An error analysis of Goertzel's (Watt's) method for computing Fourier coefficients

By W. M. Gentleman*

Goertzel's method, also known as Watt's algorithm, is one of the three standard methods of computing Fourier coefficients, and is especially commonly used when only a small number of coefficients is desired for a given sequence. This paper gives a floating-point error analysis of the technique, and shows why it should be avoided, particularly for low frequencies.

(Received November 1967, revised November 1968)

Given a sequence $f_j, j = 0, 1, \ldots, N-1$, we often require the finite Fourier coefficients, defined for any frequency ω by

$$\left.\begin{aligned} a(\omega) &= \sum_{j=0}^{N-1} f_j \cos(j\omega) \\ b(\omega) &= \sum_{j=0}^{N-1} f_j \sin(j\omega). \end{aligned}\right\} \quad (1)$$

There are three standard computational methods available.

(i) Direct evaluation of the defining formula (1). This requires N multiplications and $N-1$ additions for each coefficient. Moreover, it requires N sines and N cosines.

(ii) A method due to Goertzel (1958) and Watt (1959) (see also Goertzel (1960), Hamming (1962) and Ralston (1965)) involving the derived sequence

$$u_k = f_k + 2\cos(\omega)u_{k+1} - u_{k+2}, u_N = u_{N+1} = 0 \quad (2)$$

from which the Fourier coefficients can be obtained as

$$\left.\begin{aligned} b(\omega) &= u_1 \sin(\omega) \\ a(\omega) &= f_0 + \cos(\omega)u_1 - u_2. \end{aligned}\right\} \quad (3)$$

This is attractive not only because it requires half as many multiplications as method (i), but also because it requires only one sine and one cosine.

(iii) The fast Fourier transform (Cooley and Tukey, 1965) which computes the complete transform (i.e. all frequencies of the form $2\pi t/N$ for $t = 0, 1, \ldots, N-1$) by factorising the equivalent matrix multiplication. When the chosen factorisation is $N = \prod_{i=1}^{k} n_i$, this requires $C_1 N \sum_{i=1}^{k} n_i$ multiplications and additions and $C_2 N$ sines and cosines, where C_1 and C_2 are constants depending on programming details and are normally about 1 and $1/\max(n_i)$ respectively (Gentleman and Sande, 1966).

When the complete transform is required, the fast Fourier transform clearly takes less operations, being $O(N \log N)$ rather than $O(N^2)$ as are the other methods. If, however, only a few frequencies are required, the other methods may be more efficient. Moreover, the fast Fourier transform cannot be used if the frequencies are not of the form $2\pi t/N$, and unlike the other methods, requires space to accommodate all the data at one time.

In view of this there is still considerable interest in the first two methods.

As well as the considerations above, comparisons of roundoff error are important, especially as the computations are often done for long sequences on special-purpose hardware of low accuracy.

Gentleman and Sande (1966) used the ratio ρ of the root mean square (rms) error of the transformed sequence to the rms of the transformed sequence itself as a measure of error in the computation. For a computer using floating-point arithmetic with Wilkinson's (1963) rounding conventions, where ϵ is the largest number that can be added to unity and still produce a result indistinguishable from unity (e.g. $\epsilon = 2^{-b}$ in a binary machine with a mantissa of b bits), they showed $\rho < 1{\cdot}06\,(2N)^{3/2}\epsilon$ for method (i), and $\rho \leqslant 1{\cdot}06 \sum_i (2n_i)^{3/2}\epsilon$ for method (iii).

