The improvement of the reduced coordinate iterative procedure for simultaneous linear equations
by Timothy Mark Hodges

A thesis submitted in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE
in Mechanical Engineering
Montana State University

Abstract:
An iterative method for solving simultaneous linear equations is considered. Previously, the trial solution generation routine depended on knowledge of the physical system. A method is introduced to generate the trial solutions based on the system of equations. The new method is used to solve several differential equations with various boundary conditions, so as to show that the method is independent from the physical system. Examples of two steady-state problems and a transient problem, with 1000 unknowns, demonstrate the performance as compared to existing techniques. Results indicate that the method is superior in cases where the boundary conditions are more complicated, and the method is comparable in all other cases.

THE IMPROVEMENT OF THE REDUCED COORDINATE ITERATIVE PROCEDURE FOR SIMULTANEOUS LINEAR EQUATIONS

by

Timothy Mark Hodges

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Mechanical Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

December 1982

APPROVAL

of a thesis submitted by

Timothy Mark Hodges

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dec 27 1982
Date

Dennis O. Blackhurst
Chairperson, Graduate Committee

Approved for the Major Department

Dec 27 1982
Date

Dennis O. Blackhurst
Head, Major Department

Approved for the College of Graduate Studies

12-28-82
Date

Michael P. Malone
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in his/her absence, by the Director of Libraries when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature Timothy Mark Hodges

Date Dec 27, 1982

ACKNOWLEDGEMENTS

The author wishes to thank Dr. D. O. Blackketter and Dr. R. O. Warrington for their help and guidance in the performance of this study. The author also wishes to thank Dr. R. J. Conant for all his help in advising over the past year. And finally thanks go to Lisa for all her support and understanding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

ABSTRACT

An iterative method for solving simultaneous linear equations is considered. Previously, the trial solution generation routine depended on knowledge of the physical system. A method is introduced to generate the trial solutions based on the system of equations. The new method is used to solve several differential equations with various boundary conditions, so as to show that the method is independent from the physical system. Examples of two steady-state problems and a transient problem, with 1000 unknowns, demonstrate the performance as compared to existing techniques. Results indicate that the method is superior in cases where the boundary conditions are more complicated, and the method is comparable in all other cases.

CHAPTER I

INTRODUCTION

Many numerical methods have been developed to accurately solve a large system of simultaneous equations. A method is presented here that has the possibility of being faster and more accurate than these methods.

The Reduced Coordinate Iterative Procedure (RCIP)[1] uses a linear combination of trial solutions, with each trial solution having an unknown weighting coefficient. Each weighting coefficient is determined by minimizing an error function. This minimizing process produces a smaller set of equations which can be solved directly for these weighting coefficients. Therefore, in solving a large set of equations, the RCIP method reduces the number of equations to be solved.

The method of least squares can be related to the RCIP method, because of the minimization of the squared residual in the RCIP method.[3] The least squares method finds the equation of a curve that passes through scattered points. This curve is found so that the sum of the square distances from each point to the curve is a minimum. In the RCIP method the sum of the squares of the residuals from the equation set, the variance, is minimized. The least squares method is used for curve fitting and not for the solution of simultaneous equations as RCIP is.

Another method that RCIP is related to is the Conjugate Direction Method [4,8]. The RCIP and Conjugate Direction methods both use the minimization of an error function in the solution of the problem. The Conjugate Direction Method minimizes the error along orthogonal directions. The Conjugate Direction Method will converge in less iterations than there are unknowns. The methods have been shown to be different [1].

To evaluate the performance of the RCIP method, two existing techniques were used: The first was a Gauss-Seidel iterative scheme using successive over-relaxation (SOR) [5], and the second was the Alternating Direction Explicit Procedure (ADEP)[6]. Both are excellent methods for comparation, because of their speed of convergence and their accuracy.

The RCIP method previously has had a systematic procedure for the generation of trial solutions. However, the trial solution generation was dependent on knowledge of the physical system which the equation represented. The trial solution generation should be independent of the knowledge of the physical system.

