MATHEMATICS OF COMPUTATION, VOLUME 28, NUMBER 126, APRIL 1974, PAGES 465–473

Approximation by Aliasing with Application to "Certaine" Stiff Differential Equations

By Arthur David Snider and Gary Charles Fleming

Abstract. The usual method of finding an accurate trigonometric interpolation for a function with dominant high frequencies requires a large number of calculations. This paper shows how aliasing can be used to achieve a great reduction in the computations in cases when the high frequencies are known beforehand. The technique is applied to stiff differential equations, extending the applicability of the method of Certaine to systems with oscillatory forcing functions.

1. Introduction. In general, when one wishes to perform a Fourier analysis on a periodic function $f(t)$ using sampled data, the estimation of the Nth Fourier coefficient requires at least $2N$ data points ([1], [2]). The present paper shows that it is possible to do this with a much smaller data set under special circumstances, namely, when $f(t)$ is a sum of a smooth function and of a few harmonics of high, known frequencies. The technique involves the use of aliasing [1] to orthogonally project out the high coefficients with just a few computations.

This can be used to extend the applicability of Certaine's method ([3], [4]) in numerically solving systems of stiff differential equations of the form

$$dy/dx = My + g(y, x).$$

Here x is the independent variable, y and g are vector functions, and M is a matrix with large eigenvalues. This latter property ("stiffness") will dictate the use of an extremely fine mesh, resulting in an expensive computation, unless some special technique is used. In [3] and [4], an integrating factor $\exp(Mt)$ is introduced to overcome this difficulty, and the function g is approximated by interpolating polynomials, yielding a stable, accurate predictor-corrector scheme at reasonable mesh lengths in those cases when g is known to be smooth and slowly-varying. The trigonometric interpolation scheme which we describe herein permits an extension of this technique to cases where g is oscillatory, without destroying its basic attractive feature—its employment of reasonably-sized mesh lengths.

2. The Approximation. Let us suppose that $f(t)$ has period 2π, and that we wish to estimate the Fourier coefficients for the terms $\sin nt$ and $\cos nt$ for, say, n up to 1000, using sampled data. Normally, we would proceed as in [1]; to find a trigonometric sum of the form

$$C_N(t) = \frac{A_0}{2} + \sum_{r=1}^{N-1} (A_r \cos rt + B_r \sin rt) + \frac{A_N}{2} \cos Nt \tag{1}$$

Received February 16, 1973.
AMS (MOS) subject classifications (1970). Primary 42A08, 42A12, 65L05.

Copyright © 1974, American Mathematical Society

which approximates $f(x)$, we fit the function at the $(2N + 1)$ points t_j:

$$t_j = (j/2N)2\pi, \qquad j = 0, 1, \cdots, 2N.$$

Since the trigonometric functions are orthogonal with respect to summation over $\{t_j\}$, the coefficients are easily shown to be

$$A_r = \frac{1}{N} \sum_{j=0}^{2N-1} f(t_j) \cos rt_j,$$

$$B_r = \frac{1}{N} \sum_{j=0}^{2N-1} f(t_j) \sin rt_j.$$

Of course, if we are to detect the components $\sin 1000t$ and $\cos 1000t$, we must choose N larger than 1000, i.e., we must use more than 2000 data points. This may be undesirable for a number of reasons: time, storage, accumulation of round-off errors. Thus, it is important that in some circumstances fewer data points may be used.

Suppose we know a priori that $f(t)$ is expressible as

$$f(t) = h(t) + \sum_{m=1}^{p} c_m \cos R_m t + d_m \sin R_m t \tag{2}$$

where $h(t)$ is a smooth function whose Fourier coefficients decrease rapidly and the p (known) frequencies $R_1 < R_2 < \cdots < R_p$ are very large; specifically, in the Fourier expansion of $h(t)$,

$$h(t) = \frac{a_0'}{2} + \sum_{r=1}^{\infty} a_r' \cos rt + b_r' \sin rt \tag{3}$$

the magnitudes of a_r' and b_r' are negligible (for our purposes) when $r > L$, while each of the frequencies R_m is greater than L. Loosely speaking, we know that our function has a few high-frequency components in it, at known frequencies, and otherwise is slowly varying. It is then our goal to efficiently estimate the coefficients c_m and d_m, and the first L coefficients a_r' and b_r'.