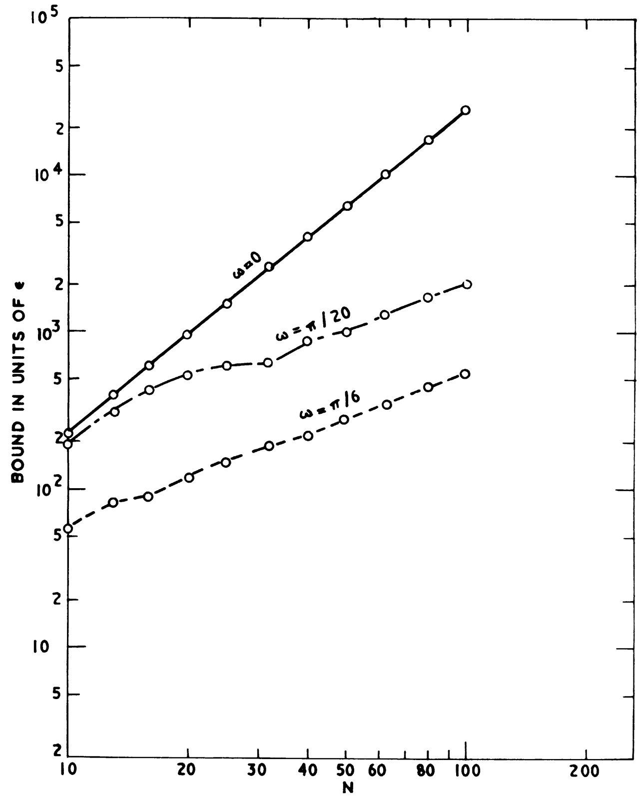

Fig. 1. Bound on $||\delta||/||f||$ in units of ϵ for ω fixed

* *Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey*
(Temporarily at Division of Numerical and Applied Mathematics, National Physical Laboratory, Teddington, Middlesex.)

Experiments on random sequences showed that the actual values of ρ are typically proportional to N for method (i), but proportional to the bound itself for method (iii). The observed values of ρ for method (ii) suggested that the best bound for this method might be proportional to N^2, but as there is practical experience (e.g. Thatcher, 1964) that method (ii) suffers from more roundoff at low frequencies than high, ρ is not a satisfactory error measure for this method and so a different analysis is given here. (Since the errors for methods (i) and (iii) do not exhibit a frequency dependence, ρ is a satisfactory measure for them.)

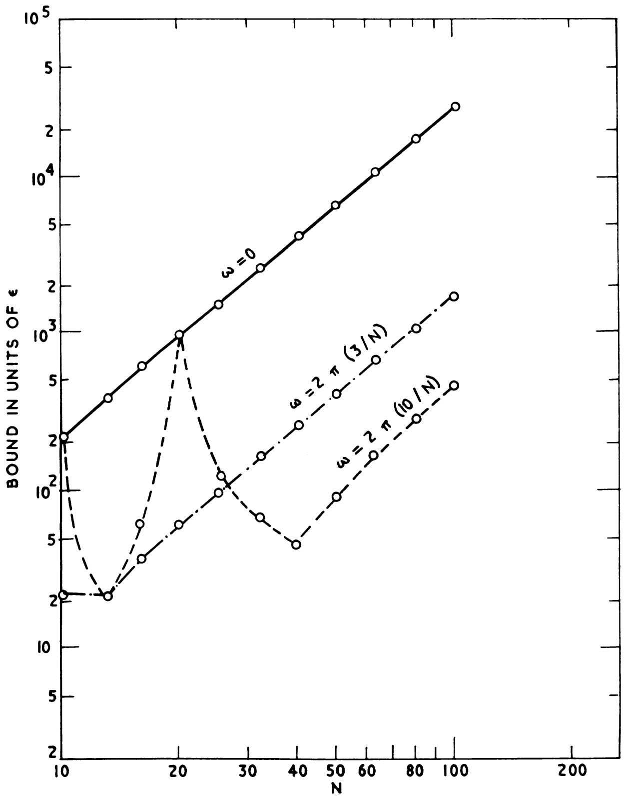

Fig. 2. Bound on $||\delta||/||f||$ in units of ϵ for ω inversely proportional to N

Derivation of the bound

We shall analyse the effects of roundoff by backward error analysis. That is, we shall find an effective sequence $f + \delta(\omega)$ such that the sequence u (and the a and b derived from u) obtained by floating-point computation from the true sequence f is also what would have been computed from the effective sequence had exact arithmetic been done. The ratio

$$\rho(\omega) = \text{rms}\,(\delta(\omega))/\text{rms}\,(f) = ||\delta(\omega)||_{L2}/||f||_{L2}$$

is then somewhat comparable to the ratio ρ used in the previous analysis.

