Relative Error Propagation in the Recursive Solution of Linear Recurrence Relations

J. Oliver

Received January 8, 1966

Abstract. This paper is concerned with the numerical solution of the general initial value problem for linear recurrence relations. An error analysis of direct recursion is given, based on relative rather than absolute error, and a theory of relative stability developed. MILLER's algorithm for second order homogeneous relations is extended to more general cases, and the propagation of errors analysed in a similar manner. The practical significance of the theoretical results is indicated by applying them to particular classes of problem.

1. Introduction

An important method for the systematic evaluation of functions obeying a linear recurrence relation with respect to one of their parameters is to solve the relation recursively, making use of known initial values. The necessity of solving such initial value problems can arise in other contexts, such as in the numerical solution of differential equations. As is well known, the accumulation of errors due to inexact starting values and to rounding can sometimes prevent any useful accuracy from being obtained in the computed solution, unless a prohibitively large number of guarding figures is used. Thus an obvious need exists to determine those circumstances in which the accumulation of errors can present a serious computational problem, and to develop alternative methods of solution which avoid this unpleasant error behaviour.

Considerable study has been made of relations with constant coefficients, of the general form

\[ a_n y(i) + a_{n-1} y(i+1) + \cdots + a_0 y(i+n) = r(i), \]

which are traditionally described as unstable when one or more of the roots \( \lambda_1, \lambda_2, \ldots, \lambda_n \) of the characteristic equation

\[ a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_0 = 0 \]

lie outside the unit circle — see for example MITCHELL and CRAGGS [5], or more recently METROPOLIS [4]. Thus if (1) is solved recursively in the direction of increasing \( i \), any propagated error will be a linear combination of the fundamental solutions \( \lambda_1, \lambda_2, \ldots, \lambda_n \) of the homogeneous form of (1), provided the roots of (2) are distinct. If one or more of these roots exceeds unity in magnitude then some or all of the propagated errors will eventually increase unboundedly in absolute value, causing the number of correct decimal places in the solution to diminish as \( i \) increases. Now although this theory of instability is quite adequate for analysing the growth of absolute errors, it has little bearing on the behaviour of
errors relative to the required solution. If we should wish to determine the number of correct significant figures in the computed values, then the absolute stability of the relation is quite irrelevant; what matters is the behaviour of the propagated errors relative, not to unity, but to the required solution.

In the first half of this paper we derive a concept of error behaviour for recursive methods of solving the general linear recurrence relation which, although analogous to the theory of absolute stability, is based on a consideration of relative rather than absolute error. This is followed by a discussion of a method, using backward recursion, which is effectively an extension of the MILLER algorithm [1] for homogeneous second order relations, but is applicable to more general cases.

2. Error Analysis of Forward Recursion

Consider the general linear recurrence relation

\[ P_n(i) y(i) + P_{n-1}(i) y(i+1) + \cdots + P_0(i) y(i+n) = R(i) \]  

(3)

where \( n \geq 1 \) and \( P_0(i), P_n(i) \neq 0 \).

We wish to compute, over some range \( i = a, a+1, \ldots, b \), a particular solution \( Y(i) \) defined by the initial conditions

\[ Y(a+r) = a_r \quad (r = 0, 1, \ldots, n-1), \]

(4)

using (3) recursively to evaluate in turn \( Y(a+n), Y(a+n+1), \ldots, Y(b) \). Note that we are concerned only with the mathematical problem defined by (3) and (4), and not with the physical system from which the problem may possibly have arisen.

Assuming initially that the exact values of \( P_r(i) \) and \( R(i) \) are used, any errors in the computed values \( Y(i) \) of the solution are caused by the use of inexact initial values \( Y(j) \);

\[ Y(j) \equiv a_j + e_j \quad (j = a, \ldots, a+n-1), \]

(5a)

and by rounding errors \( e_j \) introduced in evaluating each \( Y(j) \);

\[ Y(j) \equiv \{ R(j-n) - P_n(j-n) Y(j-n) - \cdots - P_1(j-n) Y(j-1) \}/P_0(j-n) + e_j \]

\( (j = a+n, \ldots, b). \)

(5b)

Each such \( e_j \) will give rise to propagated errors \( e_j(i) \) in the subsequently computed values \( Y(i) \), and if \( y_1(i), \ldots, y_n(i) \) constitute a fundamental system of solutions of the homogeneous form of (3) (see for example FORT [2]), then we may write

\[ e_j(i) = \{ c_{j,1} y_1(i) + \cdots + c_{j,n} y_n(i) \} e_j \]

(6)

for some constants \( c_{j,r} \).

Provided we make the assumption (to be discussed later) that \( Y(i) \) is everywhere non-zero, then we may define relative errors \( \eta_j(i) \) in \( Y(i) \) by

\[ \eta_j(i) \equiv e_j(i)/Y(i) \quad (j = a, \ldots, b), \]

\[ \eta_j(i) \equiv e_j(i)/Y(i) \quad (i = j, \ldots, b; j = a, \ldots, b). \]

(7)