The solution to this problem is accomplished through aliasing (cf. [1]). Observe that at each of the points $t = t_j$, any function $\cos R_m t$ can be replaced by $\cos r_m t$ for some $r_m \leqq N$ (and similarly for $\sin R_m t$) according to the identities

$$\cos[(2q)N + r]t_j = \cos rt_j,$$

$$\cos[(2q + 1)N + r]t_j = \cos(N - r)t_j,$$

$$\sin[(2q)N + r]t_j = \sin rt_j,$$

$$\sin[(2q + 1)N + r]t_j = -\sin(N - r)t_j.$$

(Intuitively, $\cos[(2q)N + r]t$ takes the same values as $\cos rt$ at the mesh points but oscillates faster in between.) Therefore, if we use a coarse mesh, i.e., $2N + 1$ mesh points with $N < R_1$, each of the harmonics with frequencies R_m will be "equivalent" to a harmonic with a lower frequency r_m $(< N)$, and we can use orthogonality relations to project out the coefficients (c_m, d_m). Of course, the effect of aliasing is to combine the coefficients in the following way: If the actual Fourier coefficients of $f(t)$ are (a_r, b_r) and the coefficients of the trigonometric interpolation sum are (A_r, B_r) as

in Eq. (1), then we have (cf. [1])

$$A_r = a_r + \sum_{m=1}^{\infty} (a_{2mN+r} + a_{2mN-r}),$$

$$B_r = b_r + \sum_{m=1}^{\infty} (b_{2mN+r} - b_{2mN-r}).$$

So we must choose N in such a way that none of the frequencies $r = 0, 1, 2, \cdots, L - 1, L, R_1, R_2, \cdots, R_p$ are combined: that is, their trigonometric functions must be orthogonal to each other. Clearly, we must have $N \geqq L + p$, but usually N need not be nearly as large as R_p; hence we may achieve a great saving over the usual procedure.

In summary, we choose N so that all the important frequencies are equivalent to different frequencies, each less than N, on the $(2N + 1)$ mesh points; then we perform the usual discrete Fourier analysis for this value of N; and, finally, we reidentify the higher frequencies in the result.

As an illustrative example, assume that $f(t)$ is the sum of a slowing varying function plus three harmonics of frequencies 177, 589, and 1000. To estimate the magnitude of these harmonics by the usual method would require us to use at least 2001 mesh points. However, observe that if we use $N = 52$, i.e., 105 mesh points, then

$$\begin{aligned}
\cos 1000t_j &= \cos 40t_j, \\
\sin 1000t_j &= -\sin 40t_j, \\
\cos 589t_j &= \cos 35t_j, \\
\sin 589t_j &= -\sin 35t_j, \\
\cos 177t_j &= \cos 31t_j, \\
\sin 177t_j &= -\sin 31t_j,
\end{aligned}$$

where $t_j = j\pi/52, \quad j = 0, 1, \cdots, 104$. Thus if we fit $f(t)$ at the points t_j by

$$f(t_j) = \frac{A_0}{2} + \sum_{r=1}^{51} (A_r \cos rt_j + B_r \sin rt_j) + \frac{A_{52}}{2} \cos 52t_j,$$

we can say

$$\begin{aligned}
f(t) \approx \frac{A_0}{2} &+ \sum_{r=1}^{30} (A_r \cos rt + B_r \sin rt) \\
&+ A_{31} \cos 177t - B_{31} \sin 177t \\
&+ A_{35} \cos 589t - B_{35} \sin 589t \\
&+ A_{40} \cos 1000t - B_{40} \sin 1000t,
\end{aligned}$$

and the error we make here is the truncation, after 30 terms, of the Fourier series for the slowly varying part of $f(t)$.

A quite precise error analysis of the approximation can be obtained using the techniques in [2]. There it is shown that, because of the well-known inequality $|a_r'|, |b_r'| \leqq (2/r^K) \max|h^{(K)}(t)|$ (obtained by integration by parts), the error estimates for the first L coefficients are given by

$$|A_r - a_r|, \qquad |B_r - b_r| \leqq 5 \max |h^{(K)}(t)|/N^K,$$

and quite similar reasoning shows that, for the aliased coefficients,

$$|A_{r_m} - a_{R_m}|, \qquad |B_{r_m} - b_{R_m}| \leqq \left[\frac{2}{r_m^K} + \frac{5}{N^K}\right] \max |h^{(K)}(t)|.$$

