Stability Equivalents for Power System Planning, EPSEL Report No. 38, MIT, June 1972.

27. B.P. Molinari, "Algegraic Solution of Matrix Linear Equations in Control Theory", Proc. IEEE, Vol. 116, No. 10, October, 1969.

28. M. Athans, P. Falb, Optimal Control, New York, McGraw-Hill, 1966.

29. D. Teneketzis, Perturbation Methods in Decentralized Stochastic Control, Master's Thesis, MIT, June 1976.

30. R.H. Bartels and G.W. Stewart, "Solution of the Matrix Equation AX + XB = C", Communications of the ACM, September, 1972, Vol. 15, No. 9.

31. P.V. Kokotovic, "A Riccati Equation for Block Diagonalization of Illconditioned Systems", IEEE Trans. AC-20 , December, 1975.

32. P.V. Kokotovic and R.A. Yackel, "Singular Perturbation of Linear Regulators: Basic Theorems", IEEE Trans. AC-17, No. 1, February, 1972.