It is the purpose of this paper to present a trial solution generation scheme that is independent of the knowledge of the physical system while continuing to produce rapid convergence and be competitive with existing methods.

CHAPTER II

DESCRIPTION OF THE METHOD

The RCIP method uses a linear combination of trial solution vectors with each trial solution having an unknown scalar weighting coefficient. The weighting coefficients are obtained by solving a reduced set of equations which are obtained by minimizing the variance. The reduced set is solved using an exact or direct method. The trial solutions are different at each iteration and the iterative process continues until the desired convergence is reached [2]. The formulation of RCIP proceeds as follows:

Let

$$AX=F \tag{1}$$

be a system of n equations in n unknowns. Initially choose m vectors (m is the number of trial solutions, $G(k,m)_0$ n in length, the subscript 0 indicates the first iteration)

$$G(k,1)_0, G(k,2)_0, \ldots, G(k,m)_0, \qquad k=1,\ldots,n \tag{2}$$

where $m < n$, and set

$$X_0 = \sum_{j=1}^{m} \alpha_j \, G(k,j)_0 \qquad k=1,\ldots,n \tag{3}$$

where the scalars $(\alpha_1, \alpha_2, \ldots, \alpha_m)$ are to be determined.

Let

$$V_0(\alpha_1,\ldots,\alpha_m)=\langle AX_0-F, AX_0-F\rangle=||AX_0-F||^2, \tag{4}$$

be the inner product and the square of the norm of the residual ($R_0=AX_0-F$), respectively. Find the solution $(\alpha_{0,1},\ldots,\alpha_{0,m})$ to the system

$$\frac{\partial V_0}{\partial \alpha_p} = 0 \qquad p=1,\ldots,m. \tag{5}$$

Designate

$$\bar{V}_0=V(\alpha_{0,1},\ldots,\alpha_{0,m}) \tag{6}$$

and let

$$X(1,1)=\sum_{j=1}^{m} \alpha_{0,j}\ G(k,j)_0 \qquad k=1,\ldots,n. \tag{7}$$

Then

$$V_1(\alpha_1,\ldots,\alpha_m)=||AX_1-F||^2 \tag{8}$$

solve the system

$$\frac{\partial V_1}{\partial \alpha_p} = 0 \qquad p=1,\ldots,m\ , \tag{9}$$

and if $(\alpha_{1,1},\ldots,\alpha_{1,m})$ is the solution, designate

$$\bar{V}_1=V_1(\alpha_{1,1},\ldots,\alpha_{1,m}) \tag{10}$$

and let

$$X(2,1)=\sum_{j=1}^{m} \alpha_{1,j}\ G(k,j)_1 \qquad k=1,\ldots,n\ . \tag{11}$$

At the ith stage,

$$X(i,1)=\sum_{j=1}^{m} \alpha_{i-1,j}\ G(k,j)_{i-1} \qquad k=1,\ldots,n\ . \tag{12}$$

Select vectors $G(k,2)_i,\ldots,G(k,m)_i$, and let

$$X_i = \sum_{j=1}^{m} \alpha_j\ G(k,j)_i \qquad k=1,\ldots,n\ . \tag{13}$$

Then

$$V_i(\alpha_1,\ldots,\alpha_m)=||AX_i-F||^2 \quad . \qquad (14)$$

Again solve the system

$$\frac{\partial V_i}{\partial \alpha_p} = 0 \qquad p=1,..,m. \qquad (15)$$

If this solution is $(\alpha_{i,1},\ldots,\alpha_{i,m})$, then

$$\bar{V}_i = V_i(\alpha_{i,1},\ldots,\alpha_{i,m}) \quad . \qquad (16)$$

The algorithm continues with

$$X(i+1,1)=\sum_{j=1}^{m} \alpha_{i,j}\ G(k,j)_i \qquad k=1,\ldots,n\ , \qquad (17)$$

a selection of $G(k,2)_{i+1},\ldots,G(k,m)_{i+1}$,

$$X_{i+1}=\sum_{j=1}^{m} \alpha_j\ G(k,j)_{i+1} \qquad k=1,\ldots,n\ , \qquad (18)$$

and a subsequent minimization of

$$V_{i+1}(\alpha_1,\ldots,\alpha_m)=||AX_{i+1}-F||^2. \qquad (19)$$

The solution of the original system is derived from

$$X = \mathrm{Lim}_{i->+\infty} G(i,1). \qquad (20)$$

To illustrate the theory of the above section an example problem is presented in Appendix A.