This yields (cf. [2] for details of a similar calculation)

$$\begin{aligned} &\left| f(t) - \frac{A_0}{2} - \sum_{r=1}^{L} (A_r \cos rt + B_r \sin rt) \right. \\ &\qquad \left. - \sum_{m=1}^{P} (A_{r_m} \cos R_m t + B_{r_m} \sin R_m t) \right| \\ &\leqq \left[\frac{5(2p + 2L + 1)}{N^K} + \frac{4}{(K-1)L^{K-1}} + \sum_{m=1}^{P} \frac{4}{r_m^K} \right] \max |h^{(K)}(t)|, \end{aligned} \tag{4}$$

$f(x) = -\sin 177t + \sin 589t + \cos 1000t + 2ii t - t^2$

i	Exact Coefficients a_i	Exact Coefficients b_i	Errors in third decimal for N = 1500 a_i	Errors in third decimal for N = 1500 b_i	Errors in third decimal for N = 52 a_i	Errors in third decimal for N = 52 b_i
0	13.159	-	2	-	1	-
1	-4.000	0	2	1	1	0
2	-1.000	0	0	0	1	0
3	-.444	0	0	0	2	0
4	-.250	0	0	0	1	0
5	-.160	0	0	0	1	0
6	-.111	0	0	0	1	0
7	-.082	0	0	0	1	0
8	-.063	0	0	0	1	0
9	-.049	0	0	0	2	0
10	-.040	0	0	0	1	0
20	-.010	0	0	0	1	0
30	-.004	0	0	0	2	0
31	-.004	0	0	0	-	-
35	-.003	0	0	0	-	-
40	-.003	0	0	0	-	-
177	.000	-1.0	0	1	5	0
589	.000	1.0	0	0	5	0
1000	1.0	0	1	0	4	1

$f(x) = -\sin 177t + \sin 589t + \cos 1000t + \begin{cases} t, & t \leqq \pi \\ \pi, & t > \pi \end{cases}$

i	Exact a_i	Exact b_i	Errors, N = 1500 a_i	Errors, N = 1500 b_i	Errors, N = 52 a_i	Errors, N = 52 b_i
0	4.712	-	1	-	0	-
1	-.637	-1.000	0	0	0	0
2	0	-.500	0	0	0	1
3	-.071	-.333	0	0	0	1
4	0	-.250	0	0	0	1
5	-.025	-.200	0	0	1	2
6	0	-.167	0	0	0	2
7	-.013	-.143	0	0	0	2
8	0	-.125	0	0	0	2
9	-.008	-.111	0	0	0	3
10	0	-.100	0	0	0	1
20	0	-.050	0	0	0	6
30	0	-.033	0	0	0	9
31	-.001	-.032	0	0	-	-
35	-.001	-.029	0	0	-	-
40	0	-.025	0	0	-	-
177	.000	-1.006	0	1	1	28
589	.000	.998	0	0	1	19
1000	1.000	-.001	1	1	0	14

when $h(t)$ has K derivatives. The major error, generally speaking, is the neglecting of a_{L+1}' and b_{L+1}'.

Some numerical results are presented in the table above. For the function indicated, we display the exact Fourier coefficients in the first columns, then the errors in the third decimal place resulting from computing these coefficients by the usual sampled-data method using $N = 1500$, and, finally, the errors resulting from using $N = 52$ and aliasing. The approximations are quite good for both the low and high frequencies, the errors in the latter being comparable in magnitude to the first neglected coefficients a_{31} and b_{31}.

Two remarks are in order before we turn to the application of this technique to stiff differential equations. First, the efficiency of the method hinges on the success in finding a (fairly) small integer N which "aliases" the given frequencies $R_1, R_2, \cdots, R_p$ down to *distinct* lower frequencies $r_1, \cdots, r_m$, each greater than the given frequency L. This is a highly complex number-theoretic problem for which we have found no simple solution, but in the Appendix we present an algorithm based essentially on trial-and-error (dosed with some short-cuts) which yields the best N. Our experience indicates that great savings can usually be expected except for the obvious pathological situations (p too large). The number of computations involved is proportional to N^2 (or $N \log N$ if fast Fourier transforms are used) (cf. [1]).

