Correlated round-off errors in digital integrating differential analyzers

By C. S. Wallace*

The integration method used in digital differential analyzers suffers from both truncation and round-off errors. It is shown that, if trapezoidal corrections are employed, the latter dominates. The round-off errors in successive steps of the integration are shown to be correlated whenever the function being integrated has a slope which passes through a rational fraction of small denominator. However, an analysis is presented to show that in general the correlation does not greatly affect the total round-off error. Some special cases are shown to suffer from anomalously high round-off error.

1. Introduction

The basic operation in a digital differential analyzer (D.D.A.) is the integration, by adding to an accumulator in every clock cycle, of a variable represented by a number which changes by at most one in the least-significant digit each clock cycle. If h be the step size of the independent variable (usually time), well-known techniques of trapezoidal integration (see, for example, Bradley and Genna, 1962) can reduce truncation error to the order of h^2 per unit time. A naïve analysis based on the assumption of uncorrelated errors from step to step suggests that round-off errors will be of order $h^{1\cdot 5}$ per unit time, and hence dominant. It is therefore important to examine these errors more closely. It will be shown that round-off errors can be strongly correlated whenever the rate of change of the variable being integrated approximates to a simple rational fraction. However, it is found that, except in certain well-defined exceptional cases, this correlation does not greatly affect the magnitude of the final error.

2. The basic integration process

The D.D.A. is in its operation almost exactly analogous to the perhaps more familiar mechanical analogue computer comprising many wheel-and-disc integrators, more particularly to the form in which incremental motions are transmitted among the integrators by commutators and driving stepping motors. In the mechanical integrator the output is a sequence of electric pulses announcing incremental rotations dr of the wheel caused by incremental rotations dx of the disc it touches.

In the usual mechanization, dr and dx can be either zero or a fixed positive or negative increment. The average value of dr/dt and dx/dt can be thought of as the rate of occurrence of the fixed-magnitude non-zero increments. The ratio of these rates (i.e., the average value of dr/dx) is governed by the distance y between the wheel and the centre of the disc it touches. Thus, approximately, $\frac{dr}{dt} = y\frac{dx}{dt}$, i.e. $dr = y\,dx$, and the integrator effectively integrates y with respect to x. The value of y is controlled by a stepping motor whose input dy is generally the dr output of another integrator.

In the D.D.A., the quantity y is held as a binary number in a register. Non-zero dx inputs cause y to be added or subtracted from an accumulator whose overflow or underflow pulses are the dr outputs. The register holding y is constructed as an up-or-down counter which accumulates dy input pulses. Clearly, the operation is entirely analogous to the wheel-and-disc, and again the ratio of the average values of dr/dt and dx/dt is y. The integration $dr = y\,dx$ is performed. (See Fig. 1.)

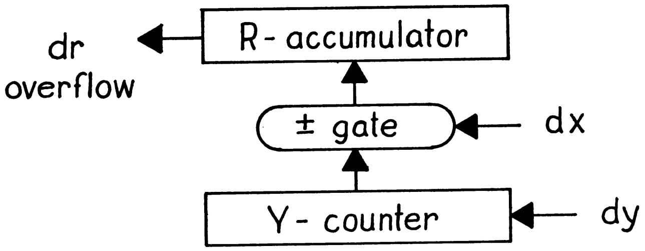

Fig. 1.—Block diagram of a D.D.A. integrator

In the usual notation (Bradley and Genna, 1962), each integrator of a D.D.A. comprises a Y register of capacity ± 1 and an R register of capacity $\pm\frac{1}{2}$. The Y register holds the instantaneous value of a time-dependent variable y in the range -1 to 1. In each machine cycle, representing an increment h in time, y is added to R. Overflows from R (positive or negative) represent increments of magnitude h in the time integral of y. Also, in each machine cycle, y, which is itself being generated as the time integral of other variables, may be increased or diminished by an amount h as the result of overflows from other R registers. (In practice, the increments of y may be a few times h if y is being generated as the sum of several integrals.) If trapezoidal corrections are used, and at the end of a cycle y is increased by dy, then $\frac{1}{2}\,dy$ is added to the R register, thus effectively