CHAPTER III

DESCRIPTION OF NEW TRIAL SOLUTION GENERATION

Previously the trial solution generation process was based on knowledge of the physical system represented. The new trial solution selection process successfully makes the RCIP method independent of the knowledge of the physical system. The new method utilizes only the system of equations and information resulting from other trial solutions to obtain sets of trial solutions.

At the beginning of the first iteration, an initial trial solution set is selected. On successive iterations the trial solution set is composed of the last best solution, and a sequence of special Gauss-Seidel solutions. Equations to be solved by Gauss-Seidel for trial solutions are formed from the same coefficient matrix as the original problem and a new right hand side. For the second trial solution the new right hand side is the residual from the equation set using the last best solution. The initial guess used for the Gauss-Seidel solution is zero. Three Gauss-Seidel iterations are performed to obtain trial solution two. The third trial solution is similar to trial solution two. The difference is that another new right hand side is formed from the residual of the second trial solution. Again a Gauss-Seidel solution scheme is used to solve for the solution to this new set. An initial guess of zero is used, with three iterations

performed. The rest of the trial solutions are calculated in the same fashion until m number of trial solutions are reached. An example, of m equal to three trial solutions, is shown below.

Let

$$A*X=F \tag{21}$$

be an nxn system of equations. For the first iteration the first trial solution is

$$G(i,1) = \text{initial estimate} \qquad i=1,\ldots,n. \tag{22}$$

Now find the residual vector for the first trial solution (or T.S. 1)

$$R_i(\text{T.S. } 1) = A*G(i,1) - F_i \qquad i=1,\ldots,n. \tag{23}$$

Form the new equation set for trial solution two

$$A*G(i,2) = R_i(\text{T.S. } 1) \qquad i=1,\ldots,n. \tag{24}$$

Solve using Gauss-Seidel with an initial guess of zero. Perform three iterations to obtain G(i,2). Now find the residual vector for the second trial solution

$$R_i(\text{T.S. } 2)= A*G(i,2) - R_i(\text{T.S. } 1) \qquad i=1,\ldots,n. \tag{25}$$

Form the new equation set for trial solution three

$$A*G(i,3) = R_i(\text{T.S. } 2) \qquad i=1,\ldots,n. \tag{26}$$

Solve this set using Gauss-Seidel with an initial guess of zero. Perform three iterations to obtain G(i,3). The three trial solutions for the first iteration have been found. For the second through the nth iterations the only change is that the first trial solution is the last best solution.

The basis for this trial solution selection process is based on the following:

$$A*G(i,1) - F - R_i(T.S.\ 1) = 0 \qquad (27a)$$

$$A*G(i,2) - R_i(T.S.\ 1) - R_i(T.S.\ 2) = 0 \qquad (27b)$$

$$A*G(i,3) - R_i(T.S.\ 2) = 0 \qquad (27c)$$

subtract (27c) from (27b)

$$A*(G(i,2) - G(i,3)) - R_i(T.S.\ 1) = 0 \qquad (28)$$

subtract (28) from (27a)

$$A*(G(i,1) - G(i,2) + G(i,3)) - F = 0\ . \qquad (29)$$

Note, if G(i,3) is an exact solution to equation (27c) and

$$X_i = G(i,1) - G(i,2) + G(i,3). \qquad (30)$$

then X_i is an exact solution to equation (21). Note that G(i,3) is approximate therefore, X_i is approximate.