Second, one may observe that if we perform the "aliased" analysis twice, with different values of N, the high-frequency coefficients would be combined with *different* low-frequency components and thus we could recover them by subtraction. This may be easier in some cases than using the "best" N, but since it is not readily adaptable to the application we have in mind, we leave the details to the interested reader.

3. Application. We now indicate how the approximation procedure described above may be used, in some situations, to extend the method described in [3] for the numerical solution of certain stiff systems of differential equations. We begin with a brief summary of Certaine's technique, as expounded by Guderley and Hsu [4].

The differential equation system is written in the form

$$y'(x) = -Dy(x) + g(y(x); x). \tag{5}$$

Here y and g are vector functions and D is a constant matrix, some of whose eigenvalues are large in magnitude. (In fact, most authors use the terminology "stiff systems" for the case when these eigenvalues are large positive, but we shall not restrict ourselves at this point.)

Using an integrating factor, (5) is recast as

$$y(x_{n+1}) = \exp(-Dh)y(x_n) + \int_{x_n}^{x_{n+1}} \exp(D(x - x_{n+1}))g(y(x); x)\, dx \tag{6}$$

where $h = x_{n+1} - x_n$. (6) is the basis for a predictor-corrector scheme for computing y_{n+1}, the approximation to $y(x_{n+1})$. The function $g(y(x); x)$ is fitted by a polynomial $g_K(x)$ of order K at the points x_{n-K}, x_{n-K+1}, x_n for the predictor, and at the points $x_{n-K+1}, x_{n-K+2}, \cdots, x_{n+1}$ for the corrector (possibly with a Newton-Raphson iteration for y_{n+1} in the corrector). (6) is thus replaced by an equation of the form

$$y_{n+1} = \exp(-Dh)y_n + \exp(-Dx_{n+1}) \int_{x_n}^{x_{n+1}} \exp(Dx)g_K(x)\, dx. \tag{7}$$

The integral can be evaluated explicitly (cf. [3] and [4] for details) and (7) is thus suitable for computation. (We note in passing that if the exponential matrices are hard to compute, D is decomposed into $D_1 + D_2$, where $\exp(-D_1)$ is computable (e.g., D_1 is diagonal), and the term $D_2 y$ is absorbed into the function g.)

The virtues of this technique can be stated, informally, in the following way. When the eigenvalues of D are large (the phenomenon of "stiffness"), the usual integration schemes would require the mesh size h to be very small in order to expect any accuracy; this in turn would require many applications of the scheme to integrate the solution over a reasonable length interval, and the resulting accumulation of round-off and truncation errors might well spoil the accuracy. But by using (7) the only error comes from the interpolation for g, and thus this interpolation alone, are not D's eigenvalues, dictate the mesh size. (This is the case unless D has eigenvalues with large negative real parts, in which case the factor $\exp(D(x - x_{n+1}))$, appearing inside the integral, must be considered together with the approximation for g. As we noted earlier, these cases are usually not regarded as "stiff".) If g is a sufficiently slowly varying function of x, a coarse mesh can be used. In fact, if g happens to be a polynomial of order less than $K + 1$, the scheme is exact; thus Dahlquist's A-stability criterion is met (cf. [4]). A detailed error analysis of the procedure is presented in [4] (however, see also [5]).

Here, we propose to extend this technique to some cases where g is oscillatory, rather than slowly varying. The obvious modification is then the employment of a trigonometric, rather than polynomial, interpolation for g. Of course, if only a few low frequencies are used in the trigonometric sum, the approximation will probably be no better than that obtained with the polynomial fit. And if high frequencies are to be used, then the usual Fourier technique would require many mesh points, i.e., a small value for h; this is precisely what we have been trying to avoid. But if we know a priori the important high frequencies contributing to g, then the aliasing approximation presented above may allow us to calculate an appropriate trigonometric approximation using a coarse mesh, and we are back in business.

The salient features of this procedure are:

(a) If $g_K(x)$ in (7) is a trigonometric sum, the integral can be evaluated explicitly, so the scheme is again suitable for computation. We do not present the formulae here; the derivation is simple but laborious.

(b) If g happens to *be* a finite trigonometric sum with less than $K + 1$ terms, the formula is exact.