At this point it is relevant to point out that Goertezl's method is just an extension of Clenshaw's method (1955) for evaluating Chebyshev series by three-term recurrence. (In fact, the cosine coefficient $a(\omega)$ corresponds to the Chebyshev series sum.) Both Clenshaw (1955) and Watt (1959) comment on the rounding error of the algorithm, but since they were thinking in terms of fixed-point arithmetic, they remark that the error involved in evaluating each term u_k arises wholly from taking the product $2\cos(\omega)u_{k+1}$, and is at most one bit in the last place. This is equivalent to changing f_k by at most one bit in the last place, and one feels the algorithm is very satisfactory.

This result is somewhat misleading, however, as we have ignored the extent to which the f_k must be scaled down to prevent the u_k from overflowing. This is more obvious when we consider floating-point arithmetic. Here, if the computed u_k is written

$$\begin{aligned} u_k &= fl(f_k + 2\cos(\omega)\,u_{k+1} - u_{k+2}) \\ &= f_k + 2\cos(\omega)u_{k+1} - u_{k+2} + \delta_k \\ &= (f_k + \delta_k) + 2\cos(\omega)u_{k+1} - u_{k+2} \end{aligned} \tag{4}$$

where δ_k is the error in this computation, we can again regard this error as an effective change in the element f_k. There are several orders in which the evaluation can be performed, but in all cases we now have

$$|\delta_k| \leqslant 3 \times 1{\cdot}06\epsilon\{|f_k| + |2\cos(\omega)u_{k+1}| + |u_{k+2}|\}. \tag{5}$$

For example, using the better of Wilkinson's rounding rules

$$\begin{aligned} t_1 &= fl(2\cos(\omega) \times u_{k+1}) = 2\cos(\omega)u_{k+1}(1 + \xi_1) \\ t_2 &= fl(f_k + t_1) = (f_k + t_1)(1 + \xi_2) \\ u_k &= fl(t_2 - u_{k+2}) = (t_2 - u_{k+2})(1 + \xi_3) \\ &\qquad \text{with } |\xi_1|, |\xi_2|, |\xi_3| \leqslant \epsilon \end{aligned}$$

yields

$$\delta_k = f_k\{(1 + \xi_2)(1 + \xi_3) - 1\} + 2\cos(\omega)u_{k+1}$$
$$\{(1 + \xi_1)(1 + \xi_2)(1 + \xi_3) - 1\} - u_{k+2}\{(1 + \xi_3) - 1\}$$

so

$$|\delta_k| \leqslant |f_k| 2 \times 1{\cdot}06\epsilon + |2\cos(\omega)u_{k+1}| 3 \times 1{\cdot}06\epsilon + |u_{k+2}|\epsilon.$$

