

THE UNIVERSITY OF MANITOBA

SOME NEW METHODS FOR THE SOLUTION OF MATRIX EQUATIONS
ARISING FROM DISCRETIZED PARTIAL DIFFERENTIAL EQUATIONS

by

DESMOND JAMES WALTON

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

WINNIPEG, MANITOBA

February 1978.

SOME NEW METHODS FOR THE SOLUTION OF MATRIX EQUATIONS
ARISING FROM DISCRETIZED PARTIAL DIFFERENTIAL EQUATIONS

BY

DESMOND JAMES WALTON

A dissertation submitted to the Faculty of Graduate Studies of
the University of Manitoba in partial fulfillment of the requirements
of the degree of

DOCTOR OF PHILOSOPHY

© 1977

Permission has been granted to the LIBRARY OF THE UNIVER-
SITY OF MANITOBA to lend or sell copies of this dissertation, to
the NATIONAL LIBRARY OF CANADA to microfilm this
dissertation and to lend or sell copies of the film, and UNIVERSITY
MICROFILMS to publish an abstract of this dissertation.

The author reserves other publication rights, and neither the
dissertation nor extensive extracts from it may be printed or other-
wise reproduced without the author's written permission.



To my parents.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Professor Hoskins, who has been a tremendous driving force behind this thesis, for his most valuable advice, constant encouragement and very useful comments and suggestions throughout my work on this project. I also wish to thank Drs. Meek, Arscott and Ponzo for their comments and suggested improvements to the manuscript, and thanks also to everyone else who have aided me in the preparation of this disertation.

ABSTRACT

The matrix equation

$$XA + AY = F$$

occurs frequently in physical and engineering applications, for example, in the discretization of some partial differential equations, and if $X = Y'$, it is known as the Lyapunov matrix equation which has a variety of applications in control theory. Some new iterative methods for the solution of this matrix equation are presented. The methods are of second or higher order convergence. Corresponding algorithms are investigated with respect to implementation considerations, storage requirements, computation time, numerical stability and regions of convergence. Theoretical and numerical comparisons are made with existing methods. It is also illustrated how these methods can be used in the numerical solution of some partial differential equations.

TABLE OF CONTENTS

	page
1. Introduction	6
1.1 Notation and conventions	6
1.2 Applications in which the matrix equation	
$XA + AY = F$ arises	7
1.2.1 Matrix differential equations	8
1.2.2 Partial differential equations	9
1.2.3 The construction of Luengerberger observers	9
1.2.4 The analysis of beam gridworks	10
1.2.5 Continuous-time dynamic systems	10
1.2.6 Computation of covariance matrices in	
filtering theory	11
1.3 Existing methods for the solution of the matrix	
equation $XA + AY = F$	11
1.3.1 Composite matrix form	12
1.3.2 Transformation to diagonal form	13
1.3.3 Transformation to Frobenius form	14
1.3.4 Transformation to Hessenberg form	16
1.3.5 Series expansion methods	17
1.3.6 Methods which use numerical integration	
techniques	25
1.3.7 Reduction to Schwarz form	26
1.3.8 Kronecker product method	27
1.3.9 The method of Beavers and Denman	28

2. New quadratically convergent iterative methods for the solution of the matrix equation $XA + AY = F$	30
2.1 The numerical solution of the matrix equation $XA + AX = F$ when X is a real symmetric positive definite matrix	30
2.1.1 An improved method for solving the matrix equation $XA + AX = F$	35
2.1.2 A method with two acceleration parameters	38
2.1.3 An alternative choice of acceleration parameters	43
2.1.4 The case when all the eigenvalues of X are greater than unity	47
2.2 The numerical solution of the matrix equation $XA + AX = F$ when X is not real or symmetric	48
2.2.1 The solution of $XA + AX = F$ when X has a full set of eigenvectors	53
2.2.2 The solution of $XA + AX = F$ when X is defective	55
2.3 The solution of $XA + AY = F$	59
2.3.1 The case when X and Y have different rates of convergence	61
3. Numerical properties of the new methods and comparisons with recently published algorithms	63
3.1 Storage organization	63
3.2 Operation count	64
3.3 Numerical examples	65