(c) Most analyses of stiff systems assume that the dominant eigenvalues of D are positive. But notice that Certaine's method will work even if D has large imaginary eigenvalues (implying oscillatory solutions $y(x)$), as long as $g(y(x); x)$ is smooth. However, if g actually does depend on y (as it certainly will if g has absorbed the term $D_2 y$, mentioned earlier), it will inherit y's oscillatory behavior, and thus be unsuitable for polynomial interpolation. In such a case the aliasing procedure may be quite appropriate; the "known" high frequencies in g would include, of course, the frequencies involved in the homogeneous solutions of (5), which can be precomputed from the eigenvalues of D.

(d) All the frequencies used in $g_K(x)$ must be integers (or at least rationals) and the important high frequencies must be known beforehand (since the mesh size must be chosen appropriately for aliasing).

(e) The method is not self-starting.

(f) An error analysis can be constructed following the same pattern as in [4], using Eq. (4). However, it is not highly illuminating; clearly, the effectiveness of the scheme hinges on the analyst's ability to correctly predict which frequencies will be important.

This procedure has been used by the senior author in studying the effect of certain hardware parameters on "gyro drift," the term used to describe a secular error in the performance of an inertial guidance system. The equations look roughly like a forced, coupled mass-spring system. The dominant frequencies in the forcing function were identified with certain environmental vibration rates, the rate of spin of the gyro wheel, and the natural frequencies of the unforced system. The data were chosen so that all of the frequencies were integers. Nominal starting values were used and the simulation was run until a steady state was achieved. The results were sufficiently good to aid in designing the instruments for optimal performance.

Appendix. In applying the method described in Section 2 to a given situation, one is confronted with the problem of choosing a suitable value for N, where $(2N + 1)$ is the number of data points. It must be chosen so that all of the desired frequencies are replaced by *different* frequencies less than N; that is, none of the important frequencies are combined. We state the problem precisely.

Definition. Given positive integers $L < R_1 < R_2 < \cdots < R_p$ and the integer N greater than L, we define ρ_i $(i = 1, \cdots, p)$ by the following: Expressing $R_i = q_i N + r_i$ via the division algorithm, set $\rho_i = r_i$ when q_i is even, and $\rho_i = N - r_i$ when q_i is odd. Then we say N separates the frequencies R_i above L *if the following conditions hold*:

(i) each $\rho_i > L$,

(ii) $i \neq j$ implies $\rho_i \neq \rho_j$.

The problem is then to find the smallest such N.

We have devised an algorithm for finding N, and it is presented in the flow chart below. The process is basically trial-and-error, but the following considerations have enabled us to proceed quite efficiently.

First of all, rather than test successive values of N, we employ the quotient q_i of R_i divided by N. Omitting subscripts for the moment, we observe that

$$\rho = R - qN, \qquad q \text{ even},$$

$$\rho = N(q + 1) - R, \qquad q \text{ odd}.$$

The condition $\rho > L$ becomes, in terms of the quotient q,

$$N < (R - L)/q, \qquad q \text{ even},$$

$$N > (R + L)/(q + 1), \qquad q \text{ odd}.$$

However, it is easy to show that, because $N > L$,

$$(R - L)/(q - 1) > N > (R + L)/(q + 2).$$

Combining these, we can show that if the quotient of R divided by N is either the even number q_e or the odd number $q_e + 1$, then in order for N to separate the frequencies above L, we must have

$$(R + L)/(q_e + 2) < N < (R - L)/q_e.$$

This inequality is the basis of our search for N. We form these intervals for de-