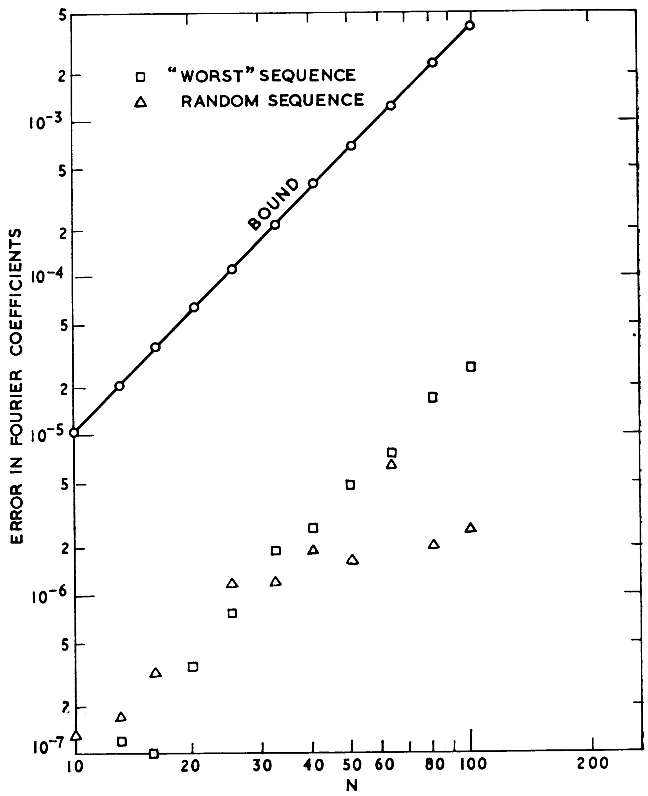

Fig. 3. Observed ratios of error norm to sequence norm, together with bound, for $\omega = 0$

If we think of the sequences as vectors, equation (5) states that the vector whose elements are $|\delta_k|$ is less, element by element, than $3 \times 1{\cdot}06\,\epsilon$ times the sum of three vectors whose elements are $|f_k|$, $|2\cos(\omega)\, u_{k+1}|$ and $|u_{k+2}|$ respectively. For such vector norms as L_1, L_2 and L_∞, that implies

$$||\delta|| \leqslant 3 \times 1{\cdot}06\epsilon\{||f|| + |2\cos(\omega)|\; ||u'|| + ||u''||\} \quad (6)$$

where u' and u'' are the vectors formed by the sequence u, shifted once and twice respectively. Clearly for these norms $||u''|| \leqslant ||u'|| \leqslant ||u||$. (There are end effects in equation (6) associated with the definitions of δ_{N-1}, δ_{N-2}, δ_1 and δ_0, but these are negligible compared to the main contributions to roundoff error, and we will ignore them.)

Thus we have

$$||\delta|| \leqslant 3 \times 1{\cdot}06\epsilon\{||f|| + (1 + |2\cos\omega|)||u||\}. \quad (7)$$

We will now replace u in this bound by U, the sequence which would have been computed from f had exact arithmetic been used. We would expect that when the error in the method is acceptably small, $||u||$ is close to $||U||$.

We recall here that in proving the validity of Goertzel's method, the explicit form of U_k is derived, namely

$$U_k = \sum_{i=k}^{N-1} \frac{\sin(\omega(i-k+1))}{\sin(\omega)} f_i. \quad (8)$$

This can be interpreted as defining the vector U to be the product of a matrix $B(\omega)$ with the vector f, where

$$B_{k,\,i}(\omega) = 0 \qquad i < k$$
$$= \frac{\sin(\omega(1-k+1))}{\sin(\omega)} \qquad i \geqslant k. \quad (9)$$