* *Basser Computing Department, The University of Sydney, Sydney, N.S.W., Australia.*

causing the average value of y over the cycle to be added to R. In effect, the change in value of the integral of y during the step of duration h is approximated by

$$yh + \tfrac{1}{2}\dot{y}h^2$$

and is hence in error by terms of order $\frac{1}{6}\ddot{y}h^3$. In unit time ($1/h$ steps) the accumulated truncation error will be of order h^2. (A full treatment of truncation errors is given by Nelson (1962).)

3. Simple round-off estimate

If it is assumed that the increments dy are chosen without error to produce an unbiassed approximation to y, so that the value of y held in Y at the end of each cycle differs by $h/2$ at most from the true value of the variable, then in each step a round-off error of at most $(h/2)h$ is introduced into the integral. If one takes as the R.M.S. value of the error $h^2/2\sqrt{3}$, and assumes that errors are completely uncorrelated from step to step, then after $1/h$ steps, the expected accumulated round-off error in the integral of y would be $(h^2/2\sqrt{3})/h$ or about $0{\cdot}29h^{1{\cdot}5}$. This is substantially larger than the truncation error, and will therefore be the dominant error. (The effect of truncation and round-off errors introduced by errors in the formation of the dy increments, that is, the interaction and propagation of errors throughout the D.D.A., will depend on the stability of the equations being solved, and is beyond the scope of this paper. However, except in pathological situations, these effects should not affect the order of magnitude of the final round-off error.)

Now, the assumption that the round-off errors in y are uncorrelated from step to step is clearly invalid. To take a simple instance, suppose that y passes through a maximum. In doing so, it may spend some considerable time within a range of values say from $(n - \frac{1}{2})\,h$ to nh (n integral). Throughout this time, it will be represented in Y by the value nh, and so there will be a large number of consecutive steps all introducing a positive round-off error. (See Fig. 2.)

4. Correlated errors

Consider a variable y whose true value is given by the function

$$y = (p/q)t + e, \qquad (1)$$

where p and q are mutually prime integers, t is time, and e is a constant.

The approximate value y_a held in Y will never differ from y by more than $h/2$, and will experience a train of

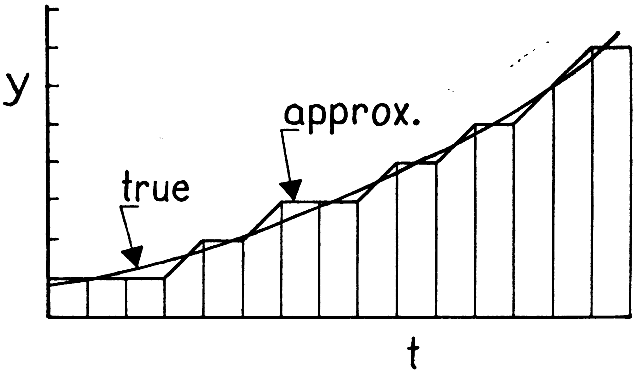

Fig. 2.—The D.D.A. trapezoidal approximation to the integral of a function

dy increments which is periodic with period q steps. For example, if $p = 2$, $q = 5$, the dy increments will repeat the cycle

$$0\,1\,0\,1\,0,\ 0\,1\,0\,1\,0,\ 0\,1\,0\,1\,0, \text{ etc.}$$

Writing $t = sh$, where s is the (integral) step number, (1) can be rewritten

$$y = h(ps)/q + kh. \qquad (2)$$

Writing $y_a = nh$, where n is integral, n can be obtained as the integer nearest to the quotient $(ps + qk)/q$. The error $y_a - y$ is then h times the remainder (positive or negative) consistent with the rounded quotient, i.e. $h((ps + qk)/q - n)$. Since p and q are mutually prime, as s increases, the remainder will cycle in some order through q values equally spaced throughout the range $-\frac{1}{2}$ to $+\frac{1}{2}$ with spacing $1/q$. For instance, if $y = h\ (2s/5 + 0{\cdot}05)$, the pattern of remainders will be

s	0	1	2	3	4	5	6	7	etc.
remainder = error/h	$-0{\cdot}05$	$-0{\cdot}45$	$0{\cdot}15$	$-0{\cdot}25$	$0{\cdot}35$	$-0{\cdot}05$	$-0{\cdot}45$	$0{\cdot}15$	etc.