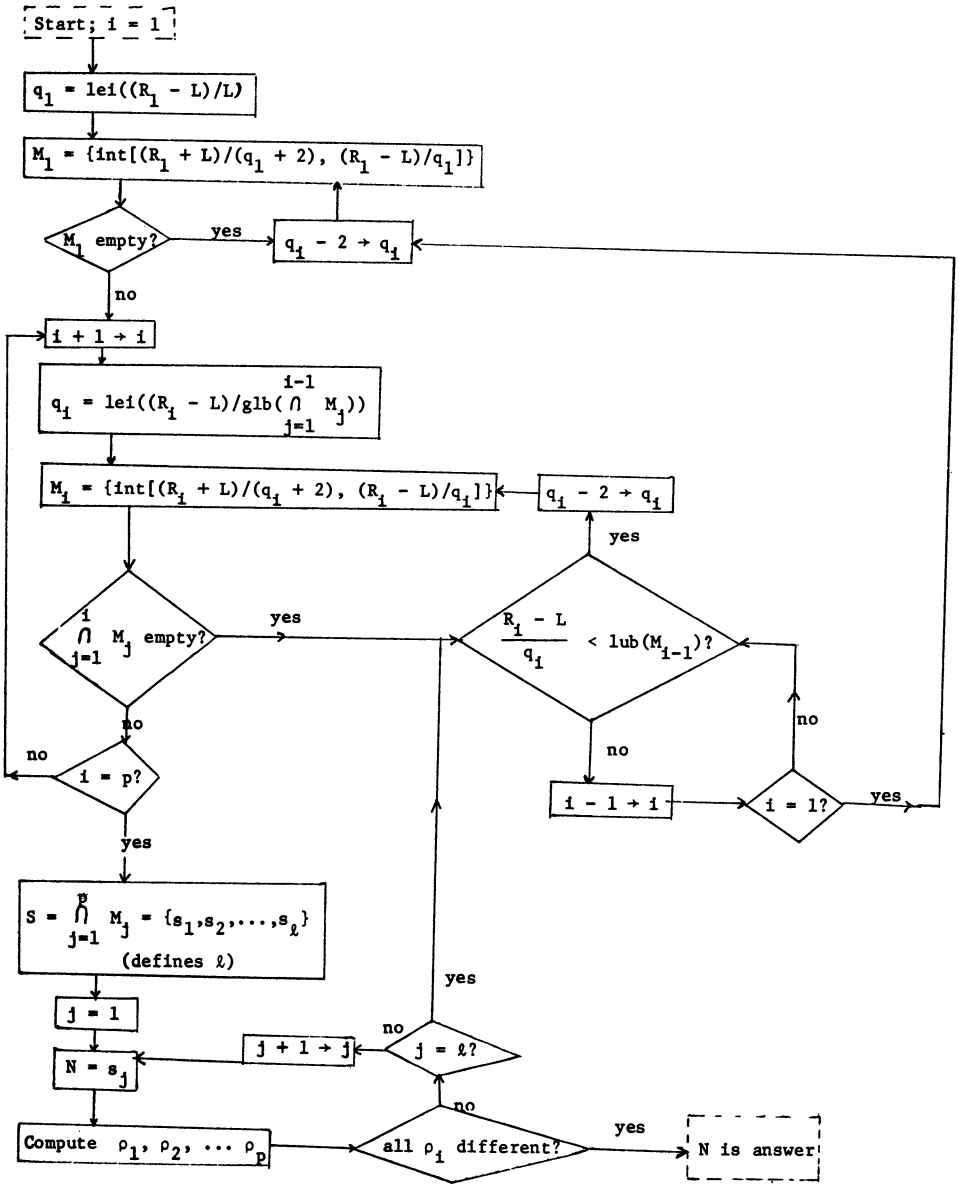
creasing even numbers q_e, and consider the integers located therein. When an integer N appears in such intervals for each frequency R_i ($i = 1, \cdots, p$), then we know that each ρ_i is greater than L, and we only have to test if all the ρ_i are different.

Of course, we start the search with the largest q_e which yields a nonempty interval; this value is the highest even number less than $(R_1 - L)/L$.

The algorithm presented in the flow chart is based on these considerations.

Abbreviations

lei(A) = largest even integer less than A.
{int[A, B]} = set of integers in the interval A, B.
lub A = greatest element in the set A.
glb A = least element in the set A.
ρ_i is computed as in the text.



Department of Mathematics
University of South Florida
Tampa, Florida 33620

1. R. W. HAMMING, *Introduction to Applied Numerical Analysis*, McGraw-Hill, New York, 1971, pp. 287–292.

2. A. D. SNIDER, "An improved estimate of the accuracy of trigonometric interpolation," *SIAM J. Numer. Anal.*, v. 9, 1972, pp. 505–508.

3. J. CERTAINE, "The solution of ordinary differential equations with large time constants," in *Mathematical Methods for Digital Computers*, A. Ralston and H. S. Wilf (Editors), Wiley, New York, 1960, pp. 128–132. MR **22** #8691.

4. K. G. GUDERLEY & C.-C. HSU, "A predictor-corrector method for a certain class of stiff differential equations," *Math. Comp.*, v. 26, 1972, pp. 51–69. MR **45** #8001.

5. A. D. SNIDER, "A remark on a paper by Guderley and Hsu." (In prep.)