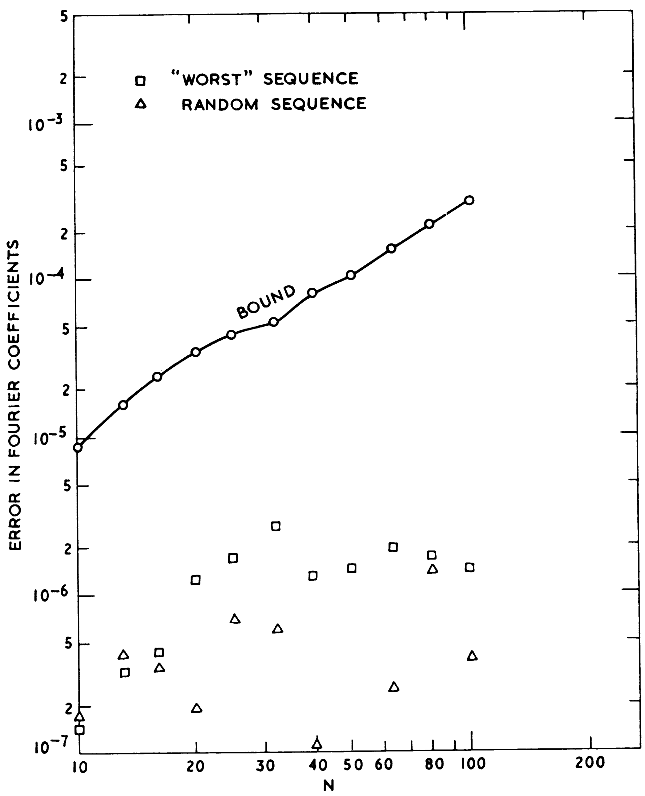

Fig. 4. Observed ratios of error norm to sequence norm, together with bound, for $\omega = \pi/20$

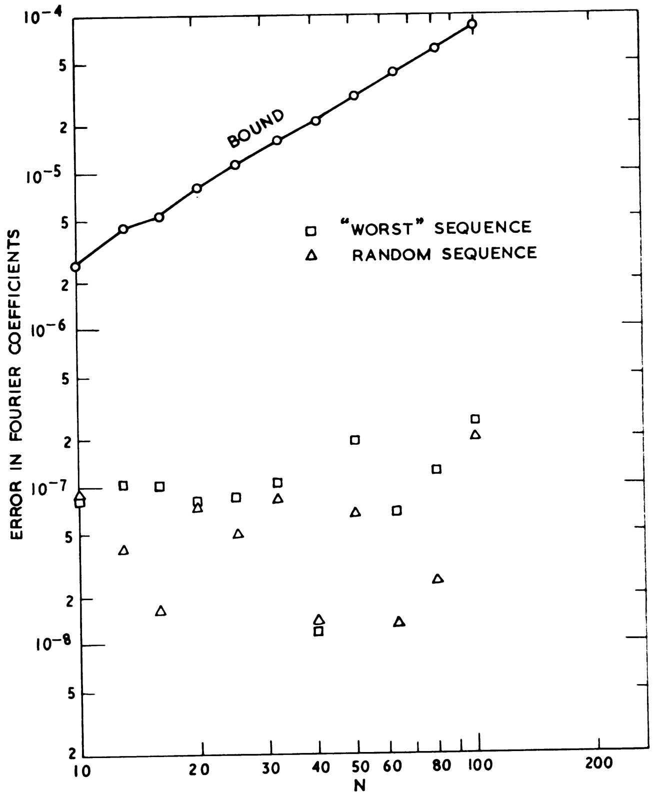

Fig. 5. Observed ratios of error norm to sequence norm, together with bound, for $\omega = \pi/6$

In view of this interpretation

$$||U|| \leqslant ||B(\omega)||\; ||f||, \text{ i.e. } ||U|| \leqslant \beta(\omega)||f|| \quad (10)$$

where $\beta(\omega)$ is the norm of $B(\omega)$. In the L_2 norm, the one usually used with Fourier transforms, $\beta(\omega)$ is the largest singular value of $B(\omega)$, that is, the square root of the largest eigenvalue of B^TB, and the equality in (10) is achieved when f is the corresponding eigenvector of B^TB.

Combining all this with equation (7) gives

$$||\delta|| \leqslant 3 \times 1{\cdot}06\epsilon\{1 + (1 + |2\cos\omega|)\beta(\omega)\}||f|| \quad (11)$$

or

$$\rho(\omega) \leqslant 3 \times 1{\cdot}06\epsilon\{1 + (1 + |2\cos\omega|)\beta(\omega)\} \quad (11')$$

in which we see that the behaviour of the bound is determined by the behaviour of $\beta(\omega)$.

Discussion of the bound

The bound of equation (11) gives rise to three questions:

(i) What are some typical values of this bound?