The numerical algorithm for the generation of the trial solutions is as follows. The first trial solution for the first iteration is

$$G(i,1) = \text{initial estimate} \qquad i=1,\ldots,n. \qquad (31)$$

The first trial solution for the second through the nth iterations is

$$G(i,1) = X_i \qquad i=1,\ldots,n. \qquad (32)$$

The second through the mth trial solution is calculated as follows

$$R_i = F_i \qquad i=1,\ldots,n \tag{33}$$

$$i=1,\ldots,n \qquad kk=2,m \qquad p=1,..,3 \tag{34a}$$

$$R_i(kk) = (A(i,k)*G(k,kk-1)) - R_i(kk-1) \tag{34b}$$

$$G(i,kk) = 0 \tag{34c}$$

$$G(i,kk)^p = R_i(kk) \,/\, A(i,i) - \sum_{k=1}^{n} A(i,k)*G(k,kk) \,/\, A(i,i) \; . \tag{34d}$$

The above algorithms will calculate m number of trial solutions. An illustrative problem is solved and included in Appendix B.

CHAPTER IV

APPLICATIONS AND RESULTS

This chapter is devoted to the solution of problems by the RCIP method and a comparison with some of the existing techniques. In particular the SOR and ADEP methods are compared with the RCIP method. The RCIP method is intended to solve linear systems of algebraic equations. Therefore, to solve the following differential equations a finite difference technique is used to model each differential equation as a system of algebraic equations.

In each problem the convergence limit used is always that the variance be less than or equal to some convergence criterion. The variance of the equation set is defined as the sum of the residuals squared, for all the equations. When an exact solution is known an average percent error can be obtained which is defined as the average of the differences between the exact solution and the approximate solution divided by the exact solution.

PROBLEMS ONE THROUGH FOUR

The following four ordinary differential equations are solved to show the generality of the new method. Various boundary conditions are

employed to illustrate a spectrum of problems. Problem one follows: Solve

$$\frac{d^2y}{dx^2} + y = 0 \tag{35}$$

for the region $0 <= x <= \pi/2$ with the derivative boundary conditions

$$y'(0) = 0 \quad , \quad y'(\pi/2) = 1 \quad . \tag{35a}$$

To illustrate the finite difference technique the above governing differential equation and boundary conditions will be modeled in algebraic form. The central difference approximation of the second order derivative is

$$\frac{d^2y}{dx^2} = \frac{y_{i+1} - 2\,y_i + y_{i-1}}{\Delta x^2} \tag{36a}$$

equation (35) becomes

$$\frac{y_{i+1} - 2\,y_i + y_{i-1}}{\Delta x^2} + y_i = 0 \quad . \tag{36b}$$

The central difference approximation of the first order derivative is

$$\frac{dy}{dx} = \frac{y_{i+1} - y_{i-1}}{2\Delta x} \tag{36c}$$

equations (35a) become

$$y_{i+1}(0) = y_{i-1}(0) \tag{36d}$$

$$y_{i+1}(\pi/2) - y_{i-1}(\pi/2) = 2\Delta x \quad . \tag{36e}$$

The above finite difference model approximates the differential equation as a system of linear equations.

The exact solution for problem one is

$$y_{exact} = - \text{COS } x \quad . \tag{36f}$$

The finite difference model for problem 1 is used to illustrate the SOR algorithm used. The algorithm follows:

$$Y_i(r+1) = (1-\lambda)*Y_i(r) + \lambda * \frac{Y_{i+1}(r+1) + Y_{i-1}(r+1)}{2 - \Delta X^2} \tag{37}$$

the variable r indicates the previous iteration, the value r+1 indicates the present iteration, and the subscript i indicates the nodal point. The variable λ is the SOR factor, which is between 1 and 2 for over relaxation and between 0 and 1 for under relaxation. The value of λ=1.0 is just the Gauss-Seidel routine. The remaining problems were modeled in the same manner as problem 1. And the SOR factor used in the following problems was the same as above. [5]