3.4 Advantages of the new algorithms	83
3.4.1 Ease of implementation	83
3.4.2 Parallel processing	83
3.4.3 Rapid convergence	83
3.4.4 Numerical stability	84
3.4.5 Versatility	84
3.5 Comparison with existing methods	85
3.5.1 The method of Beavers and Denman	85
3.5.2 The methods of Davison and Man	86
3.5.3 The method of Bartels and Stewart	87
3.6 Economizing on implementation	89
3.7 The case when X and Y are ill-conditioned	90
4. Alternative methods for the solution of the matrix equation $XA + AY = F$	92
4.1 A higher order method for choosing acceleration parameters	92
4.2 Generalization of iterative methods to higher orders	96
4.2.1 Cubic convergence	97
4.2.2 Quartic convergence	98
4.2.3 Convergence of order m	98
4.3 An alternative generalization of iterative methods to higher orders	99
4.4 Regions of convergence	101
4.5 Higher order matrix iterative methods	107
4.6 Extension of the region of convergence	109

4.7 Implementation considerations for higher order methods	113
4.7.1 Computation time	116
4.7.2 The method of Man	118
4.7.3 Hybrid methods	119
5. The numerical solution of some partial differential equations	124
5.1 The numerical solution of the equation	
$\phi u_{xx} + \theta u_{yy} + \psi u_x + \gamma u_y + (\xi + \eta)u = f$	125
5.1.1 A five point formula	127
5.1.2 Taking advantage of symmetry	131
5.1.3 The five point formula on a uniform mesh	132
5.1.4 A nine point formula	134
5.1.5 An alternative nine point formula	137
5.1.6 The solution of the equation	
$\phi u_{xx} + \theta u_{yy} + \psi u_x + \gamma u_y + (\xi + \eta)u = f$	142
5.2 The numerical solution of the equation	
$\phi u_{xxxx} + (\phi + \theta)u_{xxyy} + \theta u_{yyyy} = f$	145
5.2.1 A thirteen point formula	145
5.2.2 A twenty-five point formula	148
5.3 Partial differential equations on L-shaped regions	150
5.3.1 The solution of the equation	
$\phi u_{xx} + \theta u_{yy} = f$ on an L-shaped region	152
5.4 The numerical solution of $L_x u + L_y u + L_z u = f$	157
5.4.1 Development of the matrix equation	159

5.5 The Robbins problem	167
6. Final Remarks	171
6.1 Comparison of the new methods with some existing methods for the solution of Poisson's equation	171
6.1.1 The successive overrelaxation method	173
6.1.2 Alternating-direction implicit (ADI) methods	176
6.1.3 Block direct methods	178
6.1.4 Cyclic reduction	180
6.1.5 Other methods	182
6.2 Conclusions	183

CHAPTER 1

INTRODUCTION

The objectives of this thesis are to present some new methods for the solution of the matrix equation

$$(1.1) \quad XA + AY = F,$$

and to indicate how they may be of use in physical and engineering applications. These new methods are efficient, easily implemented as algorithms and compare favourably with methods which are discussed in the recent literature.

Applications of and existing methods for the solution of equation (1.1) are discussed in this chapter. In chapter 2 the new methods are developed and in chapter 3 they are compared with recently published methods. The methods developed in chapter 2 are iterative with quadratic convergence and in chapter 4 they are generalized to obtain methods with higher order convergence. Chapter 5 illustrates how the new methods can be used for solving some partial differential equations which occur in Physics and Engineering and in conclusion chapter 6 summarizes the advantages of the new methods.

1.1 Notation and conventions.

Unless otherwise indicated, the following notation and conventions are assumed.

Scalar variables are denoted by lowercase Roman or Greek

letters, e.g. $a, b, x, y, \alpha, \beta$.

Column vectors are denoted by underscored lowercase Roman or Greek letters, e.g. $\underline{v}, \underline{w}, \underline{z}, \underline{\gamma}, \underline{\delta}$.

Matrices are denoted by capital Roman or Greek letters, e.g. G, F, U, ϕ, ψ .

A prime is used to denote the transpose of a vector or matrix, e.g. $\underline{\alpha}', \underline{v}', \underline{z}', \underline{G}', \underline{\theta}'$.

An overbar is used to denote the complex conjugate of a scalar, vector or matrix, e.g. $\bar{a}, \bar{\beta}, \bar{v}, \bar{F}$. An asterisk is used to denote the complex conjugate transposed of a vector or a matrix, e.g. \underline{x}^*, ϕ^* . The elements of a row vector, \underline{r}' , are usually indicated as

$$\underline{r}' = (r_1, r_2, \dots).$$

The elements of a column vector, \underline{c} , are usually indicated as

$$\underline{c} = (c_1, c_2, \dots)'$$

The elements of a matrix, A , are usually indicated as

$$A = (a_{i,j}).$$

The value of a function, $f(x,y)$, evaluated at a point (x_i, y_j) is denoted as $f_{i,j}$.