(ii) How does it behave as a function of N in the two cases

(a) ω fixed? (as when we analyse more and more cycles of a periodic phenomenon sampled at a fixed rate)

(b) ω inversely proportional to N? (as when we analyse a fixed number of cycles of a periodic phenomenon sampling at finer and finer spacing)

(iii) How realistic is it? What does it mean in terms of the answers?

The first two figures show the value of this bound plotted in units of the machine precision ϵ on a log–log scale versus N. In **Fig. 1** it is plotted for several constant frequencies, in **Fig. 2** for frequencies inversely proportional to N. Before considering the behaviour as a function of N, we draw attention to the magnitudes of the numbers. Even for N as small as 63, the bound at $\omega = 0$ is larger than $10^4\epsilon$. This means, for instance, that on a machine like the IBM 360 for which $\epsilon \sim 10^{-6}$, we cannot guarantee the effective sequence will agree with the intended sequence to better than one part in a hundred, even for a sequence of 63 points. And since, as we shall see, the bound grows as N^2, this means that the technique should be highly suspect for sequences of length one thousand or more, such as are frequently analysed. Of course this is just the bound, but as we shall see shortly, it correctly predicts the actual situation.

Returning to the question of growth as a function of N, we see that the slopes of the curves in Figs. 1 and 2 suggest that for $\omega = \theta$ (a constant different from zero) the bound eventually grows linearly in N, but if $\omega = \frac{\alpha}{N}$, including the case $\alpha = 0$, the bound eventually grows as N^2. By forming B^TB explicitly and considering the integral equation which the eigenvalue problem $(B^TB)f = \beta^2 f$ approaches in the limit as $N \to \infty$, we can show that this is indeed the case. For $\omega = \theta$, the largest eigenvalue β^2 approaches $\frac{N^2}{\pi^2 \sin^2 \theta}$, and so the bound approaches

$$||\delta|| \leqslant 3 \times 1{\cdot}06\epsilon \frac{(1 + |2\cos\theta|)}{\pi \sin\theta} N||f|| \qquad (12)$$

whereas for $\omega = \frac{\alpha}{N}$, the largest eigenvalue β^2 approaches $\gamma^2 N^4$, where if $\alpha \neq 0$, γ is the largest root of

$$0 = \lambda_1\lambda_2(1 + \cos\lambda_1 \cos\lambda_2) + \alpha^2 \sin\lambda_1 \sin\lambda_2$$

with $\lambda_1 = \sqrt{(\alpha^2 - 1)/\gamma},\ \lambda_2 = \sqrt{(\alpha^2 + 1)/\gamma}$

and if $\alpha = 0$, γ is the largest root of

$$0 = \sinh^2\rho - \sin^2\rho - (\cos\rho + \cosh\rho)^2$$

with $\rho = 1/\sqrt{\gamma}$.

