Stabilizing Predictors for Weakly Unstable Correctors

By Hans J. Stetter

1. Introduction. It is well known that Milne-Simpson's method

$$y_{n+2} = y_n + \frac{h}{3}(f_n + 4f_{n+1} + f_{n+2}) \tag{1}$$

should not be used for the numerical integration of $y' = f(x, y)$ if $f_y < 0$ along the true solution $y(x)$ although the solution of (1) converges to $y(x)$ for fixed finite x as $h \to 0$ (see, e.g., [1]). In fact, rapid oscillations, with an amplitude increasing exponentially as the numerical integration proceeds, will supersede the values approximating $y(x)$ and eventually destroy the meaningfulness of the computation. This "weak unstability" occurring with (1) and similar algorithms has been well analyzed (e.g., [1, p. 248 ff.]) and procedures have been suggested to weaken its *effect* (e.g., [2]). We will show in this paper that it is quite easy to completely eliminate its *cause*: The combination of a judiciously chosen predictor with the weakly unstable corrector constitutes a strongly stable algorithm *if the corrector is not iterated.*

2. Analysis. Consider the k-step scheme

$$\rho(E)y_n - h\sigma(E)f_n = 0, \tag{2}$$

where $\rho(z) := \sum_{\nu=0}^{k} \alpha_\nu z^\nu$, $\alpha_k = 1$; $\sigma(z) := \sum_{\nu=0}^{k} \beta_\nu z^\nu$; $Ey_n := y_{n+1}$; $f_n := f(x_n, y_n)$.

(2) is called D-*stable*[1] or *stable for* $h \to 0$ if all zeros of ρ are in $|z| \leq 1$ and no multiple zeros are on $|z| = 1$. (2) is *of order* p if, for a sufficiently differentiable function y,

$$\rho(E_h)y(x) - h\sigma(E_h)y'(x) = O(h^{p+1}),$$

where $E_h y(x) := y(x + h)$.

It is well known (e.g., [1]) that the sequence y_n generated by a D-stable scheme (2) of order $p \geq 1$ *converges* in an obvious sense to the solution $y(x)$ of $y' = f(x, y)$ as $h \to 0$. It is more difficult to predict the behavior of the y_n *for finite* h as weak instabilities may occur.

Denoting by $\zeta_\nu(H)$, $\nu = 1(1)k$, the zeros of the polynomial

$$\varphi(z, H) := \rho(z) - H\sigma(z), \tag{3}$$

we know from [1, p. 238], that for a scheme (2) of order p there is one zero, which we will always denote by $\zeta_1(H)$, which satisfies

$$\zeta_1(H) = e^H + O(H^{p+1}). \tag{4}$$

For a given value of H (real) we will call a D-stable scheme (2) *strongly stable* if[2]

Received May 27, 1964. The research reported in this paper was supported in part by the U.S. Air Force under Grant No. AF EOAR 63-77 through the European Office, OAR.

[1] For Dahlquist-stable (cf. [3]).

[2] See Remark at the end of this section.

$$(5) \qquad |\zeta_\nu(H)| \leqq \zeta_1(H), \qquad \nu = 2(1)k,$$

and *weakly unstable* otherwise.

Each D-stable scheme is strongly stable for $H = 0$, by continuity there will be a largest number $H^+ \geqq 0$ and a smallest number $H^- \leqq 0$ such that (2) is strongly stable for each H from the *stability interval* $[H^-, H^+]$.[2]

It is evident for constant $g(x) := f_y(x, y(x))$ and confirmed by experience for variable g that the solution y_n of (2) simulates the behavior of $y(x)$ if hg remains within the stability interval. For a *weakly unstable scheme* (e.g., Milne-Simpson's method (1)) $H^- = 0$ and the method should not be used for $g < 0$.

If $\beta_k \neq 0$, (2) defines y_{n+k} implicitly and is usually replaced by the *predictor-corrector scheme*[3]

$$(6) \qquad \begin{aligned} y_{n+k}^{(0)} &= -\sum_{\nu=0}^{k-1} \alpha_\nu^* y_{n+\nu} + h \sum_{\nu=0}^{k-1} \beta_\nu^* f_{n+\nu}, \\ y_{n+k}^{(i)} &= -\sum_{\nu=0}^{k-1} \alpha_\nu y_{n+\nu} + h\left(\sum_{\nu=0}^{k-1} \beta_\nu f_{n+\nu} + \beta_k f(x_{n+k}, y_{n+k}^{(i-1)})\right), \qquad i = 1(1)m. \end{aligned}$$

A simple computation shows that for the algorithm (6) the polynomial (3) is transformed into[4]

$$(7) \quad \varphi^m(z, H) := (1 - B^m)(\rho(z) - H\sigma(z)) + B^m(1 - B)(\rho^*(z) - H\sigma^*(z))$$

with $B := H\beta_k$, $\rho^*(z) := \sum_{\nu=0}^{k} \alpha_\nu^* z^\nu$, $\alpha_k^* = 1$, $\sigma^*(z) := \sum_{\nu=0}^{k-1} \beta_\nu^* z^\nu$. Obviously $\lim_{m\to\infty} \varphi^m(z, H) = \varphi(z, H)$ if $|B| < 1$.