1.2 Applications in which the matrix equation $\underline{XA} + \underline{AY} = \underline{F}$ arises.

In various situations the solution of the bilinear matrix equation (1.1) is of interest, where X, Y, F are

known matrices of orders $m \times m$, $n \times n$, and $m \times n$ respectively (Kreisselmeier [39]). This matrix equation occurs, for instance, in the solution of matrix differential equations (Davison [20]), in the numerical solution of certain boundary value problems in partial differential equations (Bickley and McNamee [11]), in the construction of Luenberger observers (Luenberger [42]) and in the analysis of beam gridworks (Ma [43]).

If $X=Y'$ then equation (1.1) becomes

$$(1.2.1) \quad Y'A + AY = F,$$

which is the Lyapunov matrix equation and has applications in the stability analysis of continuous-time dynamic systems (Kalman and Bertram [35]) and in the computation of the covariance matrix in filtering theory (Kalman and Bucy [36]).

1.2.1 Matrix differential equations.

The matrix differential equation

$$(1.2.2) \quad \dot{Z}(t) = XZ + ZY - F, \quad Z(0) = C,$$

is encountered in the computation of state feedback control gains as obtained by the receding horizon method (Quintana and Fuenzalida [54]). It is easily verified by direct substitution that the solution of (1.2.2) is given by

$$Z(t) = e^{Xt} (C-A) e^{Yt} + A,$$

where A satisfies the matrix equation (1.1) (Davison [20]).

1.2.2 Partial differential equations.

Partial differential equations of the form

$$(1.2.3) \quad \frac{\partial^2}{\partial x^2} u(x,y) + \frac{\partial^2}{\partial y^2} u(x,y) = f(x,y),$$

where $u(x,y)$ is known on the boundary of a given region, have many applications in physical and engineering problems (Myint-U [51] pages 201-202). Bickley and McNamee [11] show how finite difference discretization of equation (1.2.3) leads to the matrix equation (1.1). The numerical solution of some partial differential equations are discussed in further detail in chapter 5.

1.2.3 The construction of Luenberger observers.

Assume that a plant has a single output v where

$$\dot{\underline{w}} = \underline{Y}\underline{w},$$

$$v = \underline{c}'\underline{w},$$

and that the corresponding observer is driven by v as its only input, then

$$\dot{\underline{z}} = -\underline{X}\underline{z} + \underline{b}v,$$

or

$$\dot{\underline{z}} = -\underline{X}\underline{z} + \underline{bc}'\underline{w}.$$

Under these conditions

$$\underline{z} = \underline{A}\underline{w},$$

where \underline{A} satisfies (Luenberger [42])

$$\underline{X}\underline{A} + \underline{A}\underline{Y} = \underline{bc}'.$$

which is the same as (1.1) with $F = \underline{bc}'$.

1.2.4 The analysis of beam gridworks.

In analyzing beam gridworks with various boundary conditions, the governing matrix equations all lead to the form of equation (1.1) (cf. Ma [43]). The matrices X and Y may be singular in some cases. The general solution of (1.1) consists of the complementary solution and a particular solution (Gantmacher [29] page 225, MacDuffee [44] page 92). Physically the complementary solution represents the initial internal forces of an unloaded gridwork such as may be caused by initial crookedness of the beams, uneven temperature distribution, etc. The particular solution represents the internal forces acting on the individual beams due to external loading.

1.2.5 Continuous-time dynamic systems.

Some continuous-time dynamic systems can be described by the n -th order linear system of ordinary differential equations (Kalman and Bertram [35])

$$(1.2.4) \quad \dot{\underline{z}} = \underline{Yz} .$$

The stability analysis of the system of differential equations (1.2.4) by Lyapunov's direct method leads to the matrix equation (1.2.1) (Barnett and Storey [5]).

1.2.6 Computation of covariance matrices in filtering theory.

The covariance matrix, A , for the message process in linear filtering and prediction theory satisfies the differential equation (Kalman and Bucy [36])

$$(1.2.5) \quad \frac{dA}{dt} = Y(t)'A + AY(t),$$

for a free system where Y is a matrix which represents the dynamics of the system. For a constant system, equation (1.2.5) is the same as equation (1.2.1) with $P = dA/dt$.