In this case the bound approaches

$$||\delta|| \leqslant 9 \times 1{\cdot}06\epsilon\gamma N^2||f||. \qquad (13)$$

A short table of values of γ is given as **Table 1.**

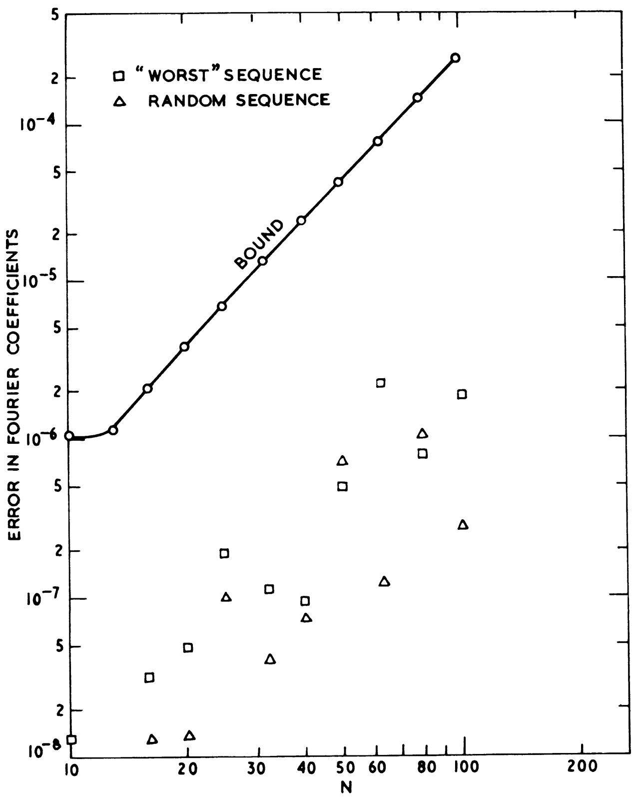

Fig. 6. Observed ratios of error norm to sequence norm, together with bound, for $\omega = 2\pi(3/N)$

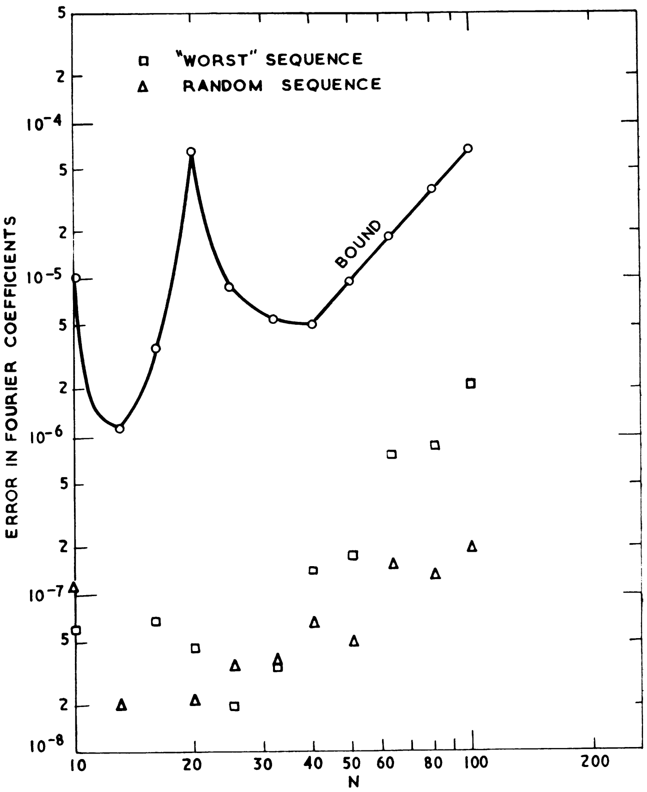

Fig. 7. Observed ratios of error norm to sequence norm, together with bo und, for $\omega = 2\pi(10/N)$

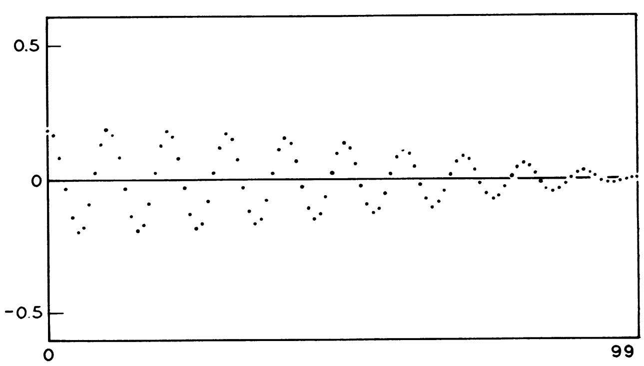

Fig. 8. "Worst" sequence, $N = 100$, $\omega = 2\pi(10/N)$

Table 1

Values of proportionality constant γ for various values of $N\omega = \alpha$

α	γ
0	$28 \cdot 44 \times 10^{-2}$
2π	$6 \cdot 031 \times 10^{-2}$
4π	$2 \cdot 754 \times 10^{-2}$
6π	$1 \cdot 784 \times 10^{-2}$
8π	$1 \cdot 319 \times 10^{-2}$
10π	$1 \cdot 055 \times 10^{-2}$

Evidently the problem, though worst at $\omega = 0$, is not unique to this 'D.C.' term, but exists at other low frequencies as well. This is very unfortunate, since it is most often low frequencies which are of interest. (The behaviour close to $\omega = \pi$ is the same as that close to $\omega = 0$, but these frequencies are not so often of interest.)