Assume that the predictor is of order $q \geqq 0$. It is clear from (3), (4), and (7) that the zeros ζ_ν^m of φ^m satisfy (after a suitable ordering)

$$(8) \qquad \begin{aligned} \zeta_1^m(H) &= e^H + O(H^{p+1}) + O(H^{q+m+1}), \\ \zeta_\nu^m(H) &= \zeta_\nu(H) + O(H^m). \end{aligned}$$

For all weakly unstable schemes of practical importance the violation of (5) for $H < 0$ is a *first-order effect in H*, hence only the zeros ζ_ν^1 of φ^1 may possibly not share the undesirable behavior of the ζ_ν.[5] Therefore we may restrict our considerations to the case $m = 1$; we will—for given weakly unstable schemes—attempt to select (ρ^*, σ^*) such that the stability interval for φ^1 has $H = 0$ as an interior point.

Remark: Some authors (e.g., [5]) replace (5) by $|\zeta_\nu(H)| \leqq 1$ in the definition of a stability interval. This seems not appropriate since, e.g., a 2-step scheme with $\zeta_2(H) = -1 - H/2 + O(H^2)$ will also generate oscillations growing exponentially relative to the true solution if used for $y' = -y$.

3. Selection of the Predictor. From now on we will only consider the polynomial $\varphi^1(x, H)$ and its zeros $\zeta_\nu^1(H)$, $\nu = 1(1)k$, hence we will omit the superscript 1. Furthermore we define $\zeta_{\nu 0} := \zeta_\nu(0)$.

---

[3] If the predictor reaches back farther than the corrector the degree k of the corrector has to be formally raised accordingly.

[4] This assumes a $P(EC)^m E$ algorithm (cf. [4]); for a $P(EC)^m$ algorithm the situation is more complicated. See footnote 5, however.

[5] Since (8) holds equally for $P(EC)^m$ algorithms (see [4]) our conclusion is also true for this case.

If $|\zeta_{\nu 0}| < 1$ for a certain $\nu > 1$, (5) has to hold in a full vicinity of $H = 0$ by continuity. Therefore it suffices to consider $\nu \in W := \{\nu: 2 \leqq \nu \leqq k, |\zeta_{\nu 0}| = 1\}$. For $\nu \in W$, let

$$(9) \qquad |\zeta_\nu(H)| = 1 + A_\nu H + B_\nu H^2 + O(H^3).$$

As $p \geqq 2$ in all cases of interest, (4) and (5) yield the following *necessary condition*:

$$(10) \qquad \begin{aligned} &(a)\ A_\nu = 1, \\ &(b)\ B_\nu \leqq \tfrac{1}{2}, \end{aligned} \qquad \text{for } \nu \in W.$$

If the equality is excluded in (10b), condition (10) is *sufficient* as well to guarantee a stability interval with $H^- < 0$, $H^+ > 0$. (For $B = \frac{1}{2}$, the third order terms would have to be investigated.) To find expressions for the A_ν and B_ν we derive, from

$$(11) \qquad \begin{aligned} \varphi^1(z, H) &= [\rho(z) - H(\sigma(z) - \beta_k \rho^*(z)) - H^2 \beta_k \sigma^*(z)](1 - B), \\ \zeta_\nu(H) &= \zeta_{\nu 0} + H \cdot \frac{\tau_\nu}{\rho_\nu'} + H^2[-\rho_\nu'' \tau_\nu^2 / 2\rho_\nu' + \tau_\nu \tau_\nu' + \beta_k \rho_\nu' \sigma_\nu^*]/\rho_\nu'^2 + O(H^3), \end{aligned}$$

where $\tau(z) := \sigma(z) - \beta_k \rho^*(z)$, the prime denotes differentiation, and $\rho_\nu := \rho(\zeta_{\nu 0})$, etc. $\rho_\nu' \neq 0$ for a D-stable scheme and $\nu \in W$. Let $\zeta_{\nu 0} = e^{i\omega_\nu}$, then (10a) becomes

$$(12a) \qquad \mathrm{Re}\left\{ e^{-i\omega_\nu} \frac{\tau_\nu}{\rho_\nu'} \right\} = 1.$$

Since τ_ν is linear in the coefficients α_ν^* of ρ^*, for given ρ, σ, condition (10a) takes the form of a linear relation between the α_ν^* (which are assumed real) for each $\nu \in W$.