Problem two follows: Solve

$$\frac{d^2y}{dx^2} + y = 0 \tag{38}$$

for the region $0 <= x <= \pi/2$ with the composite boundary conditions

$$y'(0) + y(0) = 1 \tag{39}$$

$$y'(\pi/2) - y(\pi/2) = -1 \quad . \tag{40}$$

The exact solution for equations (38-40) is

$$y_{exact} = (\text{SIN } x + \text{COS } x)/2 \quad . \tag{41}$$

Problem three follows: Solve

$$x^2 \frac{d^2y}{dx^2} - 2x \frac{dy}{dx} + 2y = x^3 \tag{42}$$

for the region 1 <= x <= 2 with the mixed boundary conditions

$$y(1) = 0 \quad , \quad y'(2) = 0. \tag{43}$$

The exact solution for equations (42-43) is

$$y_{exact} = \frac{4}{3} x - \frac{5.5}{3} x^2 + \frac{1}{2} x^3 \quad . \tag{44}$$

Problem four follows: Solve

$$(1-x^2) \frac{d^2y}{dx^2} - 6x \frac{dy}{dx} - 4y = 0 \tag{45}$$

for the region 0 <= x <= 0.5 with the Dirichlet boundary conditions

$$y(0) = -4.5 \ , \ y(0.5) = 2 \quad . \tag{46}$$

The exact solution for equations (45-46) is

$$y_{exact} = \frac{27}{6} (3x-x^2/(1-x^2)^2) - \frac{27}{6} (1/(1-x^2)^2) \quad . \tag{47}$$

Standard finite difference techniques are used to algebraically approximate the above problems. In each of the first four problems a nodal network is chosen so that there is 50 spaces for each region, which requires a different number of equations for each problem. The different number of equations resulted from the different boundary conditions. A uniform initial estimate of 0.5 is chosen to start the methods.

In problem one the convergence limit was that the variance in the 51 equation set be less than or equal to 0.1. The results of the RCIP method for various numbers of trial solutions are shown in Table 1. The SOR method, with over and then under relaxation factors, did not obtain the convergence criterion in 10,000 iterations with a user

Table 1. Results of 1-D Ordinary Differential Equation Problem 1, with Derivative Boundary Conditions, for Various Numbers of Trial Solutions.

NUMBER OF TRIAL SOLUTIONS	NUMBER OF ITERATIONS	AVERAGE % ERROR	USER EXECUTION TIME (SEC)
2,3,4	DID NOT CONVERGE AT 100 ITERATIONS.		
5	11	0.3208	12.97
6	11	1.1182	15.59
7	8	1.1586	13.40
8	6	0.7707	11.63
9	5	0.2624	11.13

The SOR Method Did Not Converge at 10,000 Iterations.

execution time of 660 seconds. Therefore, the RCIP method is superior to SOR in that RCIP obtains a solution and SOR does not. When the number of trial solutions was less than 5 the convergence criterion was not obtained with the RCIP method, it did not converge at 100 iterations with a user time of 240 seconds. This is because the smaller number of trial solutions does not contain enough information to obtain rapid convergence.

Problem 2 is very similar to problem 1. The results of the solution of 51 equations are shown in Table 2. When the number of trial solutions is less than 8 the convergence criterion was not obtained with the RCIP method , this was because of the smaller number of trial solutions not containing enough information to obtain rapid convergence. The trial solutions 2-7 did not converge at 100 iterations with a time of approximately 240 seconds. The SOR method did not obtain the convergence criterion at 10,000 iterations with a time of 660 seconds. Several relaxation factors were utilized, over and then under relaxation, with no convergence being obtained. The RCIP method again, produces a solution and SOR does not. The limit of convergence for this problem is that the variance be less than or equal to 0.001.

Problem 3 was solved and the results are shown in Table 3 for the RCIP method and in Table 4 for the SOR method. A graph of the user execution time versus the number of trial solutions for the RCIP method, and over relaxation factor for the SOR method is included (Fig. 1). Clearly, as the graph shows, the convergence to a solution requires less time for the RCIP method than for the SOR method. The