So far we have only discussed δ, the difference between the effective and true sequences. But how much can δ affect the Fourier coefficients a and b? It is easy to show that if a sequence is changed by δ, the square root of the squares of the changes in a and b can be as large as $\sqrt{N}||\delta||$. This then gives a bound

$$\left|\left|\begin{matrix} fl(a(\omega)) - a(\omega) \\ fl(b(\omega)) - b(\omega) \end{matrix}\right|\right| \leqslant \sqrt{N}\rho(\omega)||f|| \qquad (14)$$

although we might expect that on the average the error would only be about $\sqrt{\left(\frac{2}{N}\right)}$ times this, as would be the case if $\delta/||\delta||$ were randomly oriented in the unit N-sphere.

Figs. 3 to 7 show actual values of the ratio

$$\left|\left|\begin{matrix} fl(a(\omega)) - a(\omega) \\ fl(b(\omega)) - b(\omega) \end{matrix}\right|\right| / ||f||$$

observed at various frequencies and sequence lengths, for two different kinds of sequences: a random (white noise) sequence, and 'worst' sequence, i.e. the one which maximises the bound, the eigenvector corresponding to β^2. The combination of bounds (14) and (11′) is also plotted. These computations were done on the GE645 computer with the arithmetic carried out so that $\epsilon = 2^{-26} \approx 1 \cdot 5 \times 10^{-8}$. We notice that although the 'worst' sequences are not uniformly those with greatest error, the error in them is consistently high. Moreover, although the bound overestimates the actual error, it appears to be of the correct form for these 'worst' sequences. (In fact, a statistical regression analysis of the observed errors plotted in Figs. 3, 6 and 7 shows that the behaviour of the error for the 'worst' sequences must be like $N^{5/2}||f||$ rather than $N^{4/2}||f||$ or less.) The slower growth of the errors in the random sequences is presumably similar to the earlier $N^{3/2}\sqrt{N}||f||$ results (Gentleman and Sande, 1966) and is related both to the sharpness of $||U|| \leqslant \beta||f||$ and to the sharpness of the relation between $||\delta||$ and the change in the Fourier coefficients. We are, therefore, interested in what this 'worst' sequence looks like. Unfortunately, e.g. **Fig. 8,** far from being a pathological function it is exactly the sort of sequence for which we would be most likely to want to use the method, resembling a slightly damped cos ω!

Summary

Where does this leave us?

(i) We have shown the bound on $\rho(\omega)$ to be asymptotically proportional to N or N^2 depending on whether ω is constant or inversely proportional to N.

(ii) Correspondingly, we have shown the bound on the error in the Fourier coefficients to be proportional to $N^{3/2}||f||$ or $N^{5/2}||f||$.

(iii) We have observed that actual errors appear to follow these bounds, subject to the usual change of scale and division by $\sqrt{N}$, except that with ω inversely proportional to N, the full $N^{5/2}||f||$ appears to be attained by sequences of a kind likely to be of interest.

Reinsch (unpublished) has suggested that for small ω we calculate the u_k by the recurrence rewritten as

$$\begin{aligned} \Delta u_k &= f_k + \Delta u_{k+1} - 2(1 - \cos\theta)u_{k+1} \\ u_k &= u_{k+1} + \Delta u_k. \end{aligned} \qquad (15)$$

This avoids the instability near $\omega = 0$, but near $\omega = \pi$ we must calculate the sums of adjacent elements rather than the differences, and this complicates the program. Without some such modification, however, the catastrophic growth of roundoff errors make Goertzel's method inadvisable unless one can be certain of having adequate precision (such as when the sequence is short or when the frequency is well away from 0 or π).