The remainders can then be written as $-f$, $-f + \frac{1}{q}$, $-f + \frac{2}{q}$, $-f + \frac{3}{q}$, ... (This is not necessarily the order in which the remainders occur.) Now

$$-\frac{1}{2} < -f, \quad -f + \frac{q-1}{q} < \frac{1}{2}$$

whence

$$\frac{1}{2} > f > \frac{1}{2} - \frac{1}{q}. \qquad (3)$$

The value of f will depend on e in (1) or k in (2). Suppose k is such that $f = \frac{1}{2}$. Then, if so, no remainder in the division is greater than $\frac{1}{2} - 1/q$. Hence, k could be increased by nearly $1/q$ without changing any of the quotients in the division and still leave all remainders in range. In other words, the function y could be increased by nearly h/q without affecting the approximation y_a.

Thus a family of functions having the same slope but having values for a given s, covering a range h/q, will be approximated by the same y_a. Considerations of symmetry show that the function in the middle of the range will be approximated with zero average error. In fact, it can easily be shown that its integral will never be in error by more than $qh^2/4$, and this value can be reached only when $p = 1$. The extreme functions in the range will be approximated with an average error of magnitude $h/2q$. This error would introduce an error on the integral of $h^2/2q$ per step, always with the same sign. For an example, see Fig. 3.

5. Magnitude of correlation

The above result may be recast thus. If a linear function $y = (p/q)t + e$ is integrated by a D.D.A., the integral obtained will differ by not more than $h^2q/4$ from the true integral of some other parallel linear function $y_b = y + g$, where $|g| < h/2q$, and will hence accumulate error at an average rate of hg per step.

Now suppose y is not a linear function of t, but happens at some time, say $t = 0$, to have a slope $\dot{y} = p/q$. Let y_b be the linear function with slope p/q differing from y at $t = 0$ by less than $h/2q$ and such that it would be approximated in Y with zero average error.

Then, throughout the interval surrounding $t = 0$ within which $|y - y_b| < h/2q$, the sequence of values y_a by which y is approximated in Y will be the same as that by which y_b would be approximated. Thus, the integral of y over the interval will differ from the true integral of y_b over the interval by at most $h^2q/4$.

To find the consequences of this, suppose at $t = 0$, that $y = y_b + h/2q - z$, $\dot{y} = p/q$, and $\ddot{y} \gg h$, where $0 < z < h/q$. Then y can be written (neglecting higher derivatives) as

$$y_b + h/2q - z + \tfrac{1}{2}\ddot{y}_0 t^2. \qquad (4)$$

The interval during which $|y - y_b| < h/2q$ is terminated when

$$\tfrac{1}{2}\ddot{y}_0 t^2 = z, \text{ i.e. when } t = \pm\,(2z/\ddot{y}_0)^{1/2}. \qquad (5)$$

The integral $\int (y_b - y)dt$ over the interval is thus

$$2\int_0^{(2z/\ddot{y}_0)^{1/2}} (-h/2q + z - \tfrac{1}{2}\ddot{y}_0 t^2)dt$$
$$= (1/3)(4z - 3h/q)(2z/\ddot{y}_0)^{1/2}. \qquad (6)$$

Since z is of order h/q, this expression is of order $(h/q)^{1\cdot5}$.