1.3 Existing methods for the solution of the matrix equation $XA + AY = P$.

It is well-known that the matrix equation (1.1) has a unique solution if and only if X and $-Y$ have no eigenvalues in common (Bellman [10] page 239, Bickley and McNamee [11], Kucera [40], Rutherford [55], Smith [56]). This condition is satisfied if all the eigenvalues of X and Y have negative real parts. Such matrices are referred to in the literature as being stable (Barnett and Storey [5], Davison [20]) or Hurwitzian (Smith [56]) and frequently arise in applications. If equation (1.1) is multiplied through by -1 , then the preceding observations apply if all the eigenvalues of the matrices X and Y have positive real parts. The methods which follow are applicable if either X and Y or both $-X$ and $-Y$ are stable matrices.

Older methods for solving equations (1.1) and (1.2.1) include re-writing the matrix equations in composite matrix form and diagonalization of the matrices X and Y (Bickley and McNamee [11]). Later methods include transformation of the matrices X and Y to Frobenius or Hessenberg form (Kreisselmeier [39]), series expansions (Barnett and Storey [4]), using numerical integration techniques (Davison and Man [21]), reduction to Schwarz form (Power [53]) and using a Kronecker or tensor product (Barnett and Storey [5]).

1.3.1 Composite matrix form.

The matrix equation (1.1) can be written in composite or block form as (see Bickley and McNamee [11])

$$(1.3.1) \quad Z\underline{b} = \underline{c},$$

where

$$Z = \begin{pmatrix} X+y_{1,1} I & y_{2,1} I & \cdot & \cdot & \cdot & y_{n,1} I \\ y_{1,2} I & X+y_{2,2} I & \cdot & \cdot & \cdot & y_{n,2} I \\ \cdot & & & & & \\ \cdot & & & & & \\ y_{1,n} I & y_{2,n} I & \cdot & \cdot & \cdot & X+y_{n,n} I \end{pmatrix},$$

$$\underline{b} = (a_{1,1}, a_{2,1}, a_{3,1}, \dots, a_{m,1}, a_{1,2}, \dots, a_{m,n})'$$

$$\underline{c} = (f_{1,1}, f_{2,1}, f_{3,1}, \dots, f_{m,1}, f_{1,2}, \dots, f_{m,n})'$$

The entries of the solution matrix A can thus be found by solving an mn-th order system of linear equations, however due to sparseness of the matrix Z it is usually not

desirable to solve the system of equations (1.3.1).

Gaussian elimination (see Wilkinson [62], page 200) destroys many of the zero entries of Z and is unsatisfactory (Dorr [23]). Equation (1.3.1) can also be solved by a costly block elimination method (Varga [60], pages 196-197) which for large m and n is also unsatisfactory (Dorr [23]). Iterative methods for solving equation (1.3.1) include successive overrelaxation and alternating direction implicit methods (Mitchell [49], pages 68 and 103; Varga [60], pages 56-61 and 209-249), both of which can be very costly in terms of computation time when X and Y are not of a very special form.

For equation (1.2.1) MacFarlane [45] develops a system of equations such as (1.3.1) but of order $n(n+1)/2$ and Barnett and Storey [2] further reduce the order to $n(n-1)/2$ by introducing a skew-symmetric matrix.

1.3.2 Transformation to diagonal form.

If X and Y are non-defective matrices then matrices R and S can be found such that (Wilkinson [62], page 42; Bickley and McNamee [11])

$$RXR^{-1} = \begin{pmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \lambda_m \end{pmatrix} = \Lambda$$

$$S^{-1}YS = \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_n \end{pmatrix} = M.$$

Equation (1.1) can be written as

$$XR^{-1}RA + ASS^{-1}Y = F,$$

or, pre-multiplying by R and post-multiplying by S

$$RXR^{-1}RAS + RASS^{-1}YS = RFS,$$

$$(1.3.2) \quad \Lambda \bar{A} + \bar{A}M = \bar{F},$$

where

$$\bar{A} = RAS,$$

$$\bar{F} = RFS,$$

The solution of (1.3.2) is given by

$$\bar{a}_{i,j} = \bar{F}_{i,j} / (\lambda_i + \mu_j), \quad i=1(1)n, j=1(1)n,$$

and A can be found from

$$A = R^{-1}\bar{A}S^{-1}.$$

The method can however be unstable (Bartels and Stewart [6]) and the matrices R and S are not always easily found.