Condition (10b) becomes an inequality which is quadratic in the α_ν^* and linear in the β_ν^*: Using (12a) we have

$$(12b) \qquad \mathrm{Re}\,\{e^{-i\omega_\nu} \psi_\nu\} + \frac{1}{2} \left| \frac{\tau_\nu}{\rho_\nu'} \right|^2 \leqq 1,$$

where ψ_ν denotes the coefficient of H^2 in (11). Since the corrector must not be iterated according to our analysis, the order q of the predictor must be no less than $p - 1$ if the original order p of the corrector is to be maintained for the predictor-corrector scheme (6) with $m = 1$ (see, e.g., [1, p. 259 ff.]). The requirement of a certain order q for the predictor generates $q + 1$ homogeneous linear relations between the α_ν^* and β_ν^*. Thus the following procedure seems appropriate for the determination of a suitable (ρ^*, σ^*) for a given weakly unstable scheme (2): Evaluate (12a) in terms of the α_ν^*, then express ρ^* and σ^* in terms of the free parameters (if any) which are left after accounting for the order relations and (12a). Then interpret (12b) as a restriction in the space of these free parameters (or check its validity).

Remark: The same considerations can be carried through for $P(EC)^1$ algorithms. However, the details are more involved.

4. Application. For *Milne-Simpson's 2-step scheme* (1) we have $\rho = z^2 - 1$, $\sigma = (z^2 + 4z + 1)/3$, $p = 4$, and $\zeta_{20} = -1$. As we have to require $q = 3$, it seems futile to look for a stabilizing predictor with $k = 2$ since the order relations *alone* determine ρ^*, σ^* in this case:

$$\rho^* = z^2 + 4z - 5, \qquad \sigma^* = 4z + 2. \tag{13}$$

Yet by a marvelous coincidence this *is* a predictor which does the trick:

$$-1 \cdot \frac{\sigma(-1) + \beta_2 \rho^*(-1)}{\rho'(-1)} = +1,$$

$$-\psi(-1) + \frac{1}{2}\left(\frac{\tau_2}{\rho_2'}\right)^2 = -\frac{1}{3} < 1.$$

Therefore the algorithm

$$\begin{aligned} y_{n+2}^{(0)} &= -4y_{n+1} + 5y_n + 2h(2f_{n+1} + f_n), \\ y_{n+2} &= y_n + \frac{h}{3}(f_{n+2}^{(0)} + 4f_{n+1} + f_n) \end{aligned} \tag{14}$$

is a genuine 2-step method of order 4 which is strongly stable for arbitrary H (as it turns out), i.e., it can be safely used for $g < 0$ as well as for $g > 0$. Numerical results which have been obtained with (14) are shown in Section 5.

Admitting 3-step predictors, we could at first try to achieve $q = 4$: All predictors

$$\rho^* = z^3 + (8 + \alpha_0^*)z^2 - 9z - \alpha_0^*,$$

$$\sigma^* = [(17 + \alpha_0^*)z^2 + (14 + 4\alpha_0^*)z - (1 - \alpha_0^*)]/3$$

are of order 4 (see, e.g., [6, p. 201]), so it seems that we have one parameter left for the satisfaction of (12). However, upon introduction of the above ρ^* into (12a), the parameter α_0^* drops out and the necessary condition cannot be satisfied: There is no stabilizing 3-step predictor of order 4. Among the 3-step predictors with $q = 3$ the following one-parameter family is found to be stabilizing:

$$\begin{aligned} \rho^* &= z^3 + (4 + \alpha_0^*)z^2 - 5z - \alpha_0^*, \\ \sigma^* &= [(12 + \alpha_0^*)z^2 + (6 + 4\alpha_0^*)z + \alpha_0^*]/3, \end{aligned} \qquad \alpha_0^* > -3. \tag{15}$$

For $\alpha_0^* = 0$, which is well within the stabilizing region, we recover our 2-step predictor (13). Since the error term of (15) is $h^4 y^{\mathrm{IV}}/6$ *independently of* α_0^* there is no indication why one should not choose the simpler predictor (13) and discard the 3-step predictors.

5. Comparison with Runge-Kutta, Numerical Results. In the case of an equation $y' = gy$, $g =$ const, the relative discretization error

$$e_r(x_n, h) := (y_n(h) - y(x_n))/y(x_n)$$

will behave approximately[6] like $Cg^5(x - x_0)h^4$ with

$$C = \begin{cases} +\frac{1}{180} & \text{for the exact solution of (1),} \\ -\frac{1}{45} & \text{for the stabilized scheme (14),} \\ -\frac{1}{120} & \text{for the classical Runge-Kutta method.} \end{cases} \tag{16}$$

[6] (16) takes into account the first term of the asymptotic expansion of the discretization error under the assumption that the initial errors are $O(h^5)$. For the values of C, see, e.g., [1].