Since the integral formed by the D.D.A. differs from the integral of y_b over the interval by less than $h^2q/4$, the expression (6) represents the error in the integration of y over the interval to an accuracy of $h^2q/4$, and may be taken as an adequate estimate of the error provided,

$$h^2q \ll (h/q)^{1\cdot5}$$

i.e., provided

$$q^{2\cdot5} \ll (1/h)^{0\cdot5}. \qquad (7)$$

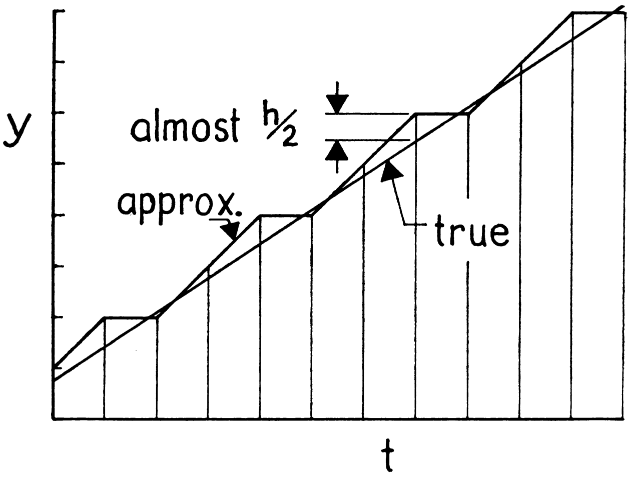

Fig. 3.—An example showing a consistent overestimation of the integral of a linear function of slope two-thirds

In typical applications of D.D.A.'s, $h \sim 10^{-6}$ to 10^{-8}, so (6) is an adequate estimate at least for $q < 10$.

Now one error of the form (6) occurs whenever the slope of y passes through a rational fraction of small denominator. To estimate the overall round-off error in an integration one can only suppose, in the absence of detailed knowledge of z and $\ddot{y}_0$ on each such occasion, that all values of z between 0 and h/q are equally probable, and that $|\ddot{y}_0|$ may be approximated by some average value R over the whole period of integration.

The squared error expected for a particular p/q, assuming a uniform distribution of z between 0 and h/q, is found from (6) to be

$$(q/h)\int_0^{h/q} (1/9)(4z - 3h/q)^2(2z/R)dz$$
$$= (1/9R)(h/q)^3. \qquad (8)$$

If the reasonable assumption is made that the signs and magnitudes of the errors occasioned by different values of p/q are uncorrelated, the average total squared error in unit time over some population of integrals can be found by summing the squared errors for the different p/q, after multiplying each by the number of times $\dot{y}$ takes the particular value p/q in unit time. In D.D.A.'s, $|\dot{y}|$ is usually made by choice of scaling to range between 0 and about 1. We will therefore consider only values of $p \leqslant q$. If one considers the integration of a periodic function such as $y = \sin t$, in each period of the function, there are four occasions when $|\dot{y}| = p/q$ for all p/q except 0/1 and 1/1, which each occur twice. We therefore estimate the frequency of occurrence of the value p/q as S, for $q \geqslant 2$, and $\frac{1}{2}S$ for $q = 1$, where S is some average value of $|\ddot{y}|$.

The sum of the squared errors in unit time can then be estimated by

$$S(h^3/9R)\left(\frac{1}{2} + \frac{1}{2} + \sum_{q=2}^{\infty} n(q)/q^3\right) \qquad (9)$$

where $n(q)$ is the number of integers less than q and mutually prime with respect to q, and the terms $\frac{1}{2}$ account for the cases 0/1 and 1/1.

The sum may be written as $S(h^3/9R)Z(2)/Z(3)$, where Z is the Riemann zeta-function (Hardy and Wright, 1938), and equals

$$0 \cdot 152h^3(S/R). \qquad (10)$$

The series in (9) converges fairly rapidly. Hence the restriction (7) on the validity of (6) and (8) to small values of q is of no great moment.

While S and R are differently weighted averages of $|\ddot{y}|$, S/R would generally be of order one. Thus the R.M.S. correlation error in unit time can be estimated as

$$0 \cdot 39h^{1 \cdot 5}.$$

This value is not much larger than the naïve estimate of $0 \cdot 29h^{1 \cdot 5}$, so it would appear that the effect of round-off error correlations is not serious.