1.3.3 Transformation to Frobenius form.

Consider a similarity transformation of Y to some special form

$$Z = T^{-1}YT.$$

Substituting

$$Y = TZT^{-1},$$

in (1.1) and multiplying (1.1) by T on the right yields

$$(1.3.3) \quad XB + BZ = G,$$

where

$$B = AT,$$

and

$$G = FT.$$

Let Z be the Frobenius representation of Y then

$$\begin{aligned} X(\underline{b}_1, \underline{b}_2, \dots, \underline{b}_n) + (\underline{b}_1, \underline{b}_2, \dots, \underline{b}_n)Z \\ = (\underline{q}_1, \underline{q}_2, \dots, \underline{q}_n), \end{aligned}$$

where \underline{b}_i and \underline{q}_i are the column vectors of the corresponding matrix and

$$Z = \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & & & & \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 \\ -z_0 & -z_1 & -z_2 & \cdot & \cdot & \cdot & -z_{n-1} \end{pmatrix}.$$

Define

$$\begin{aligned} C &= z_0 I - z_1 X + z_2 X^2 - \dots + z_{n-1} (-X)^{n-1} + (-X)^n, \\ \underline{d} &= \underline{q}_1 - X\underline{q}_2 + X^2\underline{q}_3 - \dots + (-X)^{n-1} \underline{q}_n, \end{aligned}$$

then the solution of equation (1.3.3) is given by (Kreisselmeier [39])

$$\underline{b}_n = -C^{-1} \underline{d},$$

$$\underline{b}_{i-1} = -X\underline{b}_i + z_{i-1} \underline{b}_n + \underline{q}_i, \quad i = n(-1)2,$$

and A can be found from

$$A = BT^{-1}.$$

Kreisselmeier [39] mentions that there are numerical difficulties associated with the computation of the Frobenius form (Wilkinson [62], pages 408-409).

1.3.4 Transformation to Hessenberg form.

Let the matrix Z in equation (1.3.3) be of lower Hessenberg form then (1.3.3) can be represented as

$$\begin{aligned} X(\underline{b}_1, \underline{b}_2, \dots, \underline{b}_n) + (\underline{b}_1, \underline{b}_2, \dots, \underline{b}_n)Z \\ = (\underline{q}_1, \underline{q}_2, \dots, \underline{q}_n), \end{aligned}$$

where

$$Z = \begin{pmatrix} z_{1,1} & z_{1,2} & 0 & 0 & \cdot & \cdot & 0 \\ z_{2,1} & z_{2,2} & z_{2,3} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ z_{n-1,1} & z_{n-1,2} & z_{n-1,3} & z_{n-1,4} & \cdot & \cdot & z_{n-1,n} \\ z_{n,1} & z_{n,2} & z_{n,3} & z_{n,4} & \cdot & \cdot & z_{n,n} \end{pmatrix}$$

Assuming that

$$z_{i,i+1} \neq 0, \quad i=1(1)n-1,$$

Kreisselmeier [39] shows that the solution of equation (1.3.3) is given by

$$\underline{b}_n = -C^{-1}\underline{d},$$

$$(1.3.4) \quad \underline{b}_{i-1} = -(\underline{X}\underline{b}_i + \sum_{j=i}^n z_{j,i} \underline{b}_j - \underline{q}_i) / z_{i-1,i}, \quad i=n(-1)2,$$

with

$$z_{0,1} = 1,$$

where C and \underline{d} can be calculated using the recursive algorithm (1.3.4), which generates

$$\underline{b}_0 = C\underline{b}_n + \underline{d} \quad \text{for } i=1.$$

and

$$\underline{b}_0 = \underline{d}$$

is obtained on substitution of $\underline{b}_n = 0$.

Bartels and Stewart [6] have implemented an algorithm for the solution of equation (1.1). The matrix Y is reduced to upper Hessenberg form by Householder's method (Wilkinson [62], page 347), and the upper Hessenberg matrix is in turn reduced to real Schur form by the QR algorithm (Martin, Peters and Wilkinson [48]).

1.3.5 Series expansion methods.

1.3.5.1 Infinite series expansion.

Let

$$W = (I+Y^e)(I-Y^e)^{-1}.$$

Then if Y is a stable matrix,

$$\lim_{n \rightarrow \infty} W^n = 0,$$

and equation (1.2.1) is transformed (Tausky [59]) into

$$(1.3.5) \quad A - WAW^e = M,$$

where

$$M = -(W+I)F(W^e+I)/2.$$