TABLE 1
Relative discretization error $e_r(x, h)$ for $y' = -y$

x	(14) $h = 2^{-2}$	R.-K. $h = 2^{-1}$	h	$x = 10$ (14)	$x = 10$ R.-K. (with $2h$)
2	.000 244	.001 585	2^{-1}	.0357 1363	.2113 1609
4	493	3 172	2^{-2}	.0012 4629	.0079 4948
6	744	4 762	2^{-3}	6407	4 0130
8	995	6 355	2^{-4}	377	2260
10	1 246	7 949	2^{-5}	16	138
12	1 498	9 547	2^{-6}	1	7
14	1 748	11 152			
16	1 999	12 786			
18	2 251	14 390			
20	2 503	16 002			

TABLE 2
Relative discretization error $e_r(x, h)$ for $y' = -y^2$

x	(14) $h = 2^{-5}$	R.-K. $h = 2^{-4}$	h	$x = 10$ (14)	$x = 10$ R.-K. (with $2h$)
5	$36.7 \cdot 10^{-9}$	$34.9 \cdot 10^{-9}$	2^{-1}	.0014 52234	−.0053 07526
10	20.0	19.2	2^{-2}	96792	+ 18899
15	13.9	13.4	2^{-3}	5657	4237
20	10.6	10.4	2^{-4}	334	299
			2^{-5}	20	19
			2^{-6}	1	1

Obviously, the stabilization of (1) has to be paid for by a loss in accuracy such that the stabilized version of (1) is less accurate than R.-K. However, basing the comparison on an equal number of evaluations of f for a given interval of integration (see [4]) we find that the error of (14) is only $\frac{1}{6}$ of that for R.-K. Hence we may expect that (14) is a rather effective fourth order method for the numerical integration of ordinary differential equations.

The following differential equations were solved by the predictor-corrector scheme (14) and by R.-K.: (a) $y' = -y$, (b) $y' = -y^2$, each with $y(0) = 1$, for $x \leqq 20$. The value of $y(h)$ for scheme (14) was computed by *one* execution of R.-K.; this introduces an error of $O(h^5)$.

It is clear that the usual Milne-Simpson algorithm would have failed on both equations over such a long interval.[7] With algorithm (14) not the least sign of an oscillation or an undue round-off accumulation was found on either differential equation. As to be expected from (16), for eq. (a) the error with (14) was less than

[7] Although for eq. (b) the oscillations will grow only like $h(x + 1)^{8/3}$ relative to the basic discretization error, this constitutes an intolerable disturbance for large x.

20% of that with R.-K. (and equal effort) throughout the interval and for all stepsizes used. Some numerical values are shown in Table 1.

For the nonlinear equation (b), the errors of (14) and R.-K. were practically equal for small stepsizes. For very large steps R.-K. was poorer, with decreasing h the discretization error *changed its sign* and became smaller (see Table 2). (This effect is caused by the complicated error terms of R.-K. which contain various derivatives of different order.) Due to this unsystematic behavior of the discretization error Richardson-extrapolation was *not applicable for R.-K.* while it worked well for (14) where the error decreased like h^4 approximately for large and small h.

Acknowledgment. The preparation of this paper was stimulated through some results obtained by A. Neiss in a systematic experimental investigation on the effect of predictor-corrector interaction upon stability [7].

Mathematisches Institut
Technische Hochschule
München, Germany

1. P. Henrici, *Discrete Variable Methods in Ordinary Differential Equations*, Wiley, New York, 1962. MR **24** #B1772.
2. W. E. Milne & R. R. Reynolds, "Stability of a numerical solution of differential equations. I & II," *J. Assoc. Comput. Mach.*, v. 6, 1959, p. 196–203 and v. 7, 1960, p. 46–56.
3. G. Dahlquist, "Convergence and stability in the numerical integration of ordinary differential equations," *Math. Scand.*, v. 4, 1956, p. 33–53. MR **18,** 338.
4. T. E. Hull & A. L. Creemer, "Efficiency of predictor-corrector procedures," *J. Assoc. Comput. Mach.*, v. 10, 1963, p. 291–301. MR **27** #4367.
5. H. S. Wilf, "Maximally stable numerical integration," *J. Soc. Indust. Appl. Math.*, v. 8, 1960, p. 537–540.
6. R. W. Hamming, *Numerical Methods for Scientists and Engineers*, McGraw-Hill, New York, 1962. MR **25** #735.
7. A. Neiss, *Untersuchungen zur Stabilität von Predictor-Corrector-Verfahren*, Diplomarbeit, Techn. Hochschule München, 1964.