6. Special cases

If the function y has, say, at time $t = 0$, contact of order n with a linear function $(p/q)t + e$, where n is greater than 2, then the expansion (4) of y about $t = 0$ breaks down, and must be replaced by

$$y = y_b + h/2q - z + (1/n!)y_0^{(n)}t^n.$$

In this event, the interval corresponding to (5) becomes $t = \pm (n!z/y_0^{(n)})^{1/n}$ and the integral corresponding to (6) has a value of order

$$(h/q)^{(1+1/n)}.$$

Thus such a case can give rise to a round-off error, from the neighbourhood of the high-order contact, which is larger than the estimate of the total round-off error in functions with no such singularity.

Thus, integration of a function such as $y = a \sin (t/a)$, which has a third-order contact to lines of slope one whenever $t = n\pi a$, can be expected to suffer from round-off errors anomalously large by a factor $(1/h)^{1/6}$.

Conclusions

Although correlation of round-off errors certainly occur in the integration process used in D.D.A.'s, the final round-off error in unit time is in general adequately estimated by the assumption of uncorrelated errors. However, if the function being integrated has high-order contact with a linear function of time having a slope which is zero or a rational fraction of small denominator, an anomalously large error can occur.

References

Bradley, R. E., and Genna, J. F. (1962). "Design of a One-Megacycle Iteration Rate D.D.A.," *Proceedings of A.F.I.P.S. Spring Joint Computer Conference*, 1962, p. 353.

Nelson, D. J. (1962). "D.D.A. Error Analysis Using Sampled Data Techniques," *Proceedings of A.F.I.P.S. Spring Joint Computer Conference*, 1962, p. 365.

Hardy, G. H., and Wright, E. M. (1938). *An Introduction to the Theory of Numbers* (p. 249), Oxford University Press.

Book review: Automatic control

Theory of Automatic Control, by M. A. Aizerman, 1963; 519 pages. (Oxford: *Pergamon Press Ltd.*, 80s.)

Professor Aizerman of the Moscow Institute of Automatics and Telemechanics is a leading authority on control theory in the U.S.S.R. His reputation stands high in the international field and his work, including the famous "Aizerman Problem", is well known in the West. It is therefore with great interest and expectation that one welcomes the first English translation of this work based on the author's lectures delivered to non-specialists in the Institute. It is the 1958 second edition containing substantial changes and additions which has been translated. A British control engineer, Dr. Freeman, has helped to ensure acceptable scientific phraseology.

For the ground covered the book is large, about a quarter of a million words, and is divided into five equal sections. The first is entirely descriptive dealing with various types of controllers and their characteristics. The author writes with a process-control orientation, regarding the whole subject as concerned with the design of instruments called *automatic controllers* which are applied to a process. The section concludes with a description of the various methods by which self-adaptive systems seek conditions of optimum performance.

The next section is of great importance on a vital issue sadly neglected in the majority of texts, and is concerned with the construction of an adequate linear mathematical model of the actual control process. The author stresses the inevitable non-linear characteristics of control systems and shows the essential part that linear analysis has to play in the understanding of system dynamics and stability. The third section deals with stability in a comprehensive manner including a proof of the Hurwitz-Routh conditions. The fourth section is concerned with design, chiefly for the parameter estimation to give transient performance with minimum integral error squared. There is also the author's own approximate method for determining the transient behaviour of high-order systems.

The last and by far the most interesting section takes essential non-linearities into account and deals with auto- and forced oscillations in non-linear systems. The word "harmonic" is used as a literal translation from the Russian but really means "pure sine wave". This is disconcerting when dealing with Fourier series representation of period waveforms.

Considering that the book was written in 1958, it is well up to date, including statistical work for random processes. Its main interest lies in its presentation, in that a quite different approach is made in the teaching of standard control theory